Métodos Numéricos para Engenharia

Quinta Edição



Steven C. Chapra
Raymond P. Canale

Mc Graw Hill

Métodos Numéricos para Engenharia

QUINTA EDIÇÃO

Steven C. Chapra
Berger Chair em Computação e Engenharia
Universidade de Tufts

Raymond P. Canale
Professor Emérito de Engenharia Civil
Universidade de Michigan

Tradução Técnica
Helena Castro
Professora Doutora do Instituto de Matemática e
Estatística da Universidade de São Paulo — IMEUSP

Versão impressa
desta obra: 2008

AMGH Editora Ltda.

2011

The McGraw-Hill Companies

Métodos Numéricos para Engenharia
Tradução da Quinta Edição
ISBN 978-85-86804-87-8

A reprodução total ou parcial deste volume por quaisquer formas ou meios, sem o consentimento, por escrito, da editora é ilegal e configura apropriação indevida dos direitos intelectuais e patrimoniais dos autores.

Copyright© 2008 by McGraw-Hill Interamericana do Brasil Ltda.
Todos os direitos reservados.
Av. Brigadeiro Faria Lima , 201, 17 º andar
São Paulo – SP – CEP 05426-100

Copyright © 2008 by McGraw-Hill Interamericana Editores, S.A. de C.V.
Todos os direitos reservados.
Prol. Paseo de la Reforma 1015 Torre A Piso 17, Col. Desarrollo Santa Fé, Delegación Alvaro Obregón
México 01376, D.F., México

Tradução da quinta edição em inglês de *Numerical Methods for Engineers,* publicada pela McGraw-Hill, uma unidade de negócios da McGraw-Hill Companies Inc., 1221 Avenue of the Americas, New York, NY, 10020.
© 2006 by McGraw-Hill Companies, Inc.
Edições publicadas anteriormente sob o título *Numerical Methods for Engineers*: *With Software and Programming Applications*.
© 2002, 1998, 1988, 1985 by McGraw-Hill Companies, Inc.
ISBN: 0-07-291873-X

Editora de Desenvolvimento: *Marileide Gomes*
Supervisora de Produção: *Guacira Simonelli*
Produção Editorial: *Nilcéia Esposito*
Preparação de Texto: *Marcos Soel e Maria Alice da Costa*
Design de Capa: *Ron Bissell*
Ilustração de Capa: © *Jack Novak/ SuperStock*
Diagramação: *ERJ Composição Editorial e Artes Gráficas Ltda.*

MATLAB™ é uma marca registrada de The MathWorks,Inc.

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

C467m Chapra, Steven C.
Métodos numéricos para engenharia [recurso eletrônico] / Steven C. Chapra, Raymond P. Canale ; tradução técnica: Helena Castro. – 5. ed. – Dados eletrônicos. – Porto Alegre : AMGH, 2011.

Publicado também como livro impresso em 2008.
ISBN 978-85-8055-011-5

1. Engenharia – Método numérico. I. Canale, Raymond P. II. Título.

CDU 62:51

Catalogação na publicação: Ana Paula M. Magnus – CRB 10/2052

A McGraw-Hill tem forte compromisso com a qualidade e procura manter laços estreitos com seus leitores. Nosso principal objetivo é oferecer obras de qualidade a preços justos, e um dos caminhos para atingir essa meta é ouvir o que os leitores têm a dizer. Portanto, se você tem dúvidas, críticas ou sugestões, entre em contato conosco — preferencialmente por correio eletrônico (mh_brasil@mcgraw-hill.com) — e nos ajude a aprimorar nosso trabalho. Teremos prazer em conversar com você. Em Portugal use o endereço servico_clientes@mcgraw-hill.com.



Para

Margaret e Gabriel Chapra

Helen e Chester Canale



SUMÁRIO

PREFÁCIO xv

VISÃO GERAL DO LIVRO xviii

SOBRE OS AUTORES xxi

PARTE UM

MODELAGEM, COMPUTADORES, E ANÁLISE DE ERROS 1

PT1.1 Motivação 2
PT1.2 Fundamentos Matemáticos 4
PT1.3 Orientação 6

CAPÍTULO 1
Modelagem Matemática e Resolução de Problemas de Engenharia 9

1.1 Um Modelo Matemático Simples 9
1.2 Leis de Conservação e Engenharia 15
Problemas 18

CAPÍTULO 2
Programação e Software 21

2.1 Pacotes e Programação 21
2.2 Programação Estruturada 22
2.3 Programação Modular 30
2.4 Excel 31
2.5 MATLAB 35
2.6 Outras Linguagens e Bibliotecas 38
Problemas 39

CAPÍTULO 3
Aproximações e Erros de Arredondamento 43

3.1 Algarismos Significativos 41
3.2 Exatidão e Precisão 45
3.3 Definições de Erros 46
3.4 Erros de Arredondamento 49
Problemas 62

CAPÍTULO 4
Erros de Truncamento e Séries de Taylor 63

4.1 A Série de Taylor 63
4.2 Propagação de Erros 77
4.3 Erro Numérico Total 81
4.4 Enganos, Erros de Formulação e Incerteza nos dados 82
Problemas 83

EPÍLOGO: PARTE UM 86
PT1.4 Prós e Contras 86
PT1.5 Relações e Fórmulas Importantes 89
PT1.6 Métodos Avançados e Referências Adicionais 89

PARTE DOIS

RAÍZES DE EQUAÇÕES 92

PT 2.1 Motivação 92
PT 2.2 Fundamentos Matemáticos 94
PT 2.3 Orientação 95

CAPÍTULO 5
Métodos que Isolam a Raiz em Intervalos 98

5.1 Métodos Gráficos 98
5.2 Método da Bisecção 101
5.3 O Método da Falsa Posição 108
5.4 Buscas Incrementais e Determinação de Aproximações Iniciais 113
Problemas 114

CAPÍTULO 6
Métodos Abertos 116

6.1 Iteração de Ponto Fixo Simples 116
6.2 O Método de Newton-Raphson 121
6.3 O Método da Secante 125
6.4 Raízes Múltiplas 130
6.5 Sistemas de Equações Não Lineares 132
Problemas 136

CAPÍTULO 7
Raízes de Polinômios 138

7.1 Polinômios na Engenharia e Ciência 138
7.2 Cálculos Computacionais com Polinômios 140
7.3 Métodos Convencionais 143
7.4 O Método de Miller 144
7.5 O Método de Bairstow 147
7.6 Outros Métodos 152
7.7 Localização de Raízes com Bibliotecas e Pacotes 152
Problemas 160

CAPÍTULO 8
Estudos de Caso: Raízes de Equações 162

8.1 Leis dos Gases Ideais e Não-Ideais (Engenharia Química e Bioquímica) 162
8.2 Escoamento de Canal Aberto (Engenharia Civil/Ambiental) 164
8.3 Projeto de um Circuito Elétrico 168
8.4 Análise de Vibração (Engenharia Mecânica/Aeroespacial) 170
Problemas 175

EPÍLOGO: PARTE DOIS 184
PT2.4 Prós e Contras 184
PT2.5 Relações e Fórmulas Importantes 185
PT2.6 Métodos Avançados e Referências Adicionais 185

PARTE TRÊS

EQUAÇÕES ALGÉBRICAS LINEARES 188

PT3.1 Motivação 188
PT3.2 Fundamentos Matemáticos 190
PT3.3 Orientação 196

CAPÍTULO 9
Eliminação de Gauss 199

9.1 Resolução de um Número Pequeno de Equações 199
9.2 Eliminação de Gauss Ingênua 205
9.3 Armadilhas nos Métodos de Eliminação 211
9.4 Técnicas para Melhorar Soluções 215
9.5 Sistemas Complexos 221
9.6 Sistemas de Equações Não-Lineares 222
9.7 Gauss-Jordan 223
9.8 Resumo 225
Problemas 225

CAPÍTULO 10
Decomposição *LU* e Inversão de Matrizes 228

10.1 Decomposição *LU* 228
10.2 A Matriz Inversa 236
10.3 Análise de Erro e Condicionamento do Sistema 239
Problemas 244

CAPÍTULO 11
Matrizes Especiais e Gauss-Seidel 246

11.1 Matrizes Especiais 246
11.2 Gauss-Seidel 250
11.3 Equações Algébricas Lineares com Bibliotecas e Pacotes 256
Problemas 262

CAPÍTULO 12
Estudos de Caso: Equações Algébricas Lineares 264

12.1 Análise do Estado Estacionário de um Sistema de Reatores (Engenharia Química/ Bioengenharia) 264
12.2 Análise de uma Treliça Estaticamente Determinada (Engenharia Civil/Ambiental) 267
12.3 Correntes e Voltagens em Circuitos Resistores (Engenharia Elétrica) 269
12.4 Sistemas Massa/Mola (Engenharia Mecânica/Aeroespacial) 271
Problemas 273

EPÍLOGO: PARTE TRÊS 282
PT3.4 Prós e Contras 282
PT3.5 Relações e Fórmulas Importantes 283
PT3.6 Métodos Avançados e Referências Adicionais 283

PARTE QUATRO

OTIMIZAÇÃO 285

PT4.1 Motivação 286
PT4.2 Fundamentos Matemáticos 290
PT4.3 Orientação 292

CAPÍTULO 13
Otimização Unidimensional sem Restrições 294

13.1 Busca da Razão Áurea 295
13.2 Interpolação Quadrática 300
13.3 O Método de Newton 302
Problemas 304

CAPÍTULO 14
Otimização Multidimensional sem Restrições 306

14.1 Métodos Diretos 306

14.2 Métodos Gradiente 310
Problemas 321

CAPÍTULO 15
Otimização Restrições 323

15.1 Programação Linear 323
15.2 Otimização com Restrições Não-lineares 332
15.3 Otimização com Pacotes 333
Problemas 343

CAPÍTULO 16
Estudos de Caso: Otimização 345

16.1 Projeto de Menor Custo de um Tanque (Engenharia Química/Bioengenharia) 345
16.2 Tratamento de Baixo Custo para Dejetos Líquidos (Engenharia Civil/Ambiental) 349
16.3 Potência Máxima Transferida para um Circuito (Engenharia Elétrica) 353
16.4 Projeto de *Montain Bike* (Engenharia Mecânica/Aeroespacial) 356
Problemas 358

EPÍLOGO: PARTE QUATRO 363
PT4.4 Prós e Contras 363
PT4.5 Referências Adicionais 364

PARTE CINCO

AJUSTE DE CURVAS 366

PT5.1 Motivação 366
PT5.2 Fundamentos Matemáticos 368
PT5.3 Orientação 376

CAPÍTULO 17
Regressão por Mínimos Quadrados 379

17.1 Regressão Linear 379
17.2 Regressão Polinomial 393
17.3 Regressão Linear Múltipla 396
17.4 Mínimos Quadrados Linear Geral 398
17.5 Regressão Não Linear 403
Problemas 406

CAPÍTULO 18
Interpolação 409

18.1 Polinômios Interpoladores por Diferenças Divididas de Newton 409
18.2 Polinômios Interpoladores de Lagrange 419
18.3 Coeficientes de um Polinômio Interpolador 423
18.4 Interpolação Inversa 424
18.5 Comentários Adicionais 425
18.6 Interpolação por Splines 426
Problemas 436

CAPÍTULO 19
Aproximação de Fourier 438

19.1 Ajuste de Curvas com Funções Senoidais 439
19.2 Série de Fourier Contínua 444
19.3 Domínios de Freqüência e Tempo 447
19.4 Integral e Transformada de Fourier 450
19.5 Transformada de Fourier Discreta (TFD) 452
19.6 Transformada Rápida de Fourier (FFT) 454
19.7 O Espectro de Potência 459

19.8 Ajuste de Curvas com Bibliotecas e Pacotes 460
Problemas 467

CAPÍTULO 20
Estudos de Caso: Ajuste de Curvas 470

20.1 Regressão Linear e Modelos Populacionais (Engenharia Química/Bioengenharia) 470
20.2 Uso de Splines para Estimar Transferência de Calor (Engenharia Civil/Ambiental) 473
20.3 Análise de Fourier (Engenharia Elétrica) 475
20.4 Análise de Dados Experimentais (Engenharia Mecânica/Aeroespacial) 476
Problemas 477

EPÍLOGO: PARTE CINCO 486
PT5.4 Prós e Contras 486
PT5.5 Relações e Fórmulas Importantes 487
PT5.6 Métodos Avançados e Referências Adicionais 489

PARTE SEIS

INTEGRAÇÃO DERIVAÇÃO NUMÉRICAS 491

PT 6.1 Motivação 491
PT 6.2 Fundamentos Matemáticos 499
PT 6.3 Orientação 501

CAPÍTULO 21
Fórmulas de Integração de Newton-Cotes 504

21.1 A Regra do Trapézio 505
21.2 Regras de Simpson 513
21.3 Integração com Segmentos Desiguais 521
21.4 Fórmulas de Integração Abertas 524
21.5 Integrais Múltiplas 524
Problemas 526

CAPÍTULO 22
Integração de Equações 528

22.1 Algoritmos de Newton-Cotes para Equações 528
22.2 Integração de Romberg 530
22.3 Quadratura de Gauss 534
22.4 Integrais Impróprias 540
Problemas 543

CAPÍTULO 23
Derivação Numérica 545

23.1 Fórmulas de Derivação de Alta Precisão 545
23.2 Extrapolação de Richardson 548
23.3 Derivadas de Dados Desigualmente Espaçados 549
23.4 Derivadas e Integrais para Dados com Erros 550
23.5 Integração e Derivação Numéricas com Bibliotecas e Pacotes 551
Problemas 554

CAPÍTULO 24
Estudos de Caso: Integração e Derivação Numéricas 545

24.1 Integração para Determinar a Quantidade Total de Calor (Engenharia Química/Bioengenharia) 545
24.2 Força Efetiva no Mastro de um Veleiro de Corrida (Engenharia Civil/Ambiental) 559
24.3 Corrente Eficaz por Integração Numérica (Engenharia Elétrica) 561

24.4 Integração Numérica para Calcular o Trabalho (Engenharia Mecânica/Aeroespacial) 563
Problemas 566

EPÍLOGO: PARTE SEIS 575
PT6.4 Prós e Contras 575
PT6.5 Relações e Fórmulas Importantes 576
PT6.6 Métodos Avançados e Referências Adicionais 576

PARTE SETE

EQUAÇÕES DIFERENCIAIS ORDINÁRIAS 578

PT7.1 Motivação 578
PT7.2 Fundamentos Matemáticos 582
PT7.3 Orientação 584

CAPÍTULO 25
Métodos Runge-Kutta 587

25.1 Método de Euler 588
25.2 Melhorias no Método de Euler 597
25.3 Métodos de Runge-Kutta 604
25.4 Sistemas de Equações 612
25.5 Métodos de Runge-Kutta Adaptativos 617
Problemas 623

CAPÍTULO 26
Rigidez e Métodos de Passo Mútiplo 625

26.1 Rigidez 625
26.2 Métodos de Passo Múltiplo 628
Problemas 646

CAPÍTULO 27
Problemas de Contorno e de Autovalor 648

27.1 Métodos Gerais para Problemas de Contorno 649
27.2 Problemas de Autovalor 654
27.3 EDOs e Autovalores com Bibliotecas e Pacotes 664
Problemas 670

CAPÍTULO 28
Estudos de Caso: Equações Diferenciais Ordinárias 673

28.1 Usando EDOs para Analisar a Resposta Transiente de um Reator (Engenharia Química/Bioengenharia) 673
28.2 Modelos Predador-Presa e Caos (Engenharia Civil/Ambiental) 788
28.3 Simulando Correntes Transientes para um Circuito Elétrico (Engenharia Elétrica) 682
28.4 O Pêndulo Oscilante (Engenharia Mecânica/Aeroespacial) 686
Problemas 689

EPÍLOGO: PARTE SETE 695
PT7.4 Prós e Contras 695
PT7.5 Relações e Fórmulas Importantes 696
PT7.6 Métodos Avançados e Referências Adicionais 696

PARTE OITO

EQUAÇÕES DIFERENCIAIS PARCIAIS 699

PT8.1 Motivação 699
PT8.2 Orientação 702

CAPÍTULO 29
Diferenças Finitas: Equações Elípticas 705

29.1 A Equação de Laplace 705
29.2 Técnicas de Solução 707
29.3 Condições de Contorno 712
29.4 A Abordagem por Volume de Controle 717
29.5 Software para Resolver Equações Elípticas 720
Problemas 721

CAPÍTULO 30
Diferenças Finitas: Equações Parabólicas 722

30.1 A Equação de Condução do Calor 722
30.2 Métodos Explícitos 723
30.3 Um Método Implícito Simples 726
30.4 O Método de Crank-Nicolson 730
30.5 Equações Parabólicas em Duas Dimensões Espaciais 732
Problemas 735

CAPÍTULO 31
Método dos Elementos Finitos 736

31.1 A Abordagem Geral 737
31.2 Aplicação de Elementos Finitos em uma Dimensão Um 740
31.3 Problemas Bidimensionais 748
31.4 Resolvendo EDPs com Bibliotecas e Pacotes 751
Problemas 756

CAPÍTULO 32
Estudos de Caso: Equações Diferenciais Parciais 758

32.1 Balanço de Massa Unidimensional de um Reator (Engenharia Química/Bioengenharia) 758
32.2 Deflexões em uma Placa (Engenharia Civil/Ambiental) 762
32.3 Problemas de Campo Eletrostáticos Bidimensionais (Engenharia Elétrica) 763
32.4 Solução por Elementos Finitos de uma Série de Molas (Engenharia Mecânica/Aeroespacial) 766
Problemas 769

EPÍLOGO: PARTE OITO 771
PT 8.3 Prós e Contras 771
PT 8.4 Relações e Fórmulas Importantes 771
PT 8.5 Métodos Avançados e Referências Adicionais 772

APÊNDICE A: A SÉRIE DE FOURIER 773

APÊNDICE B: INICIAÇÃO AO MATLAB 774

BIBLIOGRAFIA 781

ÍNDICE REMISSIVO 785



PREFÁCIO

Já se passaram 20 anos desde a publicação da primeira edição deste livro. Nesse período, nossa asserção de que os métodos numéricos e os computadores figurariam mais proeminentemente nos currículos de engenharia — particularmente em seus primeiros anos — se tornou dramaticamente verdadeira. Muitas universidades agora oferecem cursos tanto em introdução à computação quanto de métodos numéricos em seus primeiros três anos. Além disso, muitos de nossos colegas estão integrando problemas orientados para o computador em outros cursos em todos os níveis do currículo. Logo, esta nova edição ainda está fundamentada na premissa básica de que é preciso oferecer aos estudantes de engenharia uma introdução forte e o mais cedo possível aos métodos numéricos. Conseqüentemente, embora tenhamos expandido nossa cobertura nesta nova edição, tentamos manter muitas das características que tornaram a primeira edição acessível a estudantes tanto do início quanto do final da graduação. Elas incluem:

- **Orientação para problemas**. Os estudantes de engenharia aprendem melhor quando motivados por problemas. Isso é particularmente verdadeiro para matemática e computação. Então, adotamos uma abordagem de uma perspectiva de resolução de problemas.
- **Pedagogia Orientada para o Estudante**. Desenvolvemos diversas características para tornar este livro tão amigável ao estudante quanto possível. Elas incluem a organização geral, o uso das introduções e dos epílogos para consolidar os tópicos principais e o uso extensivo de exemplos trabalhados e estudos de casos de todas as áreas da engenharia. Tentamos também manter nossas explicações simples e orientadas para a prática.
- **Abordagem por "Caixas Transparentes".** Embora enfatizemos a resolução de problemas, acreditamos que é autolimitante para os engenheiros abordar os algoritmos numéricos como "caixas pretas". Assim, incluímos teoria suficiente para permitir que os usuários entendam os conceitos básicos por trás dos métodos. Em particular, enfatizamos a teoria relacionada à análise de erros, as limitações desses métodos e os prós e contras entre os métodos.
- **Orientação para Computadores Pessoais**. Inicialmente, quando escrevemos este livro, existia um grande abismo entre o mundo *batch* dos computadores de grande porte e o mundo interativo dos PCs. Hoje em dia, conforme o desempenho dos PCs aumenta, essa distinção está desaparecendo. Dito isso, este livro ainda enfatiza as visualizações e os cálculos iterativos, que são os marcos dos computadores pessoais.
- **Capacitação dos Estudantes**. Obviamente, introduzimos os estudantes aos recursos-padrão, prontos para o uso, de pacotes como o Excel e o software MATLAB. Entretanto, também é mostrado aos estudantes como desenvolver programas simples, bem estruturados, para estender os recursos básicos desses ambientes. Esse conhecimento se transporta para linguagens de programação padrão como Visual Basic, Fortran 90 e C/C++. Acreditamos que a fuga da programação de computadores representa um pouco de uma queda nos currículos de engenharia. O ponto principal é que, enquanto os engenheiros não estiverem satisfeitos em ficar limitados pelas ferramentas, precisarão escrever códigos. Só que agora eles podem ser chamados de "macros" ou "M-files". Este livro é projetado para capacitá-los a fazer isso.

Além desses cinco princípios originais, a principal melhoria da quinta edição é uma grande revisão e expansão dos conjuntos de problemas dos finais dos capítulos. A maioria dos problemas foi modificada, de modo que fornecem soluções numéricas diferentes das edições anteriores. Além disso, diversos novos problemas foram incluídos. Como nas edições anteriores, incluímos tanto problemas matemáticos quanto aplicados de todos os ramos da engenharia. Em todos os casos, nosso propósito é fornecer aos estudantes exercícios para verificarem seu entendimento e para ilustrar como os métodos numéricos podem ajudá-los a resolver problemas mais eficientemente.

Como sempre, nossa principal intenção ao escrever este livro é fornecer aos estudantes uma sólida introdução aos métodos numéricos. Acreditamos que os estudantes motivados que apreciem os métodos numéricos, a computação e a matemática irão, no final, tornar-se melhores engenheiros. Se nosso livro estimular um entusiasmo por esses assuntos, consideraremos nossos esforços um sucesso.

Agradecimentos. Gostaríamos de agradecer a nossos amigos na McGraw-Hill. Em particular, Amanda Green, Suzanne Jeans e Peggy Selle, que forneceram uma atmosfera positiva e de apoio para a criação desta edição. Como sempre, Beatrice Sussman fez um excelente trabalho na edição do manuscrito e Michael Ryder fez importantes contribuições durante a produção do livro. Agradecimentos especiais aos professores Wally Grant, Olga Pierrakos, Amber Phillips, Justin Griffee e Kevin Mace (Virginia Tech) e à professora Theresa Good (Texas A&M), que contribuiu com problemas para nosso livro ao longo dos anos. Como nas edições passadas, David Clough (Universidade do Colorado) e Jerry Stedinger (Universidade de Cornell) generosamente compartilharam seu discernimento e suas sugestões. Sugestões úteis também foram feitas por Bill Philpot (Universidade de Cornell), Jim Guilkey (Universidade de Utah), Dong-Il Seo (Universidade Nacional de Chungnam, Coréia) e Raymundo Cordero e Karim Muci (ITESM, México). A presente edição também se beneficiou das revisões e sugestões fornecidas pelos seguintes colegas:

Ella M. Atkins, University of Maryland
Betty Barr, University of Houston
Florin Bobaru, University of Nebraska–Lincoln
Ken W. Bosworth, Idaho State University
Anthony Cahill, Texas A&M University
Raymond C. Y. Chin, Indiana University–Purdue, Indianapolis
Jason Clark, University of California, Berkeley
John Collings, University of North Dakota
Ayodeji Demuren, Old Dominion University
Cassiano R. E. de Oliveira, Georgia Institute of Technology
Subhadeep Gan, University of Cincinnati
Aaron S. Goldstein, Virginia Polytechnic Institute and State University
Gregory L. Griffin, Louisiana State University
Walter Haisler, Texas A&M University
Don Hardcastle, Baylor University
Scott L. Hendricks, Virginia Polytechnic Institute and State University
David J. Horntrop, New Jersey Institute of Technology
Tribikram Kundu, University of Arizona
Hysuk Lee, Clemson University
Jichun Li, University of Nevada, Las Vegas
Jeffrey S. Marshall, University of Iowa
George Novacky, University of Pittsburgh
Dmitry Pelinovsky, McMaster University
Siva Parameswaran, Texas Technical University
Greg P. Semeraro, Rochester Institute of Technology
Jerry Sergent, Faifield University
Dipendra K. Sinha, San Francisco State University
Scott A. Socolofsky, Texas A&M University

Robert E. Spall, Utah State University
John C. Strikwerda, University of Wisconsin-Madison
Karsten E. Thompson, Louisiana State University
Kumar Vemaganti, University of Cincinnati
Peter Wolfe, University of Maryland
Yale Yurttas, Texas A&M University
Nader Zamani, University of Windsor
Viktoria Zoltay, Tufts University

Devemos enfatizar que, embora tenhamos recebido sugestões úteis das pessoas mencionadas, somos responsáveis por quaisquer imprecisões ou erros que possam ser detectados nesta edição. Por favor, contate Steve Chapra por e-mail se detectar algum erro.

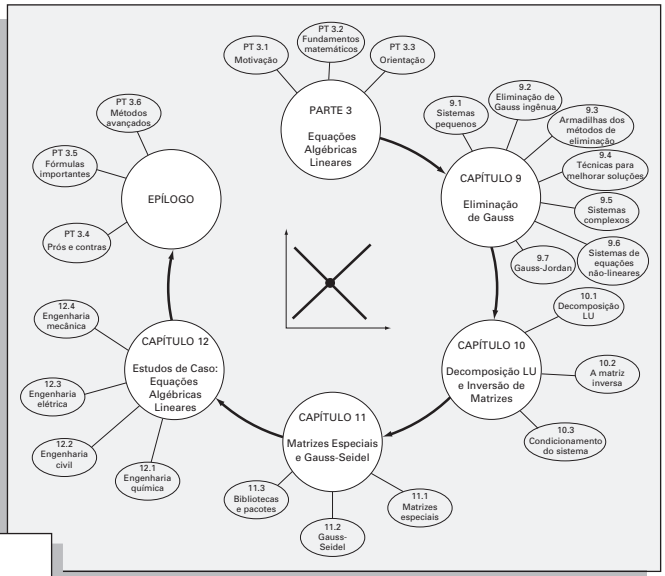
Finalmente, gostaríamos de agradecer a nossa família, amigos e estudantes por sua paciência e seu apoio permanentes. Em particular, Cyntia Chapra e Claire Canale estão sempre lá, fornecendo compreensão, perspectiva e amor.

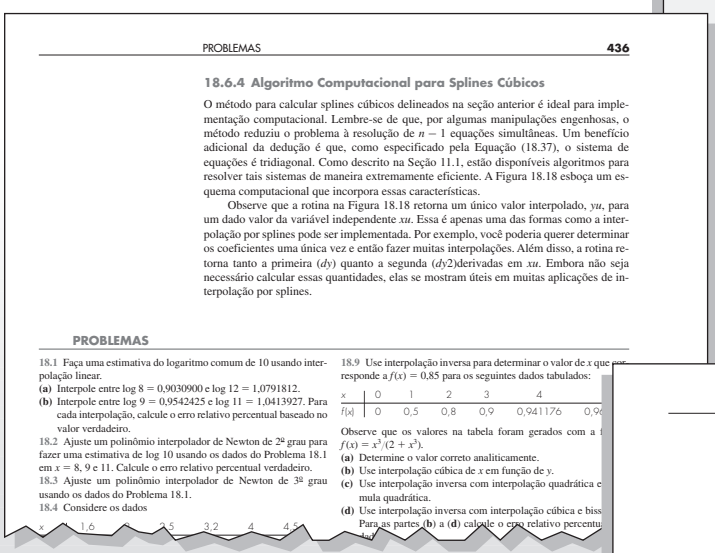
Steven C. Chapra
Medford, Massachusetts
steven.chapra@tufts.edu

Raymond P. Canale
Lake Leelanau, Michigan

VISÃO GERAL DO LIVRO

Para fornecer uma perspectiva dos métodos numéricos, organizamos o texto em partes e apresentamos **informações unificadoras nos elementos Motivação, Fundamentos Matemáticos, Orientação e Epílogo**.





PROBLEMAS 436

18.6.4 Algoritmo Computacional para Splines Cúbicos

O método para calcular splines cúbicos delineados na seção anterior é ideal para implementação computacional. Lembre-se de que, por algumas manipulações engenhosas, o método reduziu o problema à resolução de $n - 1$ equações simultâneas. Um benefício adicional da dedução é que, como especificado pela Equação (18.37), o sistema de equações é tridiagonal. Como descrito na Seção 11.1, estão disponíveis algoritmos para resolver tais sistemas de maneira extremamente eficiente. A Figura 18.18 esboça um esquema computacional que incorpora essas características.

Observe que a rotina na Figura 18.18 retorna um único valor interpolado, *yu*, para um dado valor da variável independente *xu*. Essa é apenas uma das formas como a interpolação por splines pode ser implementada. Por exemplo, você poderia querer determinar os coeficientes uma única vez e então fazer muitas interpolações. Além disso, a rotina retorna tanto a primeira (dy) quanto a segunda ($dy2$)derivadas em *xu*. Embora não seja necessário calcular essas quantidades, elas se mostram úteis em muitas aplicações de interpolação por splines.

PROBLEMAS

18.1 Faça uma estimativa do logaritmo comum de 10 usando interpolação linear.
(a) Interpole entre log 8 = 0,9030900 e log 12 = 1,0791812.
(b) Interpole entre log 9 = 0,9542425 e log 11 = 1,0413927. Para cada interpolação, calcule o erro relativo percentual baseado no valor verdadeiro.

18.2 Ajuste um polinômio interpolador de Newton de 2º grau para fazer uma estimativa de log 10 usando os dados do Problema 18.1 em $x = 8$, 9 e 11. Calcule o erro relativo percentual verdadeiro.

18.3 Ajuste um polinômio interpolador de Newton de 3º grau usando os dados do Problema 18.1.

18.4 Considere os dados

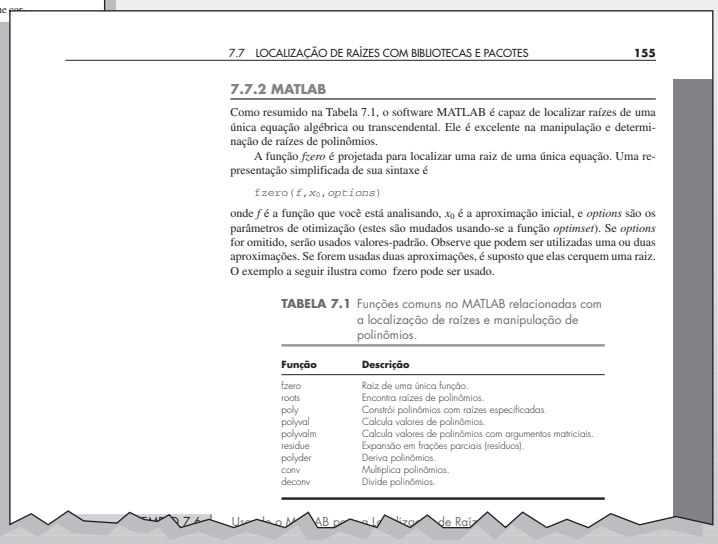
x	1,6	2	2,5	3,2	4	4,5

18.9 Use interpolação inversa para determinar o valor de x que corresponde a $f(x) = 0{,}85$ para os seguintes dados tabulados:

x	0	1	2	3	4
f(x)	0	0,5	0,8	0,9	0,941176

Observe que os valores na tabela foram gerados com a f
$f(x) = x^3/(2 + x^3)$.
(a) Determine o valor correto analiticamente.
(b) Use interpolação cúbica de x em função de y.
(c) Use interpolação inversa com interpolação quadrática e
mula quadrática.
(d) Use interpolação inversa com interpolação cúbica e biss
Para as partes (**b**) a (**d**) calcule o erro relativo percentu

Todos os capítulos contêm **problemas para casa** novos e revisados. Oitenta por cento dos problemas são novos ou revisados. No texto, são incluídos problemas desafiadores tirados de todas as disciplinas de engenharia.



7.7 LOCALIZAÇÃO DE RAÍZES COM BIBLIOTECAS E PACOTES 155

7.7.2 MATLAB

Como resumido na Tabela 7.1, o software MATLAB é capaz de localizar raízes de uma única equação algébrica ou transcendental. Ele é excelente na manipulação e determinação de raízes de polinômios.

A função *fzero* é projetada para localizar uma raiz de uma única equação. Uma representação simplificada de sua sintaxe é

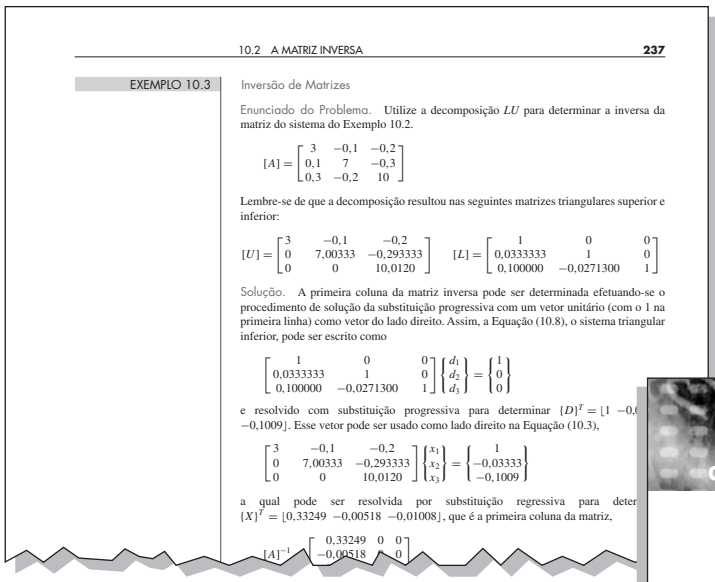
```
fzero(f,x0,options)
```

onde f é a função que você está analisando, x_0 é a aproximação inicial, e *options* são os parâmetros de otimização (estes são mudados usando-se a função *optimset*). Se *options* for omitido, serão usados valores-padrão. Observe que podem ser utilizadas uma ou duas aproximações. Se forem usadas duas aproximações, é suposto que elas cerquem uma raiz. O exemplo a seguir ilustra como fzero pode ser usado.

TABELA 7.1 Funções comuns no MATLAB relacionadas com a localização de raízes e manipulação de polinômios.

Função	Descrição
fzero	Raiz de uma única função.
roots	Encontra raízes de polinômios.
poly	Constrói polinômios com raízes especificadas.
polyval	Calcula valores de polinômios.
polyvalm	Calcula valores de polinômios com argumentos matriciais.
residue	Expansão em frações parciais (resíduos).
polyder	Deriva polinômios.
conv	Multiplica polinômios.
deconv	Divide polinômios.

Seções do texto, bem como problemas para casa, são dedicados à implementação de métodos numéricos com o **Excel da Microsoft e o software MATLAB do The MathWorks, Inc.**



10.2 A MATRIZ INVERSA 237

EXEMPLO 10.3 Inversão de Matrizes

Enunciado do Problema. Utilize a decomposição *LU* para determinar a inversa da matriz do sistema do Exemplo 10.2.

$$[A] = \begin{bmatrix} 3 & -0,1 & -0,2 \\ 0,1 & 7 & -0,3 \\ 0,3 & -0,2 & 10 \end{bmatrix}$$

Lembre-se de que a decomposição resultou nas seguintes matrizes triangulares superior e inferior:

$$[U] = \begin{bmatrix} 3 & -0,1 & -0,2 \\ 0 & 7,00333 & -0,293333 \\ 0 & 0 & 10,0120 \end{bmatrix} \quad [L] = \begin{bmatrix} 1 & 0 & 0 \\ 0,0333333 & 1 & 0 \\ 0,100000 & -0,0271300 & 1 \end{bmatrix}$$

Solução. A primeira coluna da matriz inversa pode ser determinada efetuando-se o procedimento de solução da substituição progressiva com um vetor unitário (com o 1 na primeira linha) como vetor do lado direito. Assim, a Equação (10.8), o sistema triangular inferior, pode ser escrito como

$$\begin{bmatrix} 1 & 0 & 0 \\ 0,0333333 & 1 & 0 \\ 0,100000 & -0,0271300 & 1 \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 0 \\ 0 \end{Bmatrix}$$

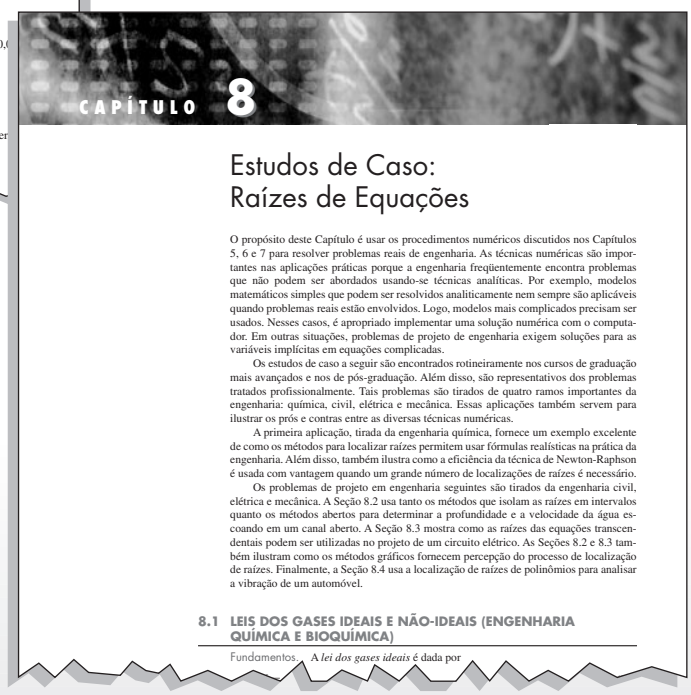
e resolvido com substituição progressiva para determinar $\{D\}^T = \lfloor 1 \ -0,($ $-0,1009 \rfloor$. Esse vetor pode ser usado como lado direito na Equação (10.3),

$$\begin{bmatrix} 3 & -0,1 & -0,2 \\ 0 & 7,00333 & -0,293333 \\ 0 & 0 & 10,0120 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ -0,03333 \\ -0,1009 \end{Bmatrix}$$

a qual pode ser resolvida por substituição regressiva para deter $\{X\}^T = \lfloor 0,33249 \ -0,00518 \ -0,01008 \rfloor$, que é a primeira coluna da matriz,

$$[A]^{-1} \begin{bmatrix} 0,33249 & 0 & 0 \\ -0,00518 & & 0 \end{bmatrix}$$

Nosso texto apresenta diversos exemplos trabalhados para fornecer ao estudante ilustrações passo a passo de como os métodos numéricos são implementados.



CAPÍTULO 8

Estudos de Caso: Raízes de Equações

O propósito deste Capítulo é usar os procedimentos numéricos discutidos nos Capítulos 5, 6 e 7 para resolver problemas reais de engenharia. As técnicas numéricas são importantes nas aplicações práticas porque a engenharia freqüentemente encontra problemas que não podem ser abordados usando-se técnicas analíticas. Por exemplo, modelos matemáticos simples que podem ser resolvidos analiticamente nem sempre são aplicáveis quando problemas reais estão envolvidos. Logo, modelos mais complicados precisam ser usados. Nesses casos, é apropriado implementar uma solução numérica com o computador. Em outras situações, problemas de projeto de engenharia exigem soluções para as variáveis implícitas em equações complicadas.

Os estudos de caso a seguir são encontrados rotineiramente nos cursos de graduação mais avançados e nos de pós-graduação. Além disso, são representativos dos problemas tratados profissionalmente. Tais problemas são tirados de quatro ramos importantes da engenharia: química, civil, elétrica e mecânica. Essas aplicações também servem para ilustrar os prós e contras entre as diversas técnicas numéricas.

A primeira aplicação, tirada da engenharia química, fornece um exemplo excelente de como os métodos para localizar raízes permitem usar fórmulas realísticas na prática da engenharia. Além disso, também ilustra como a eficiência da técnica de Newton-Raphson é usada com vantagem quando um grande número de localizações de raízes é necessário.

Os problemas de projeto em engenharia seguintes são tirados da engenharia civil, elétrica e mecânica. A Seção 8.2 usa tanto os métodos que isolam as raízes em intervalos quanto os métodos abertos para determinar a profundidade e a velocidade da água escoando em um canal aberto. A Seção 8.3 mostra como as raízes das equações transcendentais podem ser utilizadas no projeto de um circuito elétrico. As Seções 8.2 e 8.3 também ilustram como os métodos gráficos fornecem percepção do processo de localização de raízes. Finalmente, a Seção 8.4 usa a localização de raízes de polinômios para analisar a vibração de um automóvel.

8.1 LEIS DOS GASES IDEAIS E NÃO-IDEAIS (ENGENHARIA QUÍMICA E BIOQUÍMICA)

Fundamentos. A *lei dos gases ideais* é dada por

Há 28 **estudos de casos de engenharia** para ajudar os estudantes a ligar os métodos numéricos aos principais campos da engenharia.



Recursos Disponíveis no Site

O Online Learning Center (Centro de Aprendizagem Online) no endereço www.mhhe.com/chapra oferece recursos para o estudante e para o professor. Esses materiais estão disponíveis em inglês e alguns são comerciais, ou seja, você precisa comprá-los. Os professores brasileiros necessitam obter uma senha junto à McGraw-Hill do Brasil para acessarem os recursos on-line. A senha deve ser solicitada por e-mail (divulgacao_brasil@mcgraw-hill.com). Na Europa, a senha deve ser obtida junto à McGraw-Hill de Portugal (servico_clientes@mcgraw-hill.com).

Todos os softwares utilizados nos recursos on-line foram desenvolvidos pela McGraw-Hill dos EUA. A McGraw-Hill Interamericana do Brasil não oferece suporte para esses softwares nem se responsabiliza por qualquer falha que possa ocorrer durante o seu uso. Caso tenha algum problema, acesse o suporte técnico em www.mhhe.com/support *ou ainda o site do produto em* www.cosmosm.com.



SOBRE OS AUTORES

Steve Chapra leciona no Departamento de Engenharia Civil e Ambiental na Universidade de Tufts, onde ocupa a posição de Louis Berger Chair na Computação e Engenharia. Seus outros livros incluem *Surface Water-Quality Modeling and Applied Numerical Methods with MATLAB*.

Dr. Chapra recebeu diplomas de engenharia do Manhattan College e da Universidade de Michigan. Antes de se juntar ao corpo docente da Tufts, trabalhou para a Agência de Proteção Ambiental e para a Administração Nacional Atmosférica e Oceânica e lecionou na Universidade A&M do Texas e na Universidade do Colorado. Seus interesses gerais em pesquisa se concentram na modelagem da qualidade da água superficial e em aplicações computacionais avançadas na engenharia ambiental.

Ele recebeu diversos prêmios por suas contribuições acadêmicas, incluindo a medalha Rudolph Hering (ASCE) de 1993 e o Prêmio de Autor Notável da Meriam-Wiley de 1987 (Sociedade Norte-Americana para Educação em Engenharia). Também foi reconhecido como um professor destacado entre o corpo docente tanto da Universidade A&M do Texas (Prêmio Tenneco de 1986) quanto da Universidade do Colorado (Prêmio Hutchinson de 1992).

Raymond P. Canale é professor emérito da Universidade de Michigan. Durante sua carreira de mais de 20 anos na universidade, e ministrou diversos cursos na área de computadores, métodos numéricos e engenharia ambiental. Também dirigiu programas de pesquisa extensivos na área de modelagem matemática e computacional de ecossistemas aquáticos. Ele foi autor ou co-autor de diversos livros e publicou mais de cem artigos e relatórios científicos. Também projetou e desenvolveu software para computadores pessoais para facilitar a educação em engenharia e a solução de problemas de engenharia. Ele ganhou o Prêmio de Autor Notável da Meriam-Wiley da Sociedade Norte-Americana para Educação em Engenharia por seus livros e software e diversos prêmios por suas publicações técnicas.

O professor Canale está atualmente dedicando suas energias a problemas aplicados, nos quais trabalha com empresas e indústrias de engenharia e agências governamentais como consultor e perito.

PARTE UM



MODELAGEM, COMPUTADORES E ANÁLISE DE ERROS

PT1.1 MOTIVAÇÃO

Os métodos numéricos são técnicas pelas quais os problemas matemáticos são formulados de modo que possam ser resolvidos com operações aritméticas. Embora existam muitos tipos de métodos numéricos, eles têm uma característica em comum: invariavelmente envolvem grande número de cálculos aritméticos tediosos. Não é nada surpreendente que, com o desenvolvimento de computadores digitais rápidos e eficientes, o papel dos métodos numéricos na resolução de problemas de engenharia aumentou dramaticamente nos últimos anos.

PT1.1.1 Métodos Não-computacionais

Além de fornecer uma capacidade computacional aumentada, a disponibilidade muito difundida dos computadores (especialmente dos computadores pessoais) e a parceria com métodos numéricos vêm tendo uma influência significativa na resolução moderna de problemas de engenharia. Na era antes do computador, em geral havia três formas diferentes pelas quais os engenheiros abordavam a solução de problemas:

1. As soluções eram deduzidas para alguns problemas usando-se métodos analíticos ou exatos. Essas soluções eram freqüentemente úteis e forneciam uma visão excelente do comportamento de alguns sistemas. Entretanto, as soluções analíticas podem ser deduzidas apenas para uma classe limitada de problemas, que incluem os que podem ser aproximados por modelos lineares e aqueles que têm uma geometria simples e dimensão baixa. Conseqüentemente, as soluções analíticas têm valor prático limitado porque, em sua maioria, os problemas reais são não-lineares e envolvem formas e processos complexos.
2. As soluções gráficas eram usadas para caracterizar o comportamento dos sistemas; em geral, elas tinham a forma de gráficos ou nomógrafos. Embora as soluções gráficas freqüentemente possam ser usadas para resolver problemas complexos, os resultados não são muito precisos. Além disso, soluções gráficas (sem a ajuda de computadores) são extremamente tediosas e inconvenientes para implementar. Finalmente, as técnicas gráficas, com freqüência, são limitadas a problemas que podem ser descritos usando-se três dimensões ou menos.
3. Calculadoras e regras de cálculos eram usadas para implementar os métodos numéricos manualmente. Embora, em teoria, tais abordagens fossem perfeitamente adequadas para resolver problemas complexos, na realidade, eram encontradas muitas dificuldades. Os cálculos manuais são lentos e tediosos. Além disso, resultados consistentes são esquivos por causa de enganos simples que acontecem quando tarefas manuais numerosas são realizadas.

Na era pré-computador, uma quantidade significativa de energia era gasta na técnica de resolução propriamente dita, em vez de na definição e interpretação do problema (Figura PT1.1*a*). Essa situação lamentável existia porque muito tempo e trabalho pesado eram necessários para obter respostas numéricas usando-se as técnicas pré-computador.

Hoje em dia, os computadores e os métodos numéricos fornecem uma alternativa para tais cálculos complicados. Usando o poder do computador para obter diretamente soluções, é possível abordar esses cálculos sem o recurso de hipóteses simplificadoras ou

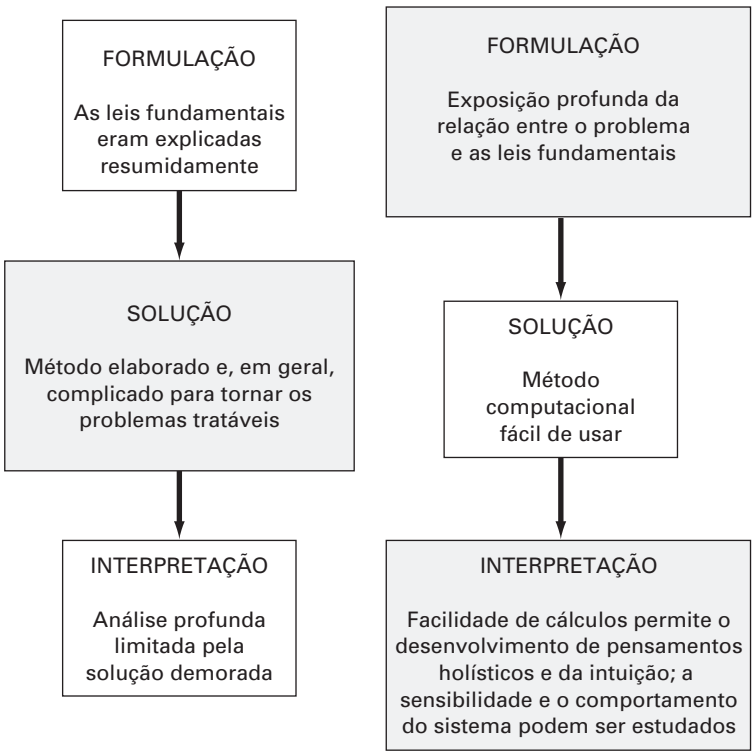


FIGURA PT1.1
Há três fases de resolução de problemas de engenharia (*a*) na era pré-computador e (*b*) na era do computador. O tamanho das caixas indica o nível de ênfase direcionada a cada fase. Os computadores facilitaram a implementação da técnica de resolução e, portanto, permitiram mais ênfase nos aspectos criativos da formulação do problema e na interpretação dos resultados.

técnicas para economizar tempo. Embora as soluções analíticas ainda sejam extremamente valiosas tanto para soluções dos problemas quanto para fornecer uma visão geral, os métodos numéricos representam alternativas que aumentam enormemente os recursos para confrontar e resolver problemas. Como resultado, mais tempo fica disponível para ser usado em habilidades criativas. Portanto, mais ênfase pode ser posta na formulação do problema e na interpretação da solução e na incorporação do sistema total, ou percepção "holística" (Figura PT1.1*b*).

PT1.1.2 MÉTODOS NUMÉRICOS E PRÁTICA DA ENGENHARIA

Desde o fim da década de 1940, a vasta disponibilidade dos computadores digitais levou a uma verdadeira explosão no uso e desenvolvimento dos métodos numéricos. Inicialmente, esse crescimento foi um pouco limitado pelo custo do acesso a computadores de grande porte — conseqüentemente, muitos engenheiros continuaram a usar a abordagem analítica simples em uma parte significativa do seu trabalho. É claro que a recente evolução dos computadores pessoais baratos forneceu pronto acesso a recursos computacionais poderosos. Existem diversas razões adicionais pelas quais se deve estudar métodos numéricos:

1. Os métodos numéricos são ferramentas extremamente poderosas na resolução de problemas. Eles são capazes de lidar com um grande número de equações, não-linearidades e geometrias complicadas que não são incomuns na prática da engenharia e, em geral, são impossíveis de resolver analiticamente. Dessa forma, eles aumentam enormemente a capacidade de resolver problemas.
2. Durante a carreira, o profissional da engenharia freqüentemente terá ocasião de usar pacotes comercialmente disponíveis, ou programas de computador "enlatados" que envolvem métodos numéricos. O uso inteligente desses programas depende, com freqüência, do conhecimento da teoria básica fundamental dos métodos.

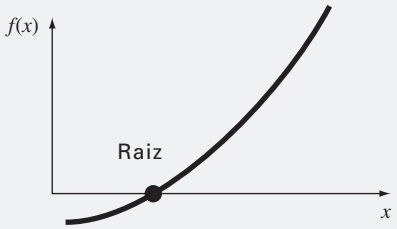
3. Muitos problemas não podem ser abordados utilizando programas enlatados. Se o engenheiro estiver familiarizado com métodos numéricos e souber programar o computador, poderá projetar seu próprio programa para resolver problemas sem ter de comprar ou contratar softwares caros.
4. Os métodos numéricos são um veículo eficiente para o aprendizado do uso de computadores. É bem conhecido que uma forma eficiente de aprender a programar é realmente escrever um programa de computador. Como os métodos numéricos são, na maior parte, projetados para implementação em computadores, eles se mostram ideais para esse propósito. Além disso, são especialmente adequados para ilustrar o poder e as limitações dos computadores. Quando se implementam com sucesso métodos numéricos em um computador e, então, eles são aplicados para resolver problemas intratáveis de outra forma, tem-se acesso a uma demonstração dramática de como os computadores podem ajudar o desenvolvimento profissional. Ao mesmo tempo, aprende-se a identificar e a controlar os erros das aproximações, que são parte essencial de cálculos numéricos em grande escala.
5. Os métodos numéricos fornecem um veículo para o profissional reforçar seu entendimento da matemática. Como uma função dos métodos numéricos é reduzir a matemática mais avançada a operações aritméticas básicas, eles chegam aos detalhes práticos de alguns tópicos que, de outra forma, seriam obscuros. Aprimoramento do entendimento e da percepção pode resultar dessa perspectiva alternativa.

PT1.2 FUNDAMENTOS MATEMÁTICOS

Todas as partes deste livro usam alguns fundamentos matemáticos. Conseqüentemente, o material introdutório para cada parte inclui uma seção, como a presente seção, sobre fundamentos matemáticos. Como a Parte Um propriamente dita é devotada à revisão de material em matemática e computadores, esta seção não envolve uma revisão de um tópico matemático específico. Em vez disso, aproveita-se a oportunidade para introduzir o tipo de assuntos da matemática cobertos neste livro. Como resumido na Figura PT1.2, esses assuntos são:

1. *Raízes de Equações* (Figura PT1.2*a*). Esses problemas dizem respeito ao valor de uma variável ou a um parâmetro que satisfaz uma única equação não-linear. Eles são especialmente importantes no contexto de projetos de engenharia, em que freqüentemente é impossível resolver explicitamente as equações para os parâmetros do projeto.
2. *Sistemas de Equações Algébricas Lineares* (Figura PT1.2*b*). Esses problemas são similares aos das raízes de equações, no sentido de que dizem respeito aos valores que satisfazem tais equações. Entretanto, em contraste a satisfazer uma única equação, é procurado um conjunto de valores que satisfaça simultaneamente um conjunto de equações algébricas lineares. Essas equações aparecem em uma grande variedade de contextos de problemas e em todas as disciplinas da engenharia. Em particular, elas se originam na modelagem matemática de grandes sistemas de elementos interconectados, como estruturas, circuitos elétricos e redes de fluidos. Entretanto, também são encontradas em outras áreas dos métodos numéricos, como ajustes de curva e equações diferenciais.
3. *Otimização* (Figura PT1.2*c*). Esses problemas envolvem a determinação de um valor ou valores de uma variável independente que corresponde ao "melhor" valor ou ao valor ótimo de uma função. Portanto, como na Figura PT1.2*c*, a otimização envolve a identificação de máximos e mínimos. Tais problemas ocorrem rotineiramente no contexto de projetos de engenharia. Eles também aparecem em uma variedade de outros métodos numéricos. A otimização sem restrições é tratada tanto em uma variável quanto em muitas variáveis. Também é descrita a otimização restrita, com ênfase particular na programação linear.

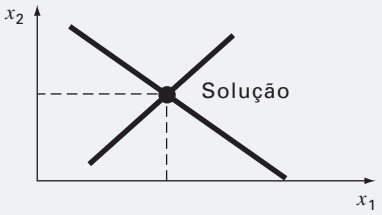
(*a*) *Parte 2:* Raízes de equações
Resolva $f(x) = 0$ determinando x.



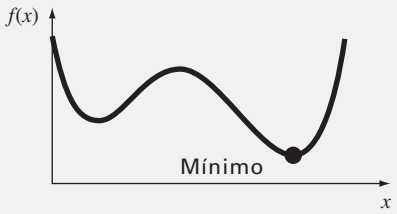
(*b*) *Parte 3:* Equações algébricas lineares
Dados os a's e os c's, resolva

$$a_{11}x_1 + a_{12}x_2 = c_1$$
$$a_{21}x_1 + a_{22}x_2 = c_2$$

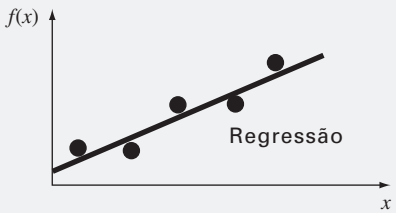
para determinar os x's.

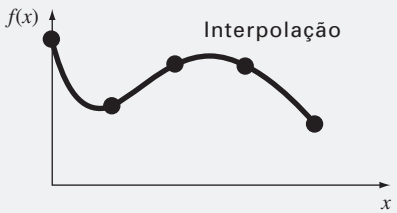


(*c*) *Parte 4:* Otimização
Determine x que dê o valor ótimo de $f(x)$.



(*d*) *Parte 5:* Ajuste de curvas

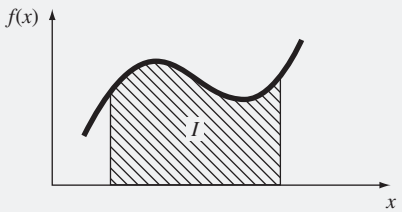




(*e*) *Parte 6:* Integração

$$I = \int_a^b f(x)\,dx$$

Encontre a área sob a curva.

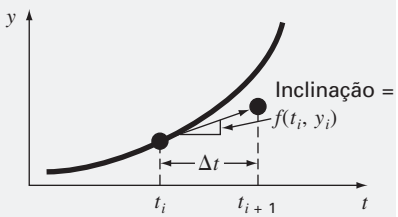


(*f*) *Parte 7:* Equações diferenciais ordinárias
Dada

$$\frac{dy}{dt} \simeq \frac{\Delta y}{\Delta t} = f(t, y)$$

resolva, encontrando y como função de t.

$$y_{i+1} = y_i + f(t_i, y_i)\,\Delta t$$



(*g*) *Parte 8:* Equações diferenciais parciais
Dada

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y)$$

resolva, encontrando u como função de x e y

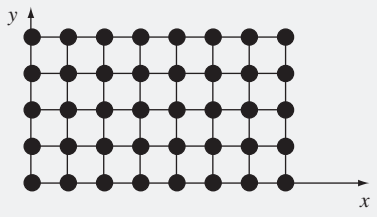


FIGURA PT1.2
Resumo dos métodos numéricos cobertos neste livro.

4. *Ajuste de Curvas* (Figura PT1.2*d*). Freqüentemente é necessário ajustar curvas a conjuntos de dados. As técnicas desenvolvidas para esse propósito podem ser divididas em duas categorias gerais: regressão e interpolação. A regressão é empregada quando existe um grau significativo de erro associado aos dados. Os resultados experimentais são geralmente desse tipo. Para tais situações, a estratégia é desenvolver uma única curva que represente a tendência geral dos dados, sem necessariamente coincidir com nenhum ponto individual. Em contraste, a interpolação é usada quando o objetivo é determinar valores intermediários entre dados relativamente livres de erros — usualmente, é o caso para informações tabuladas. Para tais situações, a estratégia é ajustar a curva diretamente pelos pontos dados e usar essa curva para prever os valores intermediários.
5. *Integração* (Figura PT1.2*e*). Como descrito, uma interpretação física de integração numérica é a determinação da área sob a curva. A integração tem muitas aplicações na prática da engenharia, variando da determinação de centróides de objetos de forma estranha a cálculos de quantidades totais baseados em conjuntos de medidas discretas. Além disso, as fórmulas de integração numérica desempenham um papel importante na solução de equações diferenciais.
6. *Equações Diferenciais Ordinárias* (Figura PT1.2*f*). As equações diferenciais ordinárias têm um grande significado na prática da engenharia, e isso ocorre porque muitas leis físicas são descritas em termos da taxa de variação de uma quantidade em vez do valor da quantidade propriamente dita. Os exemplos variam de modelos de previsão populacional (taxa de variação da população) à aceleração de um corpo em queda livre (taxa de variação da velocidade). Dois tipos de problemas são tratados: problemas de valor inicial e valor de contorno. Além disso, o cálculo de autovalores também é coberto.
7. *Equações Diferenciais Parciais* (Figura PT1.2*g*). As equações diferenciais parciais são usadas para caracterizar sistemas de engenharia nos quais o comportamento de uma quantidade física é determinado pela sua taxa de variação em relação a duas ou mais variáveis independentes. Os exemplos incluem a distribuição estacionária de temperaturas em uma chapa aquecida (duas dimensões espaciais) ou a temperatura dependendo do tempo de uma haste aquecida (tempo e uma dimensão espacial). Duas abordagens fundamentalmente diferentes são empregadas para resolver equações diferenciais parciais numericamente. Neste texto, serão enfatizados os métodos de diferença finita, que aproximam a solução de maneira pontual (Figura PT1.2*g*). Entretanto, também será apresentada uma introdução aos métodos de elementos finitos, que usam uma abordagem por partes.

PT1.3 ORIENTAÇÃO

Alguma orientação pode ser útil antes de prosseguir com a introdução aos métodos numéricos. O que segue deve ser visto como uma visão geral do material na Parte Um. Além disso, alguns objetivos foram incluídos para focalizar os esforços do aluno ao estudar este material.

PT1.3.1 Escopo e Visão Geral

A Figura PT1.3 é uma representação esquemática do material na Parte Um. Esse diagrama foi projetado para fornecer uma visão global desta parte do livro. O sentido de "perspectiva geral" é crítico no desenvolvimento da intuição nos métodos numéricos. Quando se lê um texto, é fácil perder-se nos detalhes técnicos. Sempre que o estudante sentir que está perdendo a perspectiva geral, deve consultar a Figura PT1.3 para se reorientar. Todas as partes deste livro incluem uma figura similar.

A Figura PT1.3 também serve como uma breve apresentação prévia do material coberto na Parte Um. O Capítulo 1 foi projetado para orientar o estudante nos métodos numéricos e para fornecer motivação, mostrando como essas técnicas podem ser usadas no processo de modelagem da engenharia. O Capítulo 2 é uma introdução e revisão de aspectos dos métodos numéricos relacionados com o computador e sugere o nível de ha-



FIGURA PT1.3
Esquema da organização do material da Parte Um: Modelagem, Computadores e Análise de Erros.

bilidades computacionais que é necessário adquirir para aplicar de forma eficiente as informações que vêm a seguir. Os Capítulos 3 e 4 dizem respeito ao importante tópico da análise de erros, que precisa ser entendido para o uso eficiente dos métodos numéricos. Além disso, é incluído um epílogo, no qual são introduzidos os prós e contras, que têm um grande significado para a implementação efetiva dos métodos numéricos.

PT1.3.2 Metas e Objetivos

Objetivos de Estudo. Após completar a Parte Um, o estudante deverá estar adequadamente preparado para embarcar em seus estudos sobre os métodos numéricos. Em geral, deve ter adquirido um entendimento fundamental sobre a importância dos computadores e sobre o papel das aproximações e dos erros na implementação e no desenvolvimento dos métodos numéricos. Além dessas metas gerais, deve ter dominado cada um dos objetivos de estudo específicos listados na Tabela PT1.1.

TABELA PT1.1 Objetivos de estudo específicos para a Parte Um.

1. Reconhecer a diferença entre soluções analítica e numérica.
2. Entender como as leis de conservação são empregadas para desenvolver modelos matemáticos do sistema físico.
3. Definir projeto por refinamento e modular.
4. Definir as regras que fundamentam a programação estruturada.
5. Ser capaz de compor programas estruturados e modulares em uma linguagem de computação de alto nível.
6. Saber como traduzir fluxogramas estruturados e pseudocódigos para códigos em uma linguagem de alto nível.
7. Começar a se familiarizar com todos os pacotes de software que serão usados junto com este texto.
8. Reconhecer a distinção entre erros de truncamento e de arredondamento.
9. Entender os conceitos de algarismos significativos, acurácia e precisão.
10. Reconhecer a diferença entre erro relativo verdadeiro ε_t, erro relativo aproximado ε_a, e erro aceitável ε_s, e entender como ε_a e ε_s são usados para finalizar um cálculo iterativo.
11. Entender como os números são representados em computadores digitais e como essa representação induz a erros de arredondamento. Em particular, saber a diferença entre precisão simples e estendida.
12. Reconhecer como a aritmética do computador pode introduzir e ampliar erros de arredondamento nos cálculos. Em particular, analisar o problema do cancelamento na subtração.
13. Entender como a série de Taylor e seu resto são empregados para representar funções contínuas.
14. Saber a relação entre as diferenças divididas finitas e as derivadas.
15. Ser capaz de analisar como os erros se propagam por meio das relações funcionais.
16. Estar familiarizado com os conceitos de estabilidade e condicionamento.
17. Familiarizar-se com os prós e contras delineados no Epílogo da Parte Um.

Objetivos Computacionais. Quando completar a Parte Um, o estudante deve ter adquirido habilidades computacionais suficientes para desenvolver seu próprio software para os métodos numéricos neste texto. Deve ser capaz de desenvolver programas de computador bem estruturados e confiáveis com base em pseudocódigo, fluxogramas e outras formas de algoritmos. Também deve ter desenvolvido a capacidade de documentar seus programas de forma que eles possam ser usados efetivamente pelos usuários. Finalmente, além dos seus próprios programas, o aluno pode usar pacotes de softwares juntamente com este livro. Os pacotes como Excel ou The MathWorks, Inc. MATLAB® são exemplos de tais softwares. O estudante deve estar familiarizado com esses pacotes, de forma que possa usá-los confortavelmente para resolver problemas numéricos mais adiante neste texto.

CAPÍTULO 1

Modelagem Matemática e Resolução de Problemas de Engenharia

O conhecimento e o entendimento são pré-requisitos para a implementação efetiva de qualquer ferramenta. Não importa quão incrível seja sua caixa de ferramentas, você terá dificuldades para consertar um carro se não entender como ele funciona.

Isso é particularmente verdade quando se usam computadores para resolver problemas de engenharia. Embora tenham uma utilidade potencial imensa, os computadores são praticamente inúteis sem um entendimento fundamental de como os sistemas da engenharia funcionam.

Esse conhecimento é adquirido inicialmente de forma empírica — isto é, por observação e experiência. Entretanto, embora tal informação adquirida empiricamente seja essencial, ela é apenas metade da história. Durante anos e anos de observação e experiência, os engenheiros e cientistas notaram que certos aspectos dos seus estudos empíricos ocorrem repetidamente. Tal comportamento geral pode então ser expresso como leis fundamentais que essencialmente representam o conhecimento acumulado da experiência passada. Portanto, a resolução da maioria dos problemas de engenharia usa uma abordagem com as duas vertentes, a do empirismo e a da análise teórica (Figura 1.1).

Deve ser enfatizado que as duas vertentes são intimamente ligadas. Conforme novas medidas são feitas, as generalizações podem ser modificadas, ou novas generalizações desenvolvidas. Analogamente, as generalizações podem ter uma forte influência nas experiências e observações. Em particular, as generalizações podem servir como princípios organizatórios empregados para resumir resultados de observações e experiências em uma estrutura coerente e abrangente, a partir das quais são tiradas conclusões. Da perspectiva de resolução de problemas de engenharia, tal estrutura é mais útil quando expressa na forma de um modelo matemático.

O objetivo primário deste capítulo é introduzi-lo à modelagem matemática e ao seu papel na resolução de problemas de engenharia. Vamos também ilustrar como os métodos numéricos figuram no processo.

1.1 UM MODELO MATEMÁTICO SIMPLES

Um *modelo matemático* pode ser definido, de forma geral, como uma formulação ou equação que expressa as características essenciais de um sistema ou processo físico em termos matemáticos. Em um sentido muito geral, ele pode ser representado como uma relação funcional da forma

$$\begin{matrix}\text{Variável}\\ \text{dependente}\end{matrix} = f\left(\begin{matrix}\text{Variáveis}\\ \text{independentes}\end{matrix}, \text{ parâmetros, } \begin{matrix}\text{termos}\\ \text{forçantes}\end{matrix}\right) \tag{1.1}$$

onde a *variável dependente* é uma característica que usualmente reflete o comportamento ou estado do sistema; as *variáveis independentes* usualmente são dimensões, como tempo e espaço, ao longo das quais o comportamento do sistema está sendo determinado; os parâmetros refletem propriedades ou composição do sistema; os *termos forçantes* são as influências externas agindo sobre o sistema.

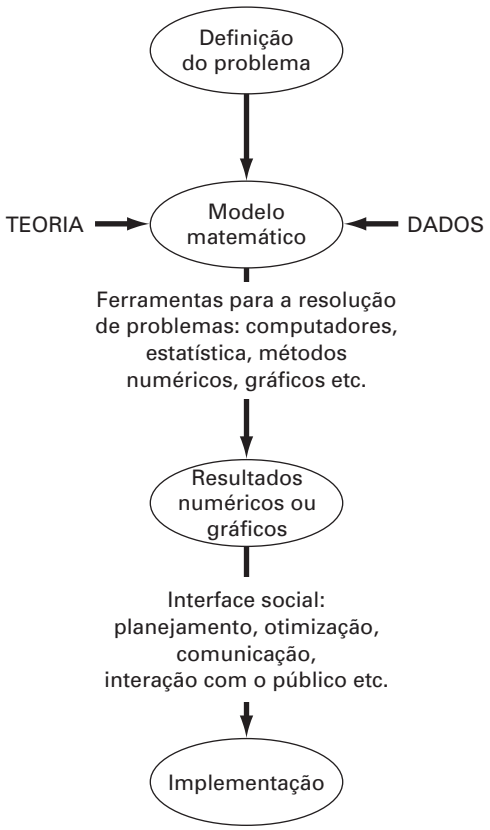


FIGURA 1.1
O processo de resolver problemas de engenharia.

A expressão matemática real da Equação (1.1) pode variar de uma simples relação algébrica a um conjunto grande e complicado de equações diferenciais. Por exemplo, com base em suas observações, Newton formulou sua segunda lei do movimento, que afirma que a taxa de variação no tempo do momento de um corpo é igual à força resultante agindo nele. A expressão matemática, ou modelo, da segunda lei é a equação bem conhecida

$$F = ma \tag{1.2}$$

onde F é a força resultante agindo no corpo (N, ou kg m/s^2), m = massa do objeto (kg) e a é a sua aceleração (m/s^2).

A segunda lei pode ser reescrita na forma da Equação (1.1) simplesmente dividindo ambos os lados por m para obter

$$a = \frac{F}{m} \tag{1.3}$$

onde a é a a variável dependente refletindo o comportamento do sistema, F é o termo forçante e m é um parâmetro representando uma propriedade do sistema. Observe que, para esse caso simples, não há nenhuma variável independente, porque não estamos prevendo como a aceleração varia no tempo ou no espaço.

A Equação (1.3) tem diversas características que são típicas de modelos matemáticos do mundo físico:

1. Ela descreve um processo ou sistema natural em termos matemáticos.
2. Ela representa uma idealização e simplificação da realidade. Isto é, o modelo ignora detalhes desprezíveis do processo natural e se concentra nas suas manifestações essenciais. Portanto, a segunda lei não inclui os efeitos da relatividade, que são de importância mínima quando aplicados a objetos e forças que interagem sobre a ou perto da superfície da Terra, com velocidades e em escalas visíveis aos humanos.

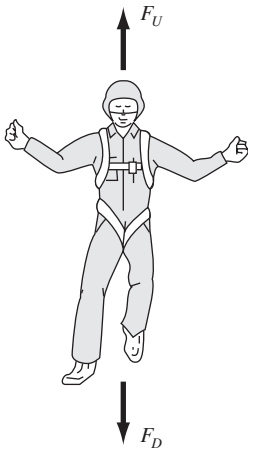


FIGURA 1.2
Diagrama esquemático das forças agindo em um pára-quedista em queda livre. F_D é a força devida à gravidade, para baixo. F_U é a força devida à resistência do ar, para cima.

3. Finalmente, ela produz resultados que podem ser reproduzidos e, conseqüentemente, ser usados para propósitos de previsão. Por exemplo, se a força sobre um objeto e a massa de um objeto são conhecidas, a Equação (1.3) pode ser usada para calcular a aceleração.

Por causa de sua forma algébrica simples, a solução da Equação (1.2) é facilmente obtida. Entretanto, outros modelos matemáticos de fenômenos físicos podem ser muito mais complexos e, ou não podem ser resolvidos exatamente, ou exigem técnicas matemáticas mais sofisticadas que a álgebra simples para sua solução. Para ilustrar um modelo mais complexo deste tipo, a segunda lei de Newton pode ser usada para determinar a velocidade terminal de um corpo em queda livre, perto da superfície da Terra. Nosso corpo em queda livre será um pára-quedista (Figura 1.2). Um modelo para esse caso pode ser deduzido expressando a aceleração como taxa de variação no tempo da velocidade (dv/dt) e substituindo-a na Equação (1.3) para obter

$$\frac{dv}{dt} = \frac{F}{m} \tag{1.4}$$

onde v é a velocidade (m/s) e t é o tempo (s). Portanto, a massa multiplicada pela taxa de variação da velocidade é igual à força resultante agindo no corpo. Se a força resultante for positiva, o objeto irá acelerar. Se for negativa, o objeto vai desacelerar. Se a força resultante for nula, a velocidade do objeto permanecerá em um nível constante.

A seguir, vamos expressar a força resultante em termos das variáveis e parâmetros mensuráveis. Para um corpo em queda livre na vizinhança da Terra (Figura 1.2), a força resultante é composta de duas forças opostas: a força gravitacional, para baixo, F_D e a força da resistência do ar, para cima, F_U:

$$F = F_D + F_U \tag{1.5}$$

Se associarmos um sinal positivo à força para baixo, a segunda lei pode ser usada para escrever a força devida à gravidade como

$$F_D = mg \tag{1.6}$$

onde g é a constante gravitacional, ou a aceleração devida à gravidade, que é aproximadamente igual a 9,8 m/s^2.

A resistência do ar pode ser formulada de diversas maneiras. Uma abordagem simples é assumir que ela é linearmente proporcional à velocidade[1] e age no sentido para cima, como em

$$F_U = -cv \tag{1.7}$$

onde c é uma constante de proporcionalidade chamada de *coeficiente de arrasto* (kg/s). Portanto, quanto maior a velocidade de queda, maior a força para cima devida à resistência do ar. O parâmetro c representa as propriedades de objetos em queda livre, como a forma ou a aspereza da superfície, que afetam a resistência do ar. No caso presente, c poderia ser uma função do tipo de macacão ou da orientação usada pelo pára-quedista durante a queda livre.

A força resultante é a diferença entre a força para baixo e a força para cima. Portanto, as Equações (1.4) até (1.7) podem ser combinadas para fornecer

$$\frac{dv}{dt} = \frac{mg - cv}{m} \tag{1.8}$$

ou, simplificando o lado direito,

$$\frac{dv}{dt} = g - \frac{c}{m}v \tag{1.9}$$

[1] Na realidade, a relação é realmente não-linear e poderia ser mais bem representada por uma relação do tipo potência, como $FU = -cv^2$. Vamos explorar como tais não-linearidades afetam o modelo em um problema no final deste capítulo.

A Equação (1.9) é um modelo que relaciona a aceleração do objeto em queda às forças agindo nele. Ela é uma *equação diferencial* porque é escrita em termos da taxa de variação diferencial (dv/dt) da variável que estamos interessados em prever. Entretanto, em contraste com a solução da segunda lei de Newton na Equação (1.3), a solução exata da Equação (1.9) para a velocidade de um pára-quedista em queda livre não pode ser obtida usando manipulação algébrica simples. Em vez disso, técnicas mais avançadas como aquelas do cálculo devem ser aplicadas para se obter uma solução exata ou analítica. Por exemplo, se o pára-quedista estiver inicialmente em repouso ($v = 0$ em $t = 0$), o cálculo pode ser usado para resolver a Equação (1.9), fornecendo

$$v(t) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{1.10}$$

Observe que a Equação (1.10) está escrita na forma geral da Equação (1.1), onde $v(t)$ é a variável independente, t é a variável independente, c e m são parâmetros, e g é o termo forçante.

EXEMPLO 1.1 Solução Analítica para o Problema do Pára-quedista em Queda Livre

Enunciado do Problema. Um pára-quedista de massa 68,1 kg pula de um balão de ar quente parado. Use a Equação (1.10) para calcular a velocidade anterior à abertura do pára-quedas. O coeficiente de arrasto é igual a 12,5 kg/s.

Solução. Inserindo os parâmetros na Equação (1.10), obtemos

$$v(t) = \frac{9{,}8(68{,}1)}{12{,}5}\left(1 - e^{-(12{,}5/68{,}1)t}\right) = 53{,}39\left(1 - e^{-0{,}18355t}\right)$$

o que pode ser usado para calcular

t, s	*v*, m/s
0	0,00
2	16,40
4	27,77
6	35,64
8	41,10
10	44,87
12	47,49
∞	53,39

De acordo com o modelo, o pára-quedista acelera rapidamente (Figura 1.3). Uma velocidade de 44,87 m/s (100,4 mi/h) é atingida após 10 s. Observe também que, após um tempo suficientemente longo, é atingida uma velocidade constante, chamada de *velocidade terminal*, de 53,39 m/s (119,4 mi/h). Essa velocidade é constante porque, eventualmente, a força da gravidade estará em equilíbrio com a resistência do ar. Portanto, a força resultante é nula e a aceleração deixa de existir.

A Equação (1.10) é chamada uma *solução analítica* ou *exata* porque ela satisfaz exatamente a equação diferencial original. Infelizmente, existem muitos modelos matemáticos que não podem ser resolvidos exatamente. Em muitos desses casos, a única alternativa é desenvolver uma solução numérica que aproxime a solução exata.

Como mencionado previamente, os *métodos numéricos* são aqueles nos quais os problemas matemáticos são reformulados de forma que possam ser resolvidos por operações aritméticas. Isso pode ser ilustrado para a segunda lei de Newton, observando que a taxa de variação no tempo da velocidade pode ser aproximada por (Figura 1.4):

$$\frac{dv}{dt} \cong \frac{\Delta v}{\Delta t} = \frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i} \tag{1.11}$$

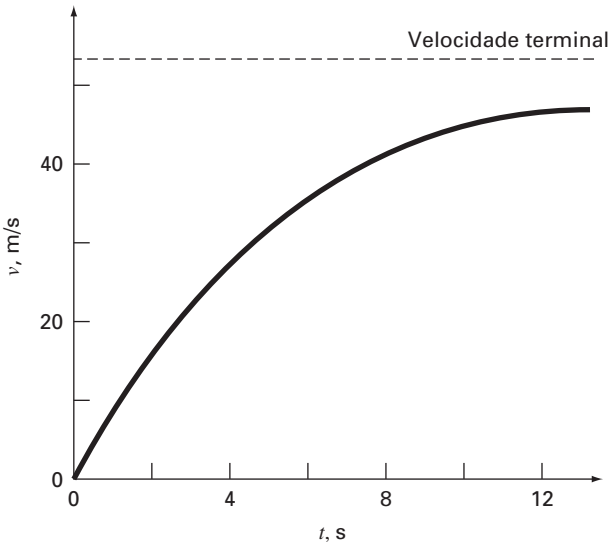


FIGURA 1.3
A solução analítica do problema do pára-quedista em queda livre, como calculada no Exemplo 1.1. A velocidade aumenta com o tempo e se aproxima assintoticamente da velocidade terminal.

onde Δv e Δt são diferenças na velocidade e no tempo, respectivamente, calculados sobre intervalos finitos, $v(t_i)$ é velocidade em um instante inicial t_i, e $v(t_{i+1})$ é velocidade em um instante posterior t_{i+1}. Observe que $dv/dt \cong \Delta v/\Delta t$ é aproximado porque Δt é finito. Lembre-se, do cálculo, que

$$\frac{dv}{dt} = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t}$$

A Equação (1.11) representa o processo reverso.

A Equação (1.11) é chamada de aproximação por *diferença dividida finita* da derivada no instante t_i. Ela pode ser substituída na Equação (1.9) para fornecer

$$\frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i} = g - \frac{c}{m} v(t_i)$$

Esta equação pode ser rearranjada para fornecer

$$v(t_{i+1}) = v(t_i) + \left[g - \frac{c}{m} v(t_i) \right](t_{i+1} - t_i) \tag{1.12}$$

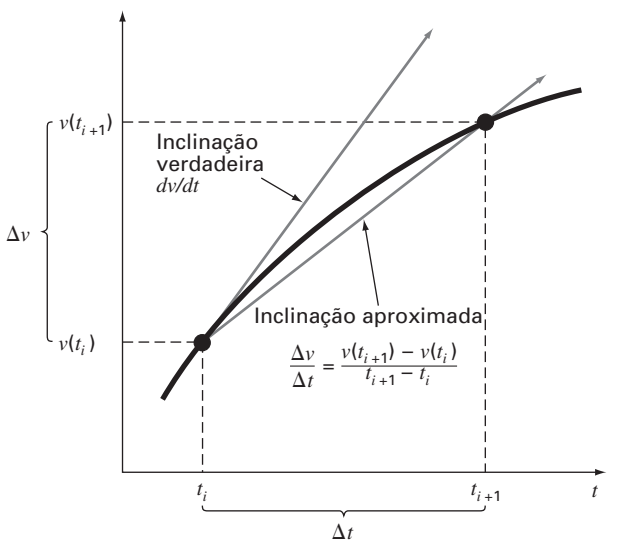


FIGURA 1.4
O uso de uma diferença finita para aproximar a primeira derivada de *v* com relação a *t*.

Observe que o termo entre colchetes é o lado direito da equação diferencial propriamente dita [Equação (1.9)]. Isto é, ela fornece um meio de calcular a taxa de variação ou a inclinação de v. Portanto, a equação diferencial foi transformada em uma equação que pode ser usada para determinar algebricamente a velocidade em t_{i+1} usando a inclinação e os valores anteriores de v e t. Se for dado um valor inicial para a velocidade em algum instante t_i, pode-se facilmente calcular a velocidade em um instante posterior t_{i+1}. Esse novo valor da velocidade em t_{i+1} pode, por sua vez, ser usado para estender o cálculo da velocidade a t_{i+2}, e assim por diante. Portanto, em qualquer instante ao longo do caminho,

Valor novo = valor velho + inclinação × tamanho do passo

Observe que essa abordagem é chamada oficialmente de *método de Euler*.

EXEMPLO 1.2 Solução Numérica para o Problema do Pára-quedista em Queda Livre

Enunciado do Problema. Faça os mesmos cálculos que no Exemplo 1.1, mas use a Equação (1.12) para calcular a velocidade. Use um passo de tamanho 2 s para os cálculos.

Solução. No início dos cálculos ($t_i = 0$), a velocidade do pára-quedista é zero. Usando essa informação e os valores dos parâmetros do Exemplo 1.1, a Equação (1.12) pode ser utilizada para calcular a velocidade em $t_{i+1} = 2$ s:

$$v = 0 + \left[9{,}8 - \frac{12,5}{68,1}(0)\right]2 = 19{,}60 \text{ m/s}$$

Para o intervalo seguinte (de $t = 2$ a 4 s), os cálculos são repetidos, com o resultado

$$v = 19{,}60 + \left[9{,}8 - \frac{12{,}5}{68{,}1}(19{,}60)\right]2 = 32{,}00 \text{ m/s}$$

Continua-se os cálculos, de forma análoga, para se obter valores adicionais:

t, s	v, m/s
0	0,00
2	19,60
4	32,00
6	39,85
8	44,82
10	47,97
12	49,96
∞	53,39

O resultado está representado graficamente na Figura 1.5, juntamente com a solução exata. Pode-se ver que o método numérico retrata as características essenciais da solução exata. Entretanto, como foram usados segmentos de retas para aproximar uma função que se curva continuamente, existe alguma discrepância entre os dois resultados. Uma forma de minimizar tais discrepâncias seria usar um passo de tamanho menor. Por exemplo, a aplicação da Equação (1.12) em intervalos de 1 s resulta em um erro menor, já que os segmentos de retas seguem a solução verdadeira mais de perto. Fazendo-se os cálculos à mão, o esforço associado ao uso de passos cada vez menores tornaria tais soluções numéricas não-práticas. Entretanto, com o auxílio do computador, grandes números de cálculos podem ser feitos facilmente. Portanto, pode-se modelar com exatidão a velocidade do pára-quedista em queda livre sem ter de resolver a equação diferencial exatamente.

Como no exemplo anterior, para resultados numéricos mais acurados, deve-se pagar o preço computacional. Cada vez que dividirmos o tamanho do passo pela metade, para obter mais acurácia teremos de fazer o dobro do número de cálculos. Assim, vemos que

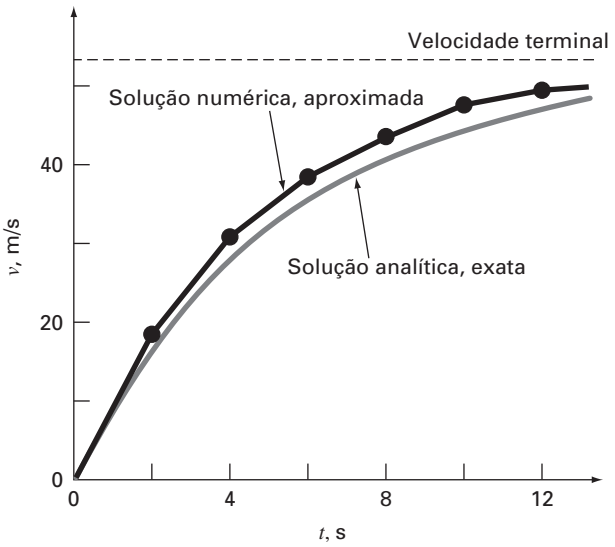


FIGURA 1.5
Comparação das soluções numéricas e analíticas do problema do pára-quedista em queda livre.

há um balanceamento entre a acurácia e o esforço computacional. Tais prós e contras figuram de forma proeminente nos métodos numéricos e constituem um tema importante neste livro. Conseqüentemente, devotamos o Epílogo da Parte Um a uma introdução a mais desses prós e contras.

1.2 LEIS DE CONSERVAÇÃO E ENGENHARIA

Além da segunda lei de Newton, existem outros princípios de organização importantes na engenharia. Entre os mais importantes deles estão as leis de conservação. Embora elas formem a base para uma variedade de modelos matemáticos complicados e poderosos, as grandes leis de conservação da ciência e da engenharia são conceitualmente fáceis de entender. Todas elas se reduzem a

$$\text{Variação} = \text{aumento} - \text{diminuição} \tag{1.13}$$

Essa é precisamente a forma empregada quando se usa a lei de Newton para deduzir o balanço de forças para um pára-quedista em queda livre [Equação (1.8)].

Embora simples, a Equação (1.13) engloba uma das formas mais fundamentais nas quais as leis de conservação são usadas em engenharia — isto é, para predizer variações com relação ao tempo. Dá-se à Equação (1.13) o nome especial de cálculo *dependente do tempo* (ou *transiente*).

Além de predizer variações, uma outra forma na qual as leis de conservação são aplicadas é no caso em que não existe a variação. Se a variação for nula, a Equação (1.13) se torna

$$\text{Variação} = 0 = \text{aumento} - \text{diminuição}$$

ou

$$\text{Aumento} = \text{diminuição} \tag{1.14}$$

Portanto, se não ocorrer nenhuma variação, o aumento e a diminuição devem estar balanceados. Esse caso, que também possui um nome especial — o cálculo de *estado estacionário* — tem muitas aplicações em engenharia. Por exemplo, para escoamentos de fluidos incompressíveis e estacionários em tubos, o escoamento entrando em uma junção deve ser balanceado pelo escoamento saindo, como em

$$\text{Escoamento entrando} = \text{escoamento saindo}$$

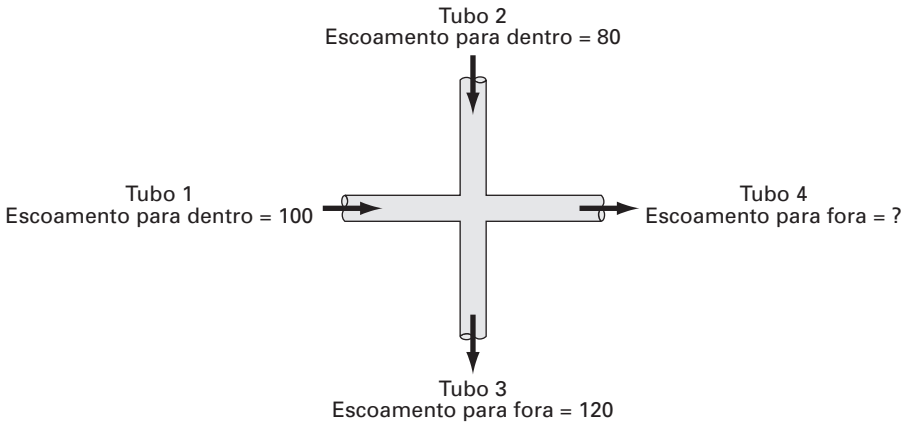


FIGURA 1.6
Um balanço de escoamento para o escoamento de um fluido incompressível e estacionário na junção de dois tubos.

Para a junção na Figura 1.6, o balanço pode ser usado para calcular que o escoamento saindo do quarto tubo deve ser 60.

Para o pára-quedista em queda livre, as condições estacionárias corresponderiam ao caso em que a força resultante fosse nula ou [Equação (1.8) com $dv/dt = 0$]

$$mg = cv \tag{1.15}$$

Desse modo, no estado estacionário, as forças para baixo e para cima estão balanceadas e a Equação (1.15) pode ser resolvida para se determinar a velocidade terminal

$$v = \frac{mg}{c}$$

Embora as Equações (1.13) e (1.14) possam parecer trivialmente simples, elas englobam as duas formas fundamentais nas quais as leis de conservação são empregadas em engenharia. Assim, elas constituirão uma parte importante de nossos esforços nos capítulos subseqüentes para ilustrar a conexão entre os métodos numéricos e a engenharia. Nossos veículos primários para fazer essa conexão são as aplicações em engenharia que aparecem no final de cada parte deste livro.

A Tabela 1.1 resume alguns dos modelos simples de engenharia e as leis de conservação associadas que formarão a base para muitas dessas aplicações em engenharia. A maioria das aplicações em engenharia química será focalizada no balanço de massa para reatores. O balanço de massa é deduzido da conservação de massa. Ele especifica que a variação de massa de um produto químico no reator depende da diferença da quantia de massa escoando para dentro e da quantia de massa escoando para fora.

As aplicações tanto em engenharia civil quanto em engenharia mecânica serão focalizadas em modelos desenvolvidos a partir da conservação do momento. Para a engenharia civil, os balanços de força são utilizados para analisar estruturas como a treliça simples na Tabela 1.1. Os mesmos princípios são usados nas aplicações em engenharia mecânica para analisar o movimento transiente para cima e para baixo ou as vibrações de um automóvel.

Finalmente, as aplicações em engenharia elétrica usam tanto balanço de corrente quanto de energia para modelar os circuitos elétricos. O balanço de corrente, que resulta da conservação da carga, é parecido — em essência — com o balanço de escoamento mostrado na Figura 1.6. Da mesma forma como o escoamento deve ser balanceado em uma junção de tubos, a corrente elétrica deve ser balanceada em uma junção de fios elétricos. O balanço de energia especifica que a variação de voltagem ao redor de qualquer laço de um circuito deve totalizar zero. As aplicações em engenharia são planejadas para ilustrar como os métodos numéricos são realmente empregados no processo de resolução de problemas de engenharia. Assim, elas vão permitir explorar questões práticas (Tabela 1.2) que aparecem em aplicações do mundo real. Fazer essas conexões entre as técnicas matemáticas, como os métodos numéricos, e a prática da engenharia é um passo crítico para usar seu verdadeiro potencial. Um exame cuidadoso das aplicações de engenharia vai ajudá-lo a dar esse passo.

TABELA 1.1 Dispositivos e tipos de balanços que são comumente usados nas quatro áreas principais da engenharia. Para cada caso, a lei de conservação na qual o balanço é baseado está especificada.

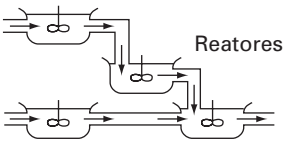
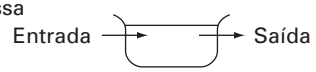
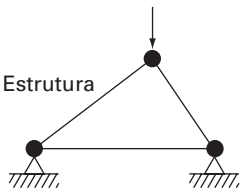
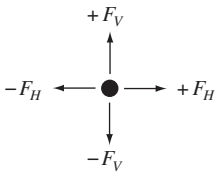
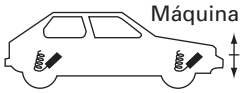
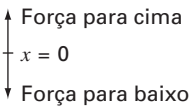
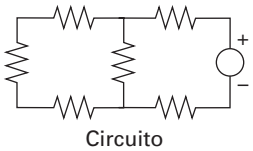
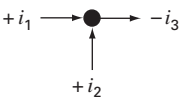
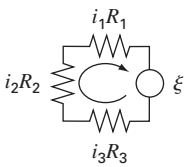
Área	Dispositivo	Princípio organizacional	Expressão matemática
Engenharia química		Conservação da massa	Balanço de massa  Sobre um período unitário de tempo Δmassa = entrada – saída
Engenharia civil		Conservação do momento	Balanço de força:  Em cada nó Σ forças horizontais (F_H) = 0 Σ forças verticais (F_V) = 0
Engenharia mecânica		Conservação do momento	Balanço de força:  $m\frac{d^2x}{dt^2}$ = força para baixo – força para cima
Engenharia elétrica		Conservação da carga	Balanço de corrente:  Para cada nó Σ corrente (i) = 0
		Conservação da energia	Balanço de voltagem:  Em torno de cada laço Σ fem's – Σ queda de voltagem por resistores = 0 Σ ξ – Σ iR = 0

TABELA 1.2 Algumas questões práticas que serão exploradas nas aplicações de engenharia no final de cada parte deste livro.

1. *Não-linear* versus *linear*. Muito da engenharia clássica depende da linearização para permitir soluções analíticas. Embora muitas vezes isso seja apropriado, pode-se aumentar a percepção ao se examinar os problemas não-lineares.
2. *Sistemas grandes* versus *pequenos*. Sem um computador, em geral não é possível examinar sistemas com mais de três componentes interagindo. Com computadores e métodos numéricos, sistemas com muitos componentes, mais realistas, podem ser examinados.
3. *Não-ideal* versus *ideal*. As leis idealizadas abundam em engenharia. Em geral, existem alternativas não-idealizadas que são mais realistas, mas que exigem mais do ponto de vista computacional. Abordagens numéricas apropriadas podem facilitar a aplicação dessas relações não-ideais.
4. *Análise de sensibilidade*. Como esse tipo de análise é bastante complicado, muitos cálculos manuais requerem uma grande quantidade de tempo e esforço para uma implementação bem-sucedida. Algumas vezes, isso desencoraja o analista de implementar os cálculos múltiplos que são necessários para examinar como o sistema responde sob diferentes condições. Tais análises de sensibilidade são facilitadas quando os métodos numéricos permitem que o computador assuma toda a carga computacional.
5. *Projeto*. Em geral, é uma tarefa simples determinar o desempenho de um sistema como uma função de seus parâmetros. Normalmente, é mais difícil resolver o problema inverso — isto é, determinar os parâmetros quando exigido é especificado. Os métodos numéricos e os computadores freqüentemente permitem que essa tarefa seja implementada de forma eficiente.

PROBLEMAS

1.1 A água é responsável por aproximadamente 60% do peso total do corpo. Assumindo que ela pode ser categorizada em seis regiões, as percentagens são como a seguir. O plasma é responsável por 4,5% do peso do corpo e por um total de 7,5% da água total do corpo. O tecido conectivo denso e as cartilagens ocupam 4,5% do peso total do corpo e 7,5% da água total do corpo. O tecido intersticial é 12% do peso do corpo, o que dá 20% da água total do corpo. A água inacessível do osso é aproximadamente 7,5% da água total do corpo e 4,5% do peso total do corpo. Se a água intracelular for 33% do peso total do corpo e a água transcelular for 2,5% da água total do corpo, que percentagem do peso total do corpo deve ser água transcelular e que percentagem da água total do corpo deve ser a água intracelular?

1.2 Um grupo de 30 alunos assiste aula em uma sala que mede 10 m por 8 m por 3 m. Cada aluno ocupa cerca de 0,075 m^3 e libera cerca de 80 W de calor (1 W = 1 J/s). Calcule o aumento da temperatura do ar durante os primeiros 15 minutos de aula, se a sala estiver completamente selada e isolada. Suponha que a capacidade térmica, C_v, do ar seja 0,718 kJ/(kg K). Suponha que o ar seja um gás ideal a 20 °C e 101,325 kPa. Observe que o calor absorvido pelo ar, Q, está associado à massa de ar m, à capacidade calorífica e à variação de temperatura pela seguinte relação:

$$Q = m\int_{T_1}^{T_2} C_v dT = mC_v(T_2 - T_1)$$

A massa de ar pode ser obtida da lei dos gases ideais:

$$PV = \frac{m}{\text{Mwt}}RT$$

onde P é a pressão do gás, V é o volume do gás, Mwt é o peso molecular do gás (para o ar, 28,97 kg/kmol) e R é a constante dos gases ideais [8,314 kPa m^3/(kmol K)].

1.3 A seguinte informação está disponível para uma conta de banco:

Data	Depósitos	Saques	Balanço
1/5			1512,33
	220,13	327,26	
1/6			
	216,80	378,61	
1/7			
	450,25	106,80	
1/8			
	127,31	350,61	
1/9			

Use a conservação do dinheiro para calcular o balanço em 1/6, 1/7, 1/8 e 1/9. Mostre cada passo dos cálculos. Esse é um cálculo estacionário ou transiente?

1.4 A vazão em volume através de um tubo é dada por $Q = vA$, onde v é a velocidade média e A é a área da seção transversal. Use a continuidade do volume para encontrar a área pedida no tubo 3.

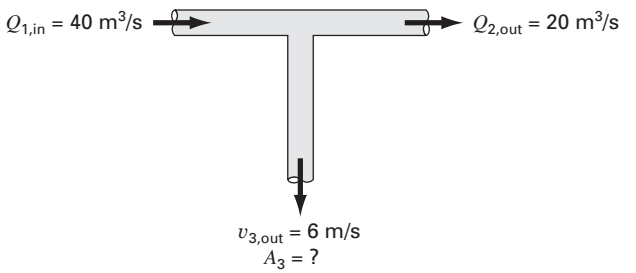


Figura P1.4

1.5 A Figura P1.5 mostra as diversas formas pelas quais um homem médio ganha e perde água durante um dia. Um litro é ingerido com a comida e o corpo metabolicamente produz 0,3 L. Respirando o ar, a troca é 0,05 L enquanto inalando, e 0,4 L enquanto exalando, em um período de um dia. O corpo também vai perder 0,2, 1,4, 0,2 e 0,35 L através de suor, urina, fezes e pela pele, respectivamente. Para manter uma condição estacionária, quanta água deve ser bebida por dia?

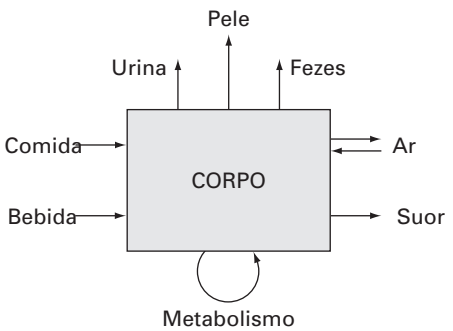


Figura P1.5

1.6 Para o pára-quedista em queda livre com arrasto linear, considere inicialmente um indivíduo de 70 kg e que tenha um coeficiente de arrasto de 12 kg/s. Se um segundo pára-quedista tem um coeficiente de arrasto de 15 kg/s e uma massa de 75 kg, quanto tempo levará para ele atingir a mesma velocidade que o primeiro atingiu em 10 s?

1.7 Use cálculo para resolver a Equação (1.9) no caso em que a velocidade inicial $v(0)$ é não-nula.

1.8 Repita o Exemplo 1.2. Calcule a velocidade até 10 s, com o tamanho do passo (**a**) 1 e (**b**) 0,5 s. Você pode fazer alguma afirmação com relação aos erros nos cálculos baseado nos resultados?

1.9 Em vez da relação linear da Equação (1.7), você poderia escolher modelar a força para cima no pára-quedista por uma relação de segundo grau,

$$F_U = -c'v^2$$

onde c' é um coeficiente de arrasto de segunda ordem (kg/m).

(a) Usando o cálculo, obtenha a solução na forma fechada, no caso em que o pára-quedista está inicialmente em repouso ($v = 0$ em $t = 0$).

(b) Repita o cálculo numérico do Exemplo 1.2 com a mesma condição inicial e os mesmos valores dos parâmetros. Use um valor de 0,225 kg/m para c'.

1.10 Calcule a velocidade de um pára-quedista em queda livre usando o método de Euler para o caso em que $m = 80$ kg e $c = 10$ kg/s. Faça os cálculos de $t = 0$ até 20 s, com um tamanho de passo 1 s. Use a condição inicial em que o pára-quedista tem uma velocidade para cima de 20 m/s em $t = 0$. Em $t = 10$ s, suponha que o pára-quedas é instantaneamente aberto, de forma que o coeficiente de arrasto pula para 50 kg/s.

1.11 No exemplo do pára-quedista em queda livre, assumimos que a aceleração devida à gravidade era uma constante de valor 9,8 m/s^2. Embora esta seja uma boa aproximação quando estamos examinando objetos caindo perto da superfície da Terra, a força gravitacional decresce quando nos movemos para cima do nível do mar. Uma representação mais geral, baseada na lei da gravitação universal de Newton, pode ser escrita como

$$g(x) = g(0)\frac{R^2}{(R+x)^2}$$

onde $g(x)$ é a aceleração da gravidade na altura x (em m) medida para cima, a partir da superfície da Terra (m/s^2), $g(0)$ é a aceleração da gravidade na superfície da Terra ($\cong 9{,}8$ m/s^2), e R é o raio da Terra ($\cong 6{,}37 \times 10^6$ m).

(a) Em uma forma similar à dedução da Equação (1.9), utilize o balanço de forças para deduzir uma equação diferencial para a velocidade como função do tempo que use essa representação mais completa da gravitação. Entretanto, para tal dedução, suponha que a velocidade para cima é positiva.

(b) Para o caso de o arrasto ser desprezível, use a regra da cadeia para expressar a equação diferencial como uma função da altura em vez do tempo. Lembre-se de que a regra da cadeia é

$$\frac{dv}{dt} = \frac{dv}{dx}\frac{dx}{dt}$$

(c) Use o cálculo para obter uma solução na forma fechada, onde $v = v_0$ em $x = 0$.

(d) Use o método de Euler para obter uma solução numérica de $x = 0$ até 100.000 utilizando um passo de 10.000 m, com velocidade inicial de 1.400 m/s para cima. Compare seu resultado com a solução analítica.

1.12 A quantidade de contaminante radioativo uniformemente distribuído contido em um reator fechado é medida pela sua concentração c (becquerel/litro ou Bq/L). O contaminante diminui a uma taxa de decaimento proporcional a sua concentração — isto é

$$\text{taxa de decaimento} = -kc$$

onde k é uma constante, com unidades de dia^{-1}. Portanto, de acordo com a Equação (1.13), o balanço de massa para o reator pode ser escrito como

$$\frac{dc}{dt} = -kc$$

$$\begin{pmatrix}\text{variação}\\ \text{na massa}\end{pmatrix} = \begin{pmatrix}\text{diminuição}\\ \text{por decaimento}\end{pmatrix}$$

(a) Use o método de Euler para resolver essa equação de $t = 0$ até 1 d com $k = 0{,}2\,\text{d}^{-1}$. Use um tamanho de passo $\Delta t = 0{,}1$. A concentração em $t = 0$ é 10 Bq/L.

(b) Trace a solução em um gráfico semilog (isto é, em c *versus* t) e determine a inclinação. Interprete seus resultados.

1.13 Um tanque de armazenamento contém um líquido à profundidade y, onde $y = 0$ quando o tanque está cheio até a metade. É tirado líquido a uma vazão constante Q, para atender à demanda. O conteúdo é reposto a uma taxa senoidal de $3Q\,\text{sen}^2(t)$.

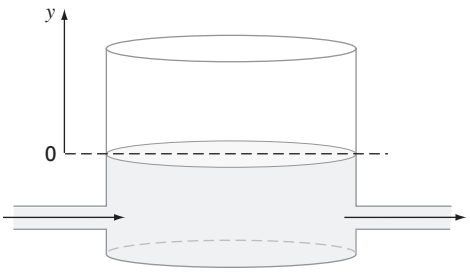


Figura P1.13

A Equação (1.13) pode ser escrita para esse sistema como

$$\frac{d(Ay)}{dx} = 3Q\,\text{sen}^2(\text{t}) - \text{Q}$$

$$\begin{pmatrix}\text{variação no}\\ \text{volume}\end{pmatrix} = \text{(escoamento p/ dentro)} - \text{(escoamento p/ fora)}$$

ou, já que a área da superfície, A, é constante

$$\frac{dy}{dx} = 3\frac{Q}{A}\,\text{sen}^2(\text{t}) - \frac{\text{Q}}{\text{A}}$$

Use o método de Euler para encontrar a profundidade y de $t = 0$ a 10 d com um tamanho de passo 0,5 d. Os valores do parâmetro são $A = 1200$ m^2 e $Q = 500$ m^3/d. Suponha que a condição inicial seja $y = 0$.

1.14 Para o mesmo tanque de armazenamento descrito no Problema 1.13, suponha que o escoamento para fora não seja constante, mas dependa da profundidade. Nesse caso, a equação diferencial para a profundidade pode ser escrita como

$$\frac{dy}{dx} = 3\frac{Q}{A}\,\text{sen}^2(\text{t}) - \frac{\alpha(1+\text{y})^{1.5}}{\text{A}}$$

Use o método de Euler para encontrar a profundidade y de $t = 0$ a 10 d com um tamanho de passo 0,5 d. Os valores do parâmetro são $A = 1200$ m^2, $Q = 500$ m^3/d, e $\alpha = 300$. Suponha que a condição inicial seja $y = 0$.

1.15 Suponha que uma gotícula esférica de líquido evapore a uma taxa que é proporcional à área de sua superfície.

$$\frac{dV}{dt} = -kA$$

onde V é o volume (mm^3), t é o tempo (h), k é a taxa de evaporação (mm/h), e A é a área da superfície (mm^2). Use o método de Euler para calcular o volume da gotícula de $t = 0$ a 10 min usando um tamanho de passo 0,25 min. Suponha que $k = 0{,}1$ mm/min e que a gotícula tenha inicialmente um raio de 3 mm. Verifique a validade de seus resultados determinando o raio do seu volume final calculado e verificando que ele é consistente com a taxa de evaporação.

1.16 A lei do resfriamento de Newton diz que a temperatura de um corpo varia a uma taxa proporcional à diferença entre a sua temperatura e a temperatura do meio que o cerca (a temperatura ambiente),

$$\frac{dT}{dt} = -k(T - T_a)$$

onde T é a temperatura do corpo (°C), t é o tempo (min.), k é a constante de proporcionalidade (por minuto) e T_a é a temperatura ambiente (°C). Suponha que uma xícara de café originalmente tenha a temperatura de 68° C. Use o método de Euler para calcular a temperatura de $t = 0$ a 10 min usando um tamanho de passo 1 min se $T_a = 21°$ C e $k = 0{,}017$/min.

1.17 As células cancerígenas crescem exponencialmente, levando 20 horas para dobrar quando têm um suprimento ilimitado de nutrientes. Entretanto, conforme as células começam a formar um tumor esférico sólido, sem suprimento de sangue, o crescimento no centro do tumor se torna limitado e, eventualmente, as células começam a morrer.

(a) O crescimento exponencial do número de células N pode ser expresso como mostrado, onde μ é a taxa de crescimento das células. Para células de câncer, encontre o valor de μ.

$$\frac{dN}{dt} = \mu N$$

(b) Escreva uma equação que descreva a taxa de variação do volume do tumor durante o crescimento exponencial, dado que o diâmetro de uma célula individual é de 20 mícrons.

(c) Depois que um tipo particular de tumor passa de 500 mícrons no diâmetro, as células no centro do tumor morrem (mas continuam a tomar espaço no tumor). Determine quanto tempo levará para que o tumor passe desse tamanho crítico.

1.18 Um fluido é bombeado para dentro da rede mostrada na Figura P1.18. Se $Q_2 = 0{,}6$, $Q_3 = 0{,}4$, $Q_7 = 0{,}2$ e $Q_8 = 0{,}3$ m^3/s determine os outros escoamentos.

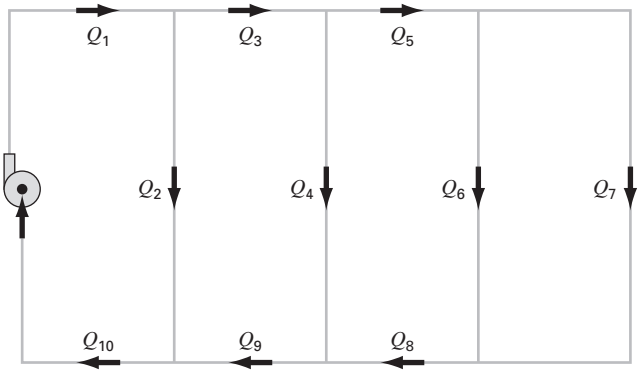


Figura P1.18

CAPÍTULO 2

Programação e Software

No capítulo anterior, usamos a força resultante para desenvolver um modelo matemático para predizer a velocidade de queda de um pára-quedista. Esse modelo tomou a forma de uma equação diferencial,

$$\frac{dv}{dt} = g - \frac{c}{m}v$$

Aprendemos também que a solução dessa equação pode ser obtida por uma aproximação numérica simples chamada método de Euler,

$$v_{i+1} = v_i + \frac{dv_i}{dt}\Delta t$$

Dadas as condições iniciais, essa equação pode ser implementada repetidamente para calcular a velocidade como uma função do tempo. Entretanto, para obter uma boa acurácia, é preciso tomar um grande número de passos, o que pode ser extremamente trabalhoso e demorado para implementar manualmente. Mas, com a ajuda de um computador, tais cálculos podem ser realizados facilmente.

Portanto, nossa tarefa é descobrir como fazer isso. O presente capítulo mostrará a você como o computador é utilizado como ferramenta para obter esse tipo de solução.

2.1 PACOTES E PROGRAMAÇÃO

Hoje, há dois tipos de usuários de software. Por um lado, existem aqueles que usam o que lhes é dado. Isto é, eles se limitam aos recursos encontrados no modo de operação padrão do software. Por exemplo, é uma proposição direta resolver um sistema de equações lineares ou gerar um gráfico de valores *xy* com o software Excel ou o MATLAB. Como isso envolve, geralmente, um mínimo de esforço, muitos usuários tendem a adotar esse modo de operação superficial. Ademais, uma vez que os projetistas desses pacotes já se anteciparam a várias necessidades dos usuários, muitos problemas importantes podem ser resolvidos dessa forma.

Mas o que acontece quando surgem problemas que estão além da capacidade padrão da ferramenta? Infelizmente, agitar as mãos e dizer "Desculpe, patrão, não dá pra fazer!" não é aceitável nos círculos da engenharia. Em muitos casos, há duas alternativas.

Primeiro, pode-se buscar um pacote diferente e verificar se ele é capaz de resolver o problema. Essa é uma das razões que nos levaram a escolher tratar tanto do Excel como do MATLAB neste livro. Como você vê, nenhum deles é completo e cada um tem potencial diferente. Estar aberto a ambos ampliará enormemente o universo de problemas que se pode resolver.

Segundo, você pode se tornar um "usuário poderoso" aprendendo a escrever Excel VBA[1] macros ou M-files em MATLAB. E o que é isso? Não passam de programas de computador que permitem que se estenda o potencial de tais ferramentas. Em razão de os engenheiros nunca aceitarem ficar limitados por ferramentas, eles farão o que for necessário para resolver seus problemas. Uma maneira de conseguir isso é aprender a escrever programas nos ambientes Excel e MATLAB. Afinal, as habilidades requeridas

[1] VBA é a sigla para *Visual Basic para Aplicações*.

para fazer macros e M-files são as mesmas usadas para desenvolver eficientemente programas em linguagens como Fortran 90 ou C.

O objetivo principal do presente capítulo é mostrar como isso pode ser feito. Entretanto, estamos assumindo que você já foi exposto aos rudimentos da programação de computadores. Assim, nossa ênfase aqui é em aspectos da programação que dizem respeito ao seu uso na solução de problemas de engenharia.

2.1.1 Programas Computacionais

Programas computacionais são, basicamente, um conjunto de instruções que direcionam o computador para executar certas tarefas. Uma vez que muitos indivíduos escrevem programas para um amplo conjunto de aplicações, muitas linguagens de computação de alto nível, como Fortran 90 e C, têm grandes recursos. Embora alguns engenheiros necessitem explorar totalmente esses recursos, a maioria necessitará simplesmente da habilidade necessária para executar cálculos numéricos orientados para engenharia.

Olhando dessa perspectiva, podemos reduzir a complexidade para um conjunto limitado de tópicos de programação. Estes são:

- Representação simples da informação (constantes, variáveis e declarações de tipo).
- Representação de informação avançada (estruturas de dados, vetores e registros).
- Fórmulas matemáticas (atribuições, regras de prioridade e funções intrínsecas).
- Entrada/saída.
- Representação lógica (seqüência, seleção e repetição).
- Programação modular (funções e sub-rotinas).

Como estamos assumindo que você teve uma exposição prévia à programação, não gastaremos muito tempo nas primeiras quatro áreas citadas. No máximo, nós as oferecemos na forma de uma lista de verificação que cobre o que é necessário saber para implementar os programas que seguem.

No entanto, dedicaremos algum tempo aos dois últimos tópicos. Enfatizamos a representação lógica porque ela é a área simples que mais influencia a coerência de um algoritmo e sua compreensão. Incluiremos programação modular porque ela também contribui grandemente para a organização dos programas. Adicionalmente, módulos provêem um meio para armazenar algoritmos úteis em um formato convencional para aplicações subseqüentes.

2.2 PROGRAMAÇÃO ESTRUTURADA

Nos primórdios da computação, os programadores usualmente não prestavam muita atenção ao fato de o programa ser claro e fácil de entender. Hoje, é reconhecido que existem muitos benefícios em escrever um código bem organizado e bem estruturado. Além do benefício óbvio de tornar o software mais fácil de ser compartilhado, isso também ajuda a garantir um desenvolvimento de programa mais eficiente.

Portanto, algoritmos bem estruturados são, invariavelmente, fáceis de corrigir e testar, resultando em programas que têm um tempo de desenvolvimento e atualização menor.

Cientistas da computação têm estudado sistematicamente os fatores e procedimentos necessários para desenvolver software de alta qualidade. Essencialmente, *programação estruturada* é um conjunto de regras que exige disciplina do programador. Embora a programação estruturada seja flexível o suficiente para permitir considerável criatividade e expressões pessoais, suas regras impõem restrições suficientes para conferir ao código resultante grande superioridade, comparada à versão não-estruturada. Em particular, o produto final é mais elegante e fácil de entender.

A idéia central por trás da programação estruturada é que qualquer algoritmo numérico pode ser composto usando-se as três estruturas de controle fundamentais: seqüência, seleção e repetição. Se nos limitarmos a essas estruturas, o código computacional resultante será claro e fácil de seguir.

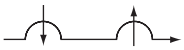
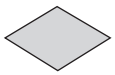
SÍMBOLO	NOME	FUNÇÃO
	Terminal	Representa o início e o final do programa.
	Linhas de fluxo	Representa o fluxo da lógica. Os morros nas flechas horizontais indicam cruzamento sem que haja conexão com o fluxo vertical.
	Processos	Representa cálculos ou manipulação de dados.
	Entrada/saída	Representa entrada ou saída de dados e informação.
	Decisão	Representa uma comparação, questão, ou decisão que determina caminhos alternativos a serem seguidos.
	Junção	Representa a confluência de linhas de fluxo.
	Conector de página	Representa uma quebra que é continuada em outra página.
	Contadores controladores de laços	Usado para ciclos que se repetem por um número predeterminado de iterações.

FIGURA 2.1
Símbolos usados nos diagramas de fluxo.

Nos parágrafos seguintes, descreveremos cada uma dessas estruturas. Para manter essa descrição genérica, empregaremos os fluxogramas e pseudocódigo. Um *fluxograma* é uma representação gráfica visual de um algoritmo, e emprega uma série de blocos e flechas, cada um dos quais representa uma operação particular ou passos de um algoritmo (Figura 2.1). As flechas representam a seqüência na qual as operações são implementadas.

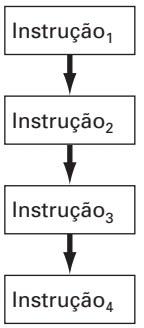
Nem todos os envolvidos em programação de computadores concordam que o fluxograma é um esquema produtivo. De fato, alguns programadores experientes não defendem o fluxograma. Entretanto, achamos que existem três boas razões para estudá-los. Primeiro, eles ainda são usados para expressar e comunicar algoritmos. Segundo, mesmo que não sejam empregados rotineiramente, existirão ocasiões nas quais se mostrarão úteis no planejamento, na resolução, ou comunicando a lógica do seu programa ou do programa de outra pessoa. Finalmente, e mais importante para o nosso propósito, eles são uma excelente ferramenta pedagógica. Do ponto de vista do ensino, fluxogramas são veículos ideais para visualizar algumas das estruturas de controle fundamentais empregadas na programação de computadores.

Uma abordagem alternativa que representa um algoritmo para cobrir a lacuna entre um fluxograma e o código computacional é chamada *pseudocódigo*, técnica que utiliza palavras semelhantes a código no lugar dos símbolos gráficos do fluxograma. Adotaremos convenções de estilo para o pseudocódigo neste livro. Palavras reservadas como IF, DO, INPUT etc. são escritas em letras maiúsculas, enquanto condições, passos de processamento e tarefas são escritas com letras minúsculas. Adicionalmente, os passos de processamento são identados. Assim, as palavras reservadas formam um "sanduíche" em torno dos passos para definir visualmente a extensão da estrutura de controle.

Uma vantagem do pseudocódigo é que é mais fácil desenvolver um programa com ele que com um fluxograma. O pseudocódigo também é mais fácil de modificar e compartilhar. Entretanto, por causa da sua forma gráfica, fluxogramas às vezes se adaptam melhor para a visualização de algoritmos complexos. Neste texto, usaremos fluxogramas por razões pedagógicas. Pseudocódigos serão o nosso principal veículo para comunicar algoritmos relacionados a métodos numéricos.

2.2.1 Representação Lógica

Seqüência. A estrutura da seqüência expressa a idéia trivial de que, exceto se direcionado de outra forma, o código computacional é implementado uma instrução por vez. Como na Figura 2.2, a estrutura pode ser expressa genericamente com um fluxograma ou um pseudocódigo.




```
Instrução1
Instrução2
Instrução3
Instrução4
```

(*a*) Fluxograma (*b*) Pseudocódigo

FIGURA 2.2
(*a*) Fluxograma
(*b*) Pseudocódigo para a estrutura-seqüência.

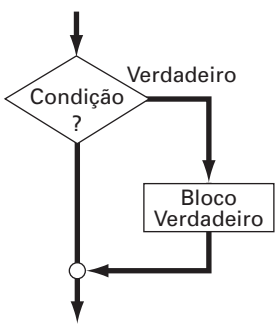
Seleção. Em contraste com a estrutura passo a passo da seqüência, a seleção provê um meio de separar o fluxo do programa em ramos com base no valor lógico da condição. A Figura 2.3 mostra as duas formas fundamentais para fazer isso.

A decisão com alternativa única, ou estrutura *IF/THEN* (Figura 2.3*a*), permite um desvio no fluxo do programa se o valor lógico da condição é verdadeiro. Caso seja falso, nada acontece e o programa vai diretamente para a próxima instrução depois do ENDIF. A decisão com dupla alternativa, ou estrutura *IF/THEN/ELSE* (Figura 2.3*b*), comporta-se da mesma maneira para a condição verdadeira. No entanto, se a condição é falsa, o programa implementa o código entre o ELSE e o ENDIF.

Embora as estruturas IF/THEN e IF/THEN/ELSE sejam suficientes para construir qualquer algoritmo numérico, duas outras variantes são comumente usadas. Suponha

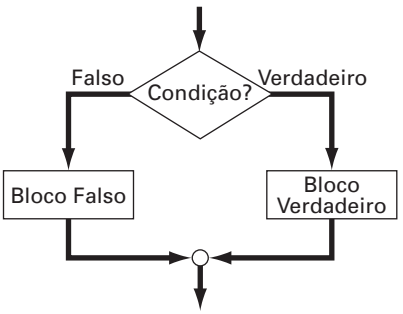
FIGURA 2.3
Fluxograma e pseudocódigo para a estrutura de seleção.
(*a*) Seleção com alternativa única (IF/THEN) e
(*b*) seleção de dupla alternativa (IF/THEN/ELSE).

Fluxograma Pseudocódigo




```
IF condição THEN
  Bloco Verdadeiro
ENDIF
```

(*a*) Estrutura com alternativa única (IF/THEN)




```
IF condição THEN
  Bloco Verdadeiro
ELSE
  Bloco Falso
ENDIF
```

(*b*) Estrutura com dupla alternativa (IF/THEN/ELSE)

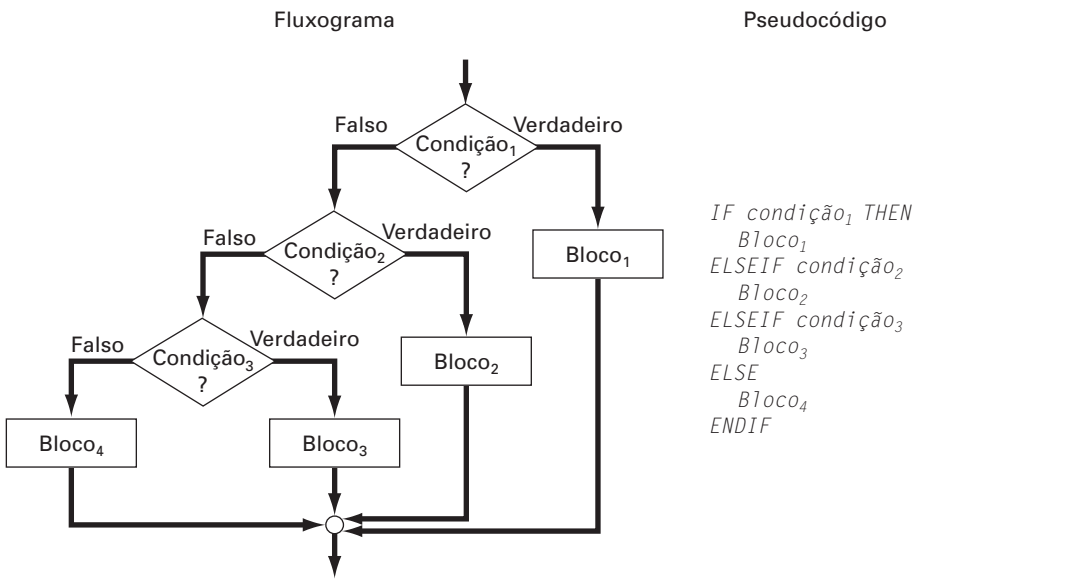
que a cláusula ELSE de um IF/THEN/ELSE contenha outro IF/THEN. Para esses casos, o ELSE e o IF podem ser combinados na estrutura *IF/THEN/ELSEIF* mostrada na Figura 2.4*a*.

Observe como na Figura 2.4*a* existe uma cadeia ou "cascata" de decisões. A primeira é um comando IF, e cada decisão sucessiva é um comando ELSEIF. Mergulhando na cadeia, a primeira condição encontrada que seja verdadeira causará um desvio para o respectivo bloco seguido do encerramento da estrutura. No final da cadeia de condições, se todas as condições se verificam falsas, um ELSE opcional pode ser incluído.

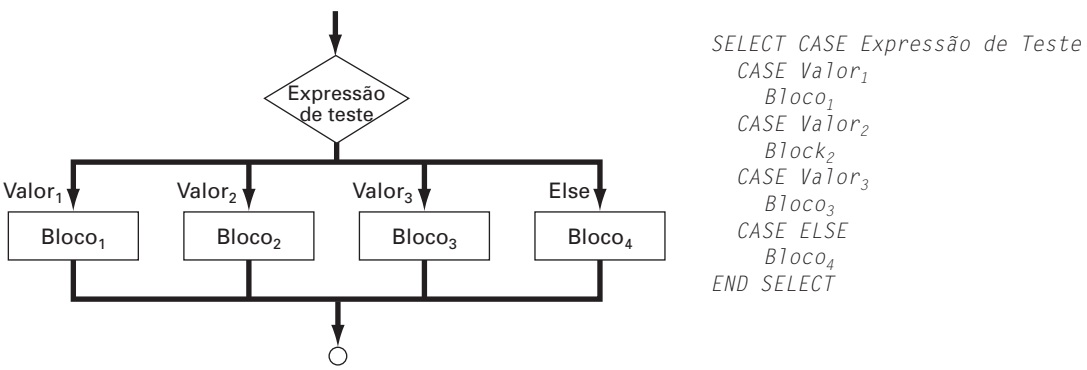
A estrutura *CASE* é uma variante desse tipo de estrutura de decisão. Em vez de testar as condições individualmente, o desvio é realizado com base no valor de uma simples *expressão de teste*. Dependendo desse valor, diferentes blocos de código podem ser executados. Além disso, um bloco opcional pode ser implementado se a expressão não assume nenhum dos valores prescritos (CASE ELSE).

Repetição. A repetição provê os meios para executar uma instrução repetidamente. A estrutura resultante, chamada *laço*, pode ter dois "aspectos" distintos, de acordo com a forma como termina.

FIGURA 2.4
Fluxograma e pseudocódigo para seleção suplementar ou desvio condicional. (*a*) Seleção de múltipla alternativa (IF/THEN/ELSEIF) e (*b*) diagrama CASE.



(*a*) Estrutura multialternativa (IF/THEN/ELSEIF)



(*b*) Estrutura CASE (SELECT ou SWITCH)

O primeiro tipo e o mais básico é o chamado *laço de decisão*, porque termina de acordo com o resultado de uma condição lógica. A Figura 2.5 mostra o tipo mais geral de *laço de decisão*, a *estrutura DOEXIT*, também chamada *laço de interrupção*. Essa estrutura se repete até que a condição lógica seja verdadeira.

Não é necessário ter dois blocos nessa estrutura. Se o primeiro bloco não é incluído, a estrutura é chamada de **laço** *com pré-teste*, porque o teste lógico é feito antes que alguma coisa ocorra. Alternativamente, se o segundo bloco é omitido, a estrutura é chamada de *laço com pós-teste*. Como os dois blocos são incluídos, o caso geral na Figura 2.5 é chamado de laço *de teste intermediário*.

Deve-se observar que o laço DOEXIT foi introduzido no Fortran 90 em um esforço para simplificar decisões em ciclos. Essa estrutura de controle é um padrão na linguagem de macros do Excel VBA, mas não é um padrão em C ou MATLAB, os quais usam a estrutura WHILE. Por acreditarmos que o DOEXIT é superior, nós o adotamos como a estrutura de repetição padrão neste livro. Para assegurar que os nossos algoritmos são diretamente implementáveis tanto no MATLAB como no Excel, mostraremos como o laço de interrupção pode ser simulado com a estrutura WHILE mais adiante neste capítulo (ver Seção 2.5).

O **laço** de interrupção na Figura 2.5 é chamado de **laço** lógico porque termina com uma condição lógica. Em contraste, o *laço DOFOR* (Figura 2.6), *controlado por contador*, executa um número especificado de repetições, ou iterações.

O **laço** controlado por contador trabalha da seguinte forma. O índice (representado por *i* na Figura 2.6) é uma variável que recebe um valor inicial *início*. O programa então testa se o *índice* é inferior ou igual ao valor final, *fim*. Se for, este executa o

FIGURA 2.5
O *laço* de interrupção DOEXIT.

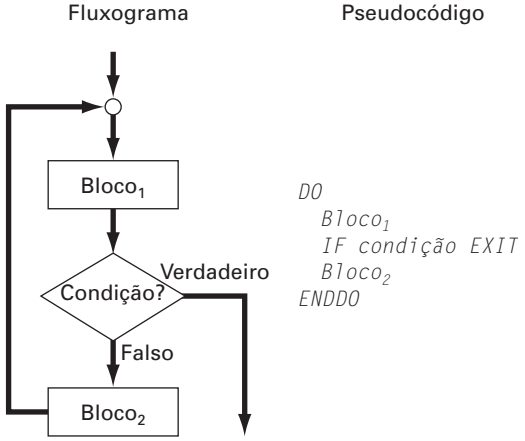
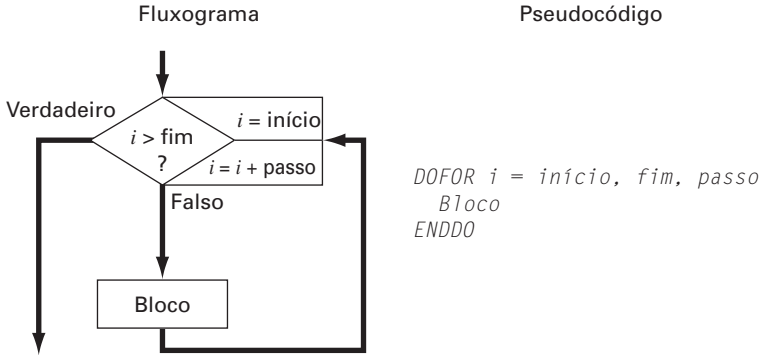


FIGURA 2.6
Estrutura controlada por contador ou DOFOR.



corpo do laço e retorna para a condição do DO. Toda vez que o ENDDO é encontrado, o índice é automaticamente incrementado do valor *passo*. Portanto, o índice atua como um contador. Assim, quando o índice é maior que o valor final (*fim*), o computador automaticamente sai do **laço** e transfere o controle para a linha seguinte ao comando ENDDO. Note que, para praticamente todas as linguagens de computação, incluindo as do Excel e MATLAB, se o valor de *passo* é omitido, o computador assume que este é igual a 1.[2]

Os algoritmos numéricos delineados nas páginas seguintes serão desenvolvidos exclusivamente das estruturas introduzidas pelas Figuras 2.2 a 2.6. O exemplo a seguir ilustra a abordagem básica para desenvolver um algoritmo para determinar as raízes de uma equação quadrática.

EXEMPLO 2.1 Algoritmo para as Raízes de uma Equação Quadrática

Enunciado do Problema. As raízes de uma equação quadrática

$$ax^2 + bx + c = 0$$

podem ser determinadas com a fórmula quadrática,

$$\begin{matrix} x_1 \\ x_2 \end{matrix} = \frac{-b \pm \sqrt{|b^2 - 4ac|}}{2a} \tag{2.1}$$

Desenvolva um algoritmo que faça o seguinte:

Passo 1: Peça ao usuário os coeficientes *a*, *b* e *c*.
Passo 2: Implemente a fórmula quadrática, evitando qualquer eventualidade (por exemplo, evitando divisão por zero e permitindo raízes complexas).
Passo 3: Mostre a solução, isto é, os valores de *x*.
Passo 4: Permita ao usuário a opção de retornar ao passo 1 e repetir o processo.

Solução. Usaremos uma abordagem *top-down* para desenvolver o algoritmo. Ou seja, refinaremos sucessivamente o algoritmo em vez de tentar resolver todos os detalhes na primeira tentativa.

Para fazer isso, vamos assumir, no momento, que a fórmula quadrática é infalível, independentemente dos valores dos coeficientes (obviamente, isso não é verdade, mas é bom o suficiente por ora). Um algoritmo estruturado para implementar o esquema é

```
DO
  INPUT a, b, c
  r1 = (−b + SQRT(b² − 4ac))/(2a)
  r2 = (−b − SQRT(b² − 4ac))/(2a)
  DISPLAY r1, r2
  DISPLAY 'Tentar novamente? Responda sim ou não'
  INPUT resposta
  IF resposta = 'não' EXIT
ENDDO
```

Uma estrutura DOEXIT é usada para implementar a fórmula quadrática repetidamente enquanto a condição é falsa. A condição depende do valor da variável de caracteres, *resposta*. Se *resposta* é igual a '*sim*', então os cálculos são executados. Caso contrário, isto é, se *resposta* = 'não', o laço termina. Assim, o usuário controla o término entrando com o valor de *resposta*.

[2] Pode ser usado um valor negativo para *passo*. Nesse caso, o laço termina quando o índice é menor que o valor final.

Embora esse algoritmo funcione em certos casos, ele não é infalível. Dependendo dos valores dos coeficientes, o algoritmo pode não funcionar. Eis o que acontece:

- Se $a = 0$, surge imediatamente um problema por causa da divisão por zero. De fato, uma inspeção mais cuidadosa da Equação (2.1) indica que dois casos podem acontecer. Isto é,
  Se $b \neq 0$, a equação se reduz à equação linear com uma raiz real, $-c/b$.
  Se $b = 0$, então não existe solução. Isto é, o problema é trivial.
- Se $a \neq 0$, ocorrem dois casos possíveis, dependendo do valor do discriminante, $d = b^2 - 4ac$. Isto é,
  Se $d \geq 0$, então ocorrem duas raízes reais.
  Se $d < 0$, ocorrem duas raízes complexas.

Observe que usamos identação para destacar a estrutura decisória por trás da matemática. Essa estrutura se traduz prontamente no conjunto de estruturas IF/THEN/ELSE acopladas que podem ser inseridas no lugar das instruções sombreadas no código anterior para dar o algoritmo final:

```
DO
  INPUT a, b, c
  r1 = 0: r2 = 0: i1 = 0: i2 = 0
  IF a = 0 THEN
    IF b ≠ 0 THEN
       r1 = −c/b
    ELSE
       DISPLAY "Solução Trivial"
    ENDIF
  ELSE
    discr = b^2 − 4 * a * c
    IF discr ≥ 0 THEN
       r1 = (−b + Sqrt(discr))/(2 * a)
       r2 = (−b − Sqrt(discr))/(2 * a)
    ELSE
       r1 = −b/(2 * a)
       r2 = r1
       i1 = Sqrt(Abs(discr))/(2 * a)
       i2 = −i1
    ENDIF
  ENDIF
  DISPLAY r1, r2, i1, i2
  DISPLAY 'Tentar novamente? Responda sim ou não'
  INPUT resposta
  IF resposta = 'não' EXIT
ENDDO
```

A abordagem do exemplo anterior pode ser empregada para desenvolver um algoritmo para o problema do pára-quedista. Lembre-se de que, dada a condição inicial para o tempo e a velocidade, o problema envolve a solução iterativa da fórmula

$$v_{i+1} = v_i + \frac{dv_i}{dt}\Delta t \tag{2.2}$$

Lembre-se também de que, se desejarmos atingir uma boa acurácia, precisaremos empregar passos pequenos. Portanto, provavelmente iremos querer aplicar a fórmula repetidamente do instante inicial até o instante final. Conseqüentemente, um algoritmo para resolver o problema se baseará em um laço.

Por exemplo, suponhamos que o cálculo se inicie em $t = 0$ e queremos predizer a velocidade em $t = 4$s usando o passo no tempo $\Delta t = 0{,}5$s. Precisaremos, portanto, aplicar a Equação (2.2) oito vezes, isto é,

$$n = \frac{4}{0{,}5} = 8$$

onde n é o número de iterações do **laço**. Como este resultado é exato, isto é, a razão é um inteiro, podemos controlar o *laço* base desse algoritmo por um contador. Eis um exemplo do pseudocódigo:

```
g = 9,8
INPUT cd, m
INPUT ti, vi, tf, dt
t = ti
v = vi
n = (tf − ti) / dt
DOFOR i = 1 TO n
  dvdt = g − (cd / m) * v
  v = v + dvdt * dt
  t = t + dt
ENDDO
DISPLAY v
```

Embora esse esquema seja simples de programar, não é infalível. Particularmente, funcionará somente se o intervalo computacional for divisível exatamente pelo passo no tempo.[3] Para cobrir esses casos, um laço de decisão pode substituir a área sombreada no pseudocódigo anterior. O resultado final é

```
g = 9,8
INPUT cd, m
INPUT ti, vi, tf, dt
t = ti
v = vi
h = dt
DO
  IF t + dt > tf THEN
     h = tf − t
  ENDIF
  dvdt = g − (cd / m) * v
  v = v + dvdt * h
  t = t + h
  IF t ≥ tf EXIT
ENDDO
DISPLAY v
```

Assim que entramos no laço, usamos uma estrutura IF/THEN para testar se a adição $t + dt$ nos levará além do fim do intervalo. Se não, que será geralmente o caso no início, não fazemos nada. Em caso afirmativo, precisaremos reduzir o intervalo fazendo a variável h igual a $tf - t$. Procedendo assim, o laço terminará porque a condição $t \geq tf$ se tornará verdadeira.

Note que, antes de entrar no laço, atribuímos o valor do passo no tempo, dt, a outra variável, h. Criamos essa variável auxiliar de modo que nossa rotina não mude o valor

[3] O problema consiste em que os computadores usam uma representação numérica binária para sua matemática interna. Conseqüentemente, uma aparente divisão exata não retorna um inteiro quando essa divisão é implementada no computador. Trataremos desse assunto no Capítulo 3.

dado de *dt* se e quando reduzirmos o passo de tempo. Fazemos isso para antecipar o fato de precisarmos usar o valor original de *dt* em algum outro lugar, no caso de esse código ser integrado a um programa maior.

É possível notar que o algoritmo também não é infalível. O usuário pode ter errado ao entrar um tamanho de passo maior que o cálculo do intervalo, por exemplo, $tf - ti = 5$ e $dt = 20$. Assim, você pode querer incluir a verificação de erro no código para detectar tais erros e então permitir que o usuário corrija o engano.

2.3 PROGRAMAÇÃO MODULAR

Seria bastante difícil estudar um texto que não tivesse capítulos, seções ou parágrafos. Dividir tarefas ou objetivos em partes mais simples é uma maneira de torná-los mais fáceis de tratar. Com esse espírito, programas de computador são divididos em subprogramas menores, ou módulos, que podem ser desenvolvidos e testados separadamente. Essa abordagem é chamada *programação modular*.

O atributo mais importante dos módulos é que sejam tão independentes e autocontidos quanto possível. Além disso, eles são tipicamente desevolvidos para desempenhar uma função específica e bem definida, e têm apenas um ponto de entrada e um ponto de saída. Assim, esses módulos são usualmente pequenos (geralmente com 50 a 100 instruções de tamanho) e altamente focados.

Nas linguagens-padrão de alto nível, como o Fortran 90 ou C, o elemento principal de programação usado para representar cada módulo é o procedimento. Um procedimento é uma série de instruções computacionais que, juntas, desempenham uma dada tarefa. Dois tipos de procedimento são empregados comumente: *funções* e *sub-rotinas*. As primeiras usualmente retornam um resultado único, enquanto que as últimas retornam vários resultados.

Além disso, deve ser mencionado que muitos programas relacionados a pacotes de software, como Excel e MATLAB, envolvem o desenvolvimento de subprogramas. Assim, macros do Excel e funções do MATLAB são desevolvidas para receber uma informação, fazer cálculos e retornar um resultado. Logo, pensar modularmente é também consistente com a maneira como os programas são implementados em ambientes de software.

A programação modular tem diversas vantagens. O uso de unidades menores e autocontidas torna a lógica subjacente mais fácil de seguir e entender, tanto para o desenvolvedor quanto para o usuário. O desenvolvimento é facilitado porque cada módulo pode ser feito isoladamente. De fato, para projetos de grande porte, diferentes programadores podem trabalhar individualmente nas partes. O projeto modular também aumenta a facilidade com a qual um programa pode ser testado e corrigido, porque os erros são isolados mais facilmente. Finalmente, a modificação e a manutenção dos programas são facilitadas, devido ao fato de que novos módulos podem ser desenvolvidos para desempenhar tarefas adicionais e facilmente inseridos em um esquema já coerente e organizado.

Embora todos esses atributos sejam razão suficiente para usar módulos, a razão mais importante, relacionada à resolução de problemas numéricos em engenharia, é que eles permitem que você mantenha sua própria biblioteca de módulos úteis para uso posterior em outros programas. Esta será a filosofia deste livro: todos os algoritmos serão apresentados como módulos.

Tal abordagem é ilustrada na Figura 2.7, que mostra uma função desenvolvida para implementar o método de Euler. Note que a aplicação dessa função difere das versões anteriores em como se manipula a entrada/saída. Nas versões anteriores, entrada e saída vinham diretamente do usuário (via atribuições de entrada ou comandos DISPLAY). Na função, as entradas são inseridas na FUNCTION via uma lista de argumentos.

```
Function Euler(dt, ti, tf, yi)
```

e as saídas são retornadas via uma instrução de atribuição

```
y = Euler(dt, ti, tf, yi)
```

Além disso, observe quão genérica a rotina se tornou. Não existe nenhuma referência às especificidades do problema do pára-quedista. Por exemplo, em vez de chamar a variável dependente v de velocidade, um símbolo mais genérico, y, é usado na função.

```
FUNCTION Euler(dt, ti, tf, yi)
t = ti
y = yi
h = dt
DO
  IF t + dt > tf THEN
    h = tf − t
  ENDIF
  dydt = dy(t, y)
  y = y + dydt * h
  t = t + h
  IF t ≥ tf EXIT
ENDDO
Euler = y
END
```

FIGURA 2.7
Pseudocódigo para a função que resolve a equação diferencial usando o método de Euler.

Mas antes, observe que a derivada não é calculada dentro da função por uma equação explícita – outra função, *dy*, deve ser invocada para calculá-la, pois poderemos querer usar essa função em muitos problemas diferentes, além de aplicá-la para resolver o problema da velocidade do pára-quedista.

2.4 EXCEL

Excel é uma planilha produzida pela Microsoft Inc. Planilhas são um tipo especial de software matemático que permite ao usuário entrar com linhas e colunas de dados e fazer cálculos sobre elas. Assim, são uma versão computadorizada de uma grande folha contábil na qual cálculos extensos e interconectados podem ser mostrados. Todo o cálculo é atualizado quando um valor é modificado, e por isso planilhas são ideais para o tipo de análise "e se?".

O Excel tem alguns recursos numéricos internos, incluindo a resolução de equações, traçado de curvas e otimização. Está também incluída a VBA como uma linguagem de macros que pode ser usada para implementar cálculos numéricos. Finalmente, o Excel possui várias ferramentas de visualização, como gráficos e desenhos de superfícies tridimensionais, que servem como um valioso acessório para a análise numérica. Nesta seção, mostraremos como tais recursos podem ser usadas para resolver o problema do pára-quedista.

Para isso, vamos primeiro preparar uma planilha simples. Como mostrado a seguir, o primeiro passo envolve a inserção de legendas e números nas células da planilha.

	A	B	C	D
1	**Problema do Pára-quedista**			
2				
3	**m**	**68.1**	**kg**	
4	**cd**	**12.5**	**kg/s**	
5	**dt**	**0.1**	**s**	
6				
7	**t**	**vnum (m/s)**	**vanal (m/s)**	
8	**0**	**0.000**		
9	**2**			

Antes de escrever um programa macro para calcular o valor numérico, podemos tornar nosso trabalho subseqüente mais fácil associando nomes aos valores dos parâmetros. Para isso, selecionamos as células A3:B5 (a maneira mais fácil de fazê-lo é movendo

o mouse para A3, pressionando o botão esquerdo e arrastando até B5). Em seguida, vamos ao cabeçalho e selecionamos, nesta ordem,

```
Insert Name Create Left column OK
```

Para constatarmos que tudo funcionou adequadamente, selecionamos a célula B3 e verificamos se um rótulo "m" aparece na caixa de nomes (localizada do lado esquerdo da folha logo abaixo da barra de instruções).

Movemos para a célula C8 e inserimos a solução analítica (Equação 1.9),

```
=9,8*m/cd*(1-exp(-cd/m*A8))
```

Quando essa fórmula é introduzida, o valor 0 aparece em C8. Então, copiamos a fórmula logo abaixo, na célula C9, e vemos que o valor é 16,405 m/s.

Isso que foi dito é típico do uso padrão do Excel. Por exemplo, nesse ponto é possível mudar o valor dos parâmetros e ver como a solução analítica muda.

Agora ilustraremos como macros VBA podem ser usadas para estender os recursos padrão. A Figura 2.8 lista o pseudocódigo lado a lado com o código Excel VBA para todas as estruturas de controle descritas na seção anterior (Figuras 2.2 a 2.6). Observe que, apesar da diferença nos detalhes, a estrutura do pseudocódigo e do código VBA é idêntica.

Podemos agora usar algumas das construções da Figura 2.8 para escrever uma macrofunção para calcular a velocidade numericamente. Abra o VBA selecionando[4]

```
Tools Macro Visual Basic Editor
```

Uma vez dentro do Visual Basic *Editor* (VBE), selecione

```
Insert Module
```

e uma janela de código se abrirá. A seguinte função VBA pode ser desenvolvida diretamente do pseudocódigo na Figura 2.7. Digite esta função na janela de edição.

```
Option Explicit
---------------------------------------------------------------
Function Euler(dt, ti, tf, yi, m, cd)

Dim h As Single, t As Single, y As Single, dydt As Single
t = ti
y = yi
h = dt
Do
  If t + dt > tf Then
   h = tf - t
  End If
  dydt = dy(t, y, m, cd)
  y = y + dydt * h
  t = t + h
  If t >= tf Then Exit Do
Loop
Euler = y
End Function
```

Compare essa macro com o pseudocódigo da Figura 2.7 e verifique quão similares eles são. Observe também como expandimos a lista dos argumentos da função para incluir os parâmetros necessários para o modelo da velocidade do pára-quedista. A velocidade resultante, v, é então devolvida para a planilha via o nome da função.

[4] Existe uma combinação de teclas Alt-F11 que é ainda mais rápida!

Note ainda como incluímos outra função para calcular a derivada, que pode ser incluída no mesmo módulo digitando-se diretamente a seguir da função Euler:

```
Function dy(t, v, m, cd)
Const g As Single = 9,8
dy = g - (cd / m) * v
End Function
```

FIGURA 2.8
Estruturas de controle fundamentais em (*a*) pseudocódigo e (*b*) Excel VBA.

(*a*) **Pseudocódigo**	(*b*) **Excel VBA**
IF/THEN:	
IF condição THEN	`If b <> 0 Then`
True bloco	`  r1 = -c / b`
ENDIF	`End If`
IF/THEN/ELSE:	
IF condição THEN	`If a < 0 Then`
True bloco	`  b = Sqr(Abs(a))`
ELSE	`Else`
False bloco	`  b = Sqr(a)`
ENDIF	`End If`
IF/THEN/ELSEIF:	
IF condição$_1$ THEN	`If class = 1 Then`
Bloco$_1$	`  x = x + 8`
ELSEIF condição$_2$	`ElseIf class < 1 Then`
Bloco$_2$	`  x = x - 8`
ELSEIF condição$_3$	`ElseIf class < 10 Then`
Bloco$_3$	`  x = x - 32`
ELSE	`Else`
Bloco$_4$	`  x = x - 64`
ENDIF	`End If`
CASE:	
SELECT CASE Expressão de teste	`Select Case a + b`
CASE Valor$_1$	`  Case Is < -50`
Bloco$_1$	`    x = -5`
CASE Valor$_2$	`  Case Is < 0`
Bloco$_2$	`    x = -5 - (a + b) / 10`
CASE Valor$_3$	`  Case Is < 50`
Bloco$_3$	`    x = (a + b) / 10`
CASE ELSE	`  Case Else`
Bloco$_4$	`    x = 5`
END SELECT	`End Select`
DOEXIT:	
DO	`Do`
Bloco$_1$	`  i = i + 1`
IF condição EXIT	`  If i >= 10 Then Exit Do`
Bloco$_2$	`  j = i*x`
ENDIF	`Loop`
COUNT-CONTROLLED LOOP:	
DOFOR i = início, fim, passo	`For i = 1 To 10 Step 2`
Bloco	`  x = x + i`
ENDDO	`Next i`

O passo final é retornar à planilha e invocar a função entrando a seguinte fórmula na célula B9:

```
=Euler(dt,A8,A9,B8,m,cd)
```

O resultado da integração numérica, 16,531, aparecerá na célula B9.

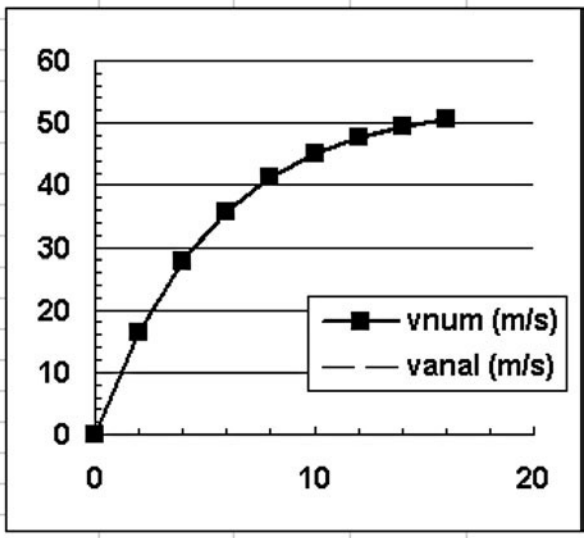
É possível avaliar o que aconteceu aqui. Quando você entra com o código da função na célula da planilha, os parâmetros são repassados para o programa VBA onde o cálculo é feito, e o resultado então é enviado para a célula no monitor. De fato, a linguagem VBA da macro permite que se use o Excel como um mecanismo de entrada e saída. Vários benefícios advêm desse fato.

Por exemplo, agora que você preparou os cálculos, pode manipulá-los. Suponha que o saltador era muito pesado, digamos, $m = 100$ kg (cerca de 220 libras). Inserir 100 na célula B3 da planilha atualizará o valor apresentado para 17,438 na célula B9. Mude a massa de volta para 68,1 kg e o resultado anterior, 16,531, automaticamente reaparecerá na célula B9.

Agora vamos processar um passo adiante preenchendo mais alguns números para o tempo. Insira os números 4, 6,... 16 nas células A10 a A16. Agora copie as fórmulas das células B9:C9 para as linhas de 10 a 16. Note que o programa VBA calcula os resultados numéricos corretamente para cada nova linha. (Para verificar isso, mude o *dt* para 2 e compare com o resultado previamente computado no Exemplo 1.2). Como um desenvolvimento adicional, pode ser feito um gráfico *xy* com os resultados, usando-se o Excel Chart Wizard.

A planilha final é mostrada a seguir. Acabamos de criar uma ferramenta interessante para solução de problemas. Você pode fazer análises de sensibilidade mudando os valores de cada um dos parâmetros. Assim que um novo valor for inserido, a computação e o gráfico serão atualizados automaticamente. Essa natureza interativa faz do Excel uma ferramenta poderosa. No entanto, a habilidade para resolver esses problemas repousa em poder escrever as macros em VBA.

	A	B	C
1	**Problema do Pára-quedista**		
2			
3	**m**	**68.1**	**kg**
4	**cd**	**12.5**	**kg/s**
5	**dt**	**0.1**	**s**
6			
7	**t**	**vnum (m/s)**	**vanal (m/s)**
8	**0**	**0.000**	**0.000**
9	**2**	**16.531**	**16.405**
10	**4**	**27.943**	**27.769**
11	**6**	**35.822**	**35.642**
12	**8**	**41.262**	**41.095**
13	**10**	**45.017**	**44.873**
14	**12**	**47.610**	**47.490**
15	**14**	**49.400**	**49.303**
16	**16**	**50.635**	**50.559**
17			
18			



É a combinação do Excel com a linguagem de programação VBA que verdadeiramente abre um mundo de possibilidades para a solução de problemas em engenharia. Nos capítulos seguintes, ilustraremos como isso pode ser feito.

2.5 MATLAB

MATLAB é o produto carro-chefe da Math Works Inc., que foi fundada pelos analistas numéricos Cleve Moler e John N. Little. Como o nome diz, MATLAB foi originalmente desenvolvido como um *lab*oratório para *mat*rizes. Atualmente, o elemento central do MATLAB ainda é matriz. Manipulações matemáticas de matrizes são muito convenientemente implementadas em ambientes interativos fáceis de usar. A essas manipulações de matrizes, o MATLAB adicionou diversas funções numéricas, cálculos simbólicos e ferramentas de visualização. Como conseqüência, a versão corrente representa um ambiente técnico computacional quase completo.

O MATLAB possui uma variedade de funções e operadores que permitem implementar convenientemente muitos dos métodos numéricos desenvolvidos neste livro. Tais métodos serão descritos em detalhes nos capítulos que seguem. Além disso, programas podem ser escritos nos chamados *M-files*, que podem ser usados para implementar cálculos numéricos. Vamos explorar como isso é feito.

Primeiro, você pode ver que o uso normal do MATLAB está intimamente relacionado à programação. Por exemplo, suponha que nós queremos determinar a solução analítica para o problema do pára-quedista, o que pode ser obtido com a seguinte série de comandos MATLAB:

```
>> g=9,8;
>> m=68,1;
>> cd=12,5;
>> tf=2;
>> v=g*m/cd*(1-exp(-cd/m*tf))
```

cujo resultado é mostrado como

```
v =
   16,4050
```

Portanto, a seqüência de comandos é semelhante à seqüência de instruções em uma linguagem de programação típica.

Agora, digamos que você queira se desviar da estrutura seqüencial. Embora existam algumas maneiras claras de inserir funcionalidades não-seqüenciais no modo padrão de comando, a inclusão de decisões e **laços** é mais bem implementada criando-se um documento MATLAB chamado M-file. Para isso, faça a seguinte seleção na barra de instruções

```
File New Mfile
```

e uma nova janela se abrirá com o cabeçalho "MATLAB Editor/Debugger". Nessa janela, você pode digitar e editar programas MATLAB. Digite o seguinte código:

```
g=9,8;
m=68,1;
cd=12,5;
tf=2;
v=g*m/cd*(1-exp(-cd/m*tf))
```

Observe que os comandos são escritos da mesma maneira como seriam escritos na interface do MATLAB. Salve o programa com o nome analpara. O MATLAB automaticamente agregará a extensão .m para denotá-lo como um M-file: analpara.m.

Para executar esse programa, você deve voltar ao modo de comando. A maneira mais direta de fazê-lo é clicar no botão "MATLAB Command Window" na barra de tarefas (a qual está usualmente na parte inferior do monitor).

O programa pode ser executado digitando-se o nome do M-file, analpara, como mostrado a seguir:

```
>> analpara
```

Se você fez tudo corretamente, o MATLAB deverá responder com a resposta correta:

```
v =
   16,4050
```

O problema, agora, é que tudo foi configurado para rodar um caso somente. Você pode tornar o programa mais flexível fazendo o usuário acionar mais algumas variáveis de entrada. Por exemplo, suponha que você deseja avaliar o impacto da massa na velocidade em 2 s. O M-file pode ser reescrito da seguinte forma para atender a essa análise:

```
g=9,8;
m=input('massa (kg):');
cd=12,5;
tf=2;
v=g*m/cd*(1-exp(-cd/m*tf))
```

Salve esse programa como analpara2.m. Se você digitou analpara2 enquanto estava no modo comando, o prompt deve aparecer como

```
mass (kg):
```

O usuário pode então entrar com um valor como 100, e o resultado será mostrado como

```
v =
   17,3420
```

Agora está bastante claro como podemos programar uma solução numérica com um M-file. Para fazer isso, devemos primeiro entender como MATLAB manipula logicamente as estruturas lógicas e de laço. A Figura 2.9 traz uma lista de pseudocódigos lado a lado com o código MATLAB para todas as estruturas de controle da seção anterior. Embora as estruturas do pseudocódigo e do código MATLAB sejam muito parecidas, existem algumas pequenas diferenças que devem ser notadas.

Em particular, observe como representamos a estrutura DOEXIT. No lugar do DO, usamos a instrução WHILE(1). MATLAB interpreta o número 1 como correspondendo a "verdadeiro", portanto, essa instrução será repetida infinitamente da mesma maneira como a instrução DO. O laço terminará com um comando *break*. Esse comando transfere o controle para a instrução seguinte à instrução *end* que finaliza o laço.

Observe também que os parâmetros do laço controlado por contador são ordenados de forma diferente. Para o pseudocódigo, os parâmetros do laço são especificados como *início, fim, passo.* Para o MATLAB, os parâmetros são ordenados como *início*:passo:fim.

O seguinte M-file MATLAB pode ser desenvolvido do pseudocódigo na Figura 2.7. Digite-o no MATLAB Editor/Debugger:

```
g=9,8;
m=input('massa (kg):');
cd=12,5;
ti=0;
tf=2;
vi=0;
dt=0,1;
t = ti;
v = vi;
h = dt;
while (1)
  if t + dt > tf
    h = tf - t;
  end
  dvdt = g - (cd / m) * v;
  v = v + dvdt * h;
  t = t + h;
  if t >= tf, break, end
end
disp('velocidade (m/s):')
disp(v)
```

(*a*) **Pseudocódigo**	(*b*) **MATLAB**
IF/THEN:	
IF condição THEN	`if b ~= 0`
True bloco	`  r1 = -c / b;`
ENDIF	`end`
IF/THEN/ELSE:	
IF condição THEN	`if a < 0`
True bloco	`  b = sqrt(abs(a));`
ELSE	`else`
False bloco	`  b = sqrt(a);`
ENDIF	`end`
IF/THEN/ELSEIF:	
IF condição$_1$ THEN	`if class == 1`
Bloco$_1$	`  x = x + 8;`
ELSEIF condição$_2$	`elseif class < 1`
Bloco$_2$	`  x = x - 8;`
ELSEIF condição$_3$	`elseif class < 10`
Bloco$_3$	`  x = x - 32;`
ELSE	`else`
Bloco$_4$	`  x = x - 64;`
ENDIF	`end`
CASE:	
SELECT CASE Expressão de Teste	`switch a + b`
CASE Valor$_1$	`  case 1`
Bloco$_1$	`    x = -5;`
CASE Valor$_2$	`  case 2`
Bloco$_2$	`    x = -5 - (a + b) / 10;`
CASE Valor $_3$	`  case 3`
Bloco$_3$	`    x = (a + b) / 10;`
CASE ELSE	`  otherwise`
Bloco$_4$	`    x = 5;`
END SELECT	`end`
DOEXIT:	
DO	`while (1)`
Bloco$_1$	`  i = i + 1;`
IF condição EXIT	`  if i >= 10, break, end`
Bloco$_2$	`  j = i*x;`
ENDIF	`end`
COUNT-CONTROLLED LOOP:	
DOFOR i = start, finish, step	`for i = 1:2:10`
Bloco	`  x = x + i;`
ENDDO	`end`

FIGURA 2.9
Estruturas de controle fundamentais em (*a*) pseudocódigo e (*b*) linguagem de programação MATLAB.

Salve esse programa como numpara.m e retorne para o modo comando para executá-lo, digitando numpara. Isso deve resultar na seguinte saída:

```
massa (kg): 100

velocidade (m/s):
   17,4381
```

Como passo final nesse desenvolvimento, vamos tomar o M-file e convertê-lo na função apropriada, o que resulta no seguinte M-file, baseado no pseudocódigo da Figura 2.7:

```
function yy = euler(dt,ti,tf,yi,m,cd)
t = ti;
y = yi;
h = dt;
while (1)
  if t + dt > tf
    h = tf - t;
  end
  dydt = dy(t, y, m, cd);
  y = y + dydt * h;
  t = t + h;
  if t >= tf, break, end
end
yy = y;
```

Salve essa função como euler.m e crie outro M-file para calcular a derivada,

```
function dydt = dy(t, v, m, cd)
g = 9,8;
dydt = g - (cd / m) * v;
```

Salve esse arquivo como dy.m e retorne ao modo comando. Para invocar a função e ver o resultado, você pode digitar os seguintes comandos:

```
>> m=68,1;
>> cd=12,5;
>> ti=0;
>> tf=2.;
>> vi=0;
>> dt=0.1;
>> euler(dt,ti,tf,vi,m,cd)
```

Quando o último comando for inserido, a resposta será mostrada como

```
ans =
16,5309
```

Essa é a combinação do ambiente MATLAB com a linguagem de programação dos M-file que de fato abre um mundo de possibilidades para a resolução de problemas de engenharia. No próximo capítulo, ilustraremos como isso é feito.

2.6 OUTRAS LINGUAGENS E BIBLIOTECAS

Nas seções anteriores, mostramos como desenvolver funções e procedimentos em Excel e MATLAB para implementar o método de Euler partindo de algoritmos expressos em pseudocódigo. Você deve notar que funções similares podem ser escritas em linguagens de alto nível como Fortran 90 e C++. Por exemplo, uma função em Fortran 90 para o método de Euler seria

```
Function Euler(dt, ti, tf, yi, m, cd)

REAL dt, ti, tf, yi, m, cd
Real h, t, y, dydt
```

```
t = ti
y = yi
h = dt
Do
  If (t + dt > tf) Then
    h = tf - t
  End If
  dydt = dy(t, y, m, cd)
  y = y + dydt * h
  t = t + h
  If (t >= tf) Exit
End Do
Euler = y
End Function
```

Para C, o resultado seria similar ao da função MATLAB. O ponto é que, uma vez tendo desenvolvido um algoritmo bem estruturado na forma de pseudocódigo, ele pode ser prontamente implementado em uma variedade de ambientes de programação.

Neste livro, nossa abordagem será a de muní-lo com procedimentos bem estruturados escritos em pseudocódigo. Essa coleção de algoritmos constitui uma biblioteca numérica que pode ser acessada para implementar tarefas numéricas específicas nem diversas ferramentas de software ou linguagens de programação.

Além dos seus programas, você deve estar ciente que bibliotecas de programas comerciais também trazem procedimentos numéricos úteis. Por exemplo, a biblioteca *Numerical Recipe* inclui uma grande variedade de algoritmos escritos em Fortran e C.[5] Tais procedimentos são descritos em livros (por exemplo, Press et al., 1992) e em formato eletrônico.

Para Fortran, o *IMSL* (*I*nternational *M*athematical and *S*tatistical *L*ibrary) provê mais de 700 procedimentos cobrindo todas as áreas numéricas mencionadas neste livro. Em razão do uso bem difundido do Fortran em engenharia, incluímos aplicações IMSL neste texto.

PROBLEMAS

2.1 Escreva o pseudocódigo para implementar o fluxograma mostrado na Figura P2.1. Certifique-se de incluir a identação apropriada para deixar o programa bem claro.

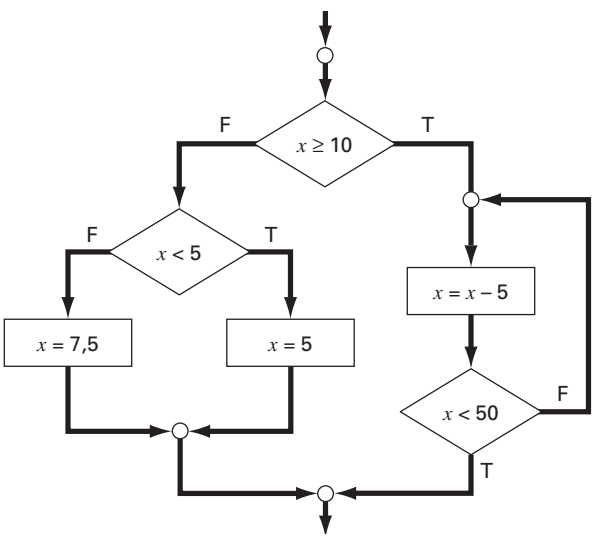


Figura P2.1

2.2 Reescreva o seguinte pseudocódigo usando a identação apropriada.

```
DO
i = i + 1
IF z > 50 EXIT
x = x + 5
IF x > 5 THEN
y = x
ELSE
y = 0
ENDIF
z = x + y
ENDDO
```

2.3 Um valor para a concentração de poluentes em um lago é registrado em cada cartão em uma seqüência de cartões indexados. Um cartão especial marcado com "fim dos dados" é colocado no final da seqüência. Escreva um algoritmo que determine a soma, a média e o máximo desses valores.

2.4 Escreva um fluxograma estruturado para o Problema 2.3.

2.5 Desenvolva, corrija e documente um programa para determinar as raízes de uma equação quadrática, $ax^2 + bx + c$, em uma linguagem de programação de alto nível ou uma linguagem de macro

[5] Os procedimentos do Numerical Recipe estão disponíveis em livro e em formato eletrônico para Pascal, MS BASIC e MATLAB. Informações sobre os produtos Numerical Recipe podem ser encontradas em http://www.nr.com/.

da sua escolha. Use sub-rotinas para calcular as raízes (sejam reais ou complexas). Faça testes de execução para os casos **(a)** $a = 1$, $b = 6$, $c = 2$; **(b)** $a = 0$, $b = -4$, $c = 1{,}6$; **(c)** $a = 3$, $b = 2{,}5$, $c = 7$;

2.6 A função cosseno pode ser calculada pela seguinte série infinita:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

Escreva um algoritmo para implementar essa fórmula de modo a calcular e imprimir os valores de cos x assim que cada termo da série for acrescentado. Em outras palavras, calcule e imprima a seqüência de valores para

$$\cos x = 1$$
$$\cos x = 1 - \frac{x^2}{2!}$$
$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!}$$

até o termo de ordem n escolhido. Para cada valor especificado, calcule e mostre o erro relativo porcentual como

$$\%\ \text{erro} = \frac{\text{valor real} - \text{aproximação}}{\text{valor real}} \times 100\%$$

2.7 Escreva um algoritmo para o Problema 2.6 como (**a**) um fluxograma estruturado, (**b**) pseudocódigo.

2.8 Desenvolva, corrija e documente um programa para o Problema 2.6 em uma linguagem de programação de alto nível ou em uma linguagem de macro de sua escolha. Empregue funções de biblioteca para o cosseno em seu computador para determinar os valores reais. Faça o programa imprimir as aproximações da série e o erro em cada etapa. Como um teste, faça o programa calcular o cos(1,25) até o termo $x^{10}/10!$ Interprete os resultados.

2.9 O algoritmo seguinte foi desenvolvido para determinar a nota para um curso cuja avaliação consiste em testes, exercícios para casa e exames:

Passo 1: Entre com o número e o nome do curso.
Passo 2: Entre com os fatores de peso para os testes (PT), exercício para casa (PE) e para o exame final (PEF).
Passo 3: Entre com a nota dos testes e determine a média das notas dos testes (MT).
Passo 4: Entre com a nota dos exercícios e determine a sua média (AH).
Passo 5: Se o curso tem uma nota final, continue para o Passo 6. Se não, vá para o Passo 9.
Passo 6: Entre com a nota do exame final (EF).
Passo 7: Determine o valor de AG segundo a equação

$$\text{AG} = \frac{\text{PT} \times \text{MT} + \text{PE} \times \text{ME} + \text{PEF} \times \text{EF}}{\text{PT} + \text{PE} + \text{PEF}} \times 100\%$$

Passo 8: Vá para o Passo 10.
Passo 9: Determine o valor de AG segundo a fórmula

$$\text{AG} = \frac{\text{PT} \times \text{MT} + \text{PE} \times \text{ME}}{\text{PT} + \text{PE}} \times 100\%$$

Passo 10: Imprima o número do curso, o nome e a média geral.
Passo 11: Termine o programa.

(a) Escreva um pseudocódigo bem estruturado para implementar esse algoritmo.

(b) Escreva, corrija e documente um programa estruturado baseado nesse algoritmo. Teste o programa utilizando os seguintes dados para calcular a nota sem o exame final e a nota com o exame final: PT = 35; PE = 30; PEF = 35; testes = 98, 85, 90, 65, 99; exercício para casa = 95, 90, 87, 100, 92, 77 exame final = 92.

2.10 O método da "divisão e média", um método antigo para calcular a raiz quadrada de um número positivo a, pode ser formulado como

$$x = \frac{x + a/x}{2}$$

(a) Escreva um pseudocódigo bem estruturado para implementar esse algoritmo como mostrado na Figura P2.10. Use a identação apropriada para que a estrutura fique clara.

(b) Desenvolva, corrija e documente o programa que implementa essa equação em uma linguagem de alto nível ou uma linguagem de macro à sua escolha. Estruture seu código como na Figura P2.10.

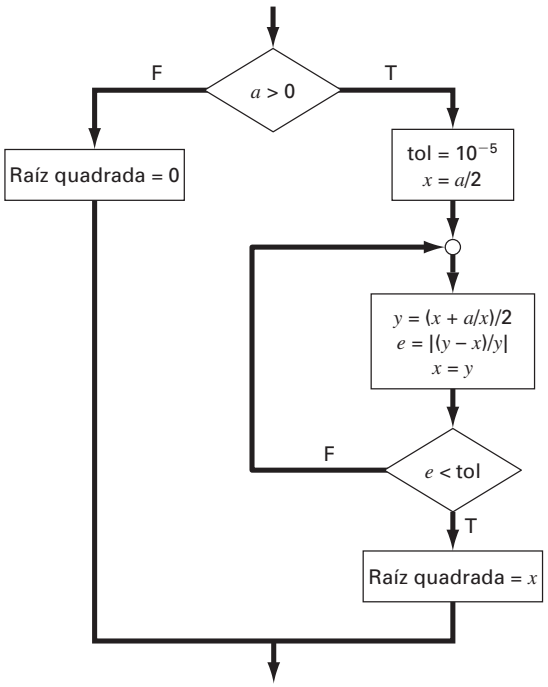


Figura P2.10

2.11 Uma quantidade de dinheiro P é investida em uma conta cujos juros são computados no final do período. O valor de retorno R fornecido por uma taxa de juros i, depois de n períodos, deve ser determinado pela seguinte fórmula:

$$R = P(1 + i)^n$$

Escreva um programa para calcular o valor de retorno do investimento para cada ano de 1 a n. Os argumentos da função devem incluir o investimento inicial P, a taxa de juros i (um valor decimal) e o número de anos n para o qual o valor de retorno deve ser calculado. A saída deve consistir em uma tabela com cabeçalho e colunas para n e R. Execute o programa para $P = \$100.000$, $i = 0{,}06$ e $n = 5$ anos.

2.12 Fórmulas econômicas estão disponíveis para calcular pagamentos anuais para empréstimos. Suponha que você tenha tomado emprestada uma quantidade de dinheiro P e tenha concordado em pagar em n pagamentos anuais com uma taxa de juros i. A fórmula para calcular o pagamento anual A é

$$A = P\frac{i(1+i)^n}{(1+i)^n - 1}$$

Escreva um programa que calcule A. Teste o programa para P = \$55.000 e uma taxa de juros de 6,6% (i = 0,066). Calcule os resultados para n = 1, 2, 3, 4 e 5 anos e mostre o resultado em uma tabela com cabeçalho e colunas para n e A.

2.13 A temperatura média por dia de uma determinada área pode ser aproximada pela seguinte função:

$$T = T_{\text{média}} + (T_{\text{máx}} - T_{\text{média}}) \cos(\omega\,(t - t_{\text{máx}}))$$

onde $T_{\text{média}}$ é a temperatura média anual, $T_{\text{máx}}$ é a temperatura máxima, ω = freqüência de variação anual ($= 2\pi/365$), e $t_{\text{máx}}$ são os dias de temperatura máxima ($\cong$ 205 d). Desenvolva um programa que calcule a temperatura média entre dois dias do ano para uma cidade particular. Teste o programa para (**a**) janeiro-fevereiro (t = 0 a 59) em Miami, Flórida ($T_{\text{média}}$ = 22,1ºC; $T_{\text{máx}}$ = 28,3ºC), e (**b**) julho-agosto (t = 180 a 242) em Boston, Massachusetts ($T_{\text{média}}$ = 10,7ºC; $T_{\text{máx}}$ = 22,9ºC).

2.14 Desenvolva, corrija e teste um programa em uma linguagem de alto nível ou em uma linguagem de macro à sua escolha para calcular a velocidade de queda de um pára-quedista como mostrado no Exemplo 1.2. Desenvolva o programa de modo a permitir que o usuário entre com valores para o coeficiente de arrasto e a massa. Teste o programa duplicando os resultados do Exemplo 1.2. Repita os cálculos mas use tamanho de passo 1 a 0,5 s. Compare os resultados com a solução analítica obtida previamente no Exemplo 1.1. Um menor passo tem resultados melhores ou piores? Explique os resultados.

2.15 A *bubble sort* é uma técnica de ordenação ineficiente, porém fácil de programar. A idéia por trás desse algoritmo é mover por um vetor comparando pares adjacentes e trocando os valores se estes estiverem fora de ordem. Seguindo esse método, para ordenar um vetor completamente, é necessário repetir os passos muitas vezes. A cada passo, para uma ordenação desconto, os elementos menores do vetor parecem subir para o fim dos vetores como bolhas. Eventualmente, haverá uma passagem pelo vetor em que nenhuma troca é necessária. Então, o vetor estará ordenado. Depois da primeira passada, o maior valor do *vetor* vai diretamente para o vetor fundo. Conseqüentemente, a segunda passada tem de posicionar o penúltimo valor, e assim por diante. Desenvolva um programa para montar um vetor com 20 números aleatórios e ordená-los em ordem decrescente com o *bubble sort* (Figura P2.15).

2.16 A Figura P2.16 mostra um tanque cilíndrico de base cônica. Se o nível de líquido é muito baixo na parte cônica, o volume é simplesmente o volume cônico de líquido. Se o nível de líquido está no meio da parte cilíndrica, o volume total de líquido inclui a parte cônica preenchida e a parte parcialmente ocupada do cilindro. Escreva uma função bem estruturada para calcular o volume do tanque como função dos valores dados de R e d. Use estruturas de controle e decisão (como If/Then, ElseIf, Else, EndIf). Desenvolva a função de modo que esta retorne o volume para todos os casos em que a profundidade é menor que 3R. Retorne uma mensagem de erro ("Transbordo") se você transbordou o tanque, isto é, $d > 3R$. Teste a função com os dados seguintes:

R	1	1	1	1
d	0,5	1,2	3,0	3,1

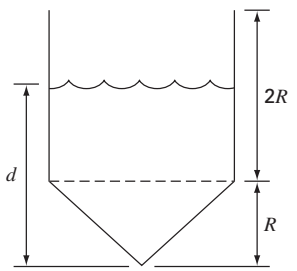


Figura P2.16

2.17 Duas distâncias são necessários para especificar a localização de um ponto relativamente à origem no espaço bidimensional (Figura P2.17):

- As distâncias horizontal e vertical (x, y) em coordenadas cartesianas.
- O raio e o ângulo (r, θ) em coordenadas polares.

Figura P2.15

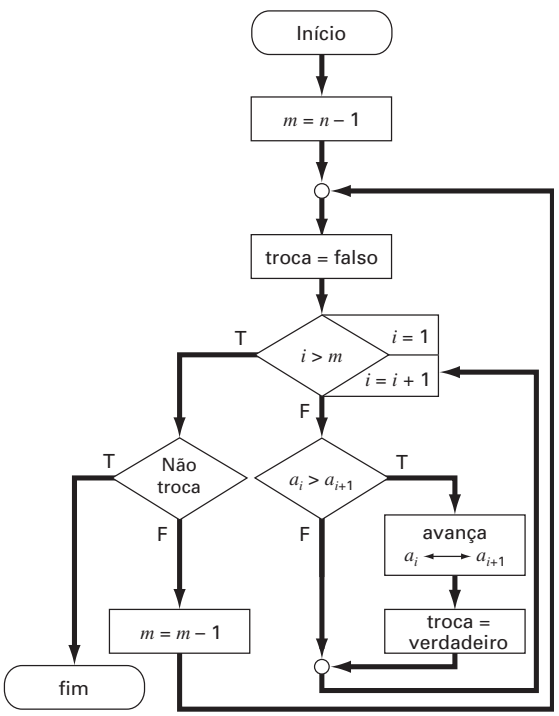
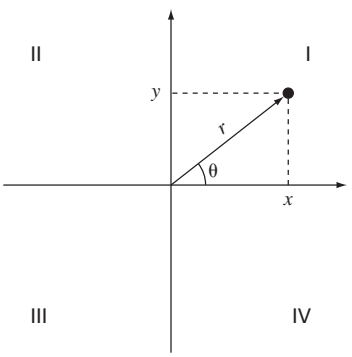


Figura P2.17



É relativamente simples calcular as coordenadas cartesianas (x, y) com base das coordenadas polares (r, θ). O inverso não é tão simples. O raio pode ser calculado pela seguinte fórmula:

$$r = \sqrt{x^2 + y^2}$$

Se as coordenadas estiverem no primeiro e no quarto quadrantes (isto é, $x > 0$), então uma fórmula simples pode ser utilizada,

$$\theta = \text{tg}^{-1}\left(\frac{y}{x}\right)$$

A dificuldade surge para os demais casos. A tabela a seguir sintetiza as possibilidades:

x	y	θ
<0	>0	$\tan^{-1}(y/x) + \pi$
<0	<0	$\tan^{-1}(y/x) - \pi$
<0	$=0$	π
$=0$	>0	$\pi/2$
$=0$	<0	$-\pi/2$
$=0$	$=0$	0

(a) Escreva um fluxograma bem estruturado para um procedimento que calcule r e θ como função de x e y. Expresse os resultados finais para θ em graus.

(b) Escreva uma função bem estruturada baseada no seu fluxograma. Teste seu programa usando-o para preencher a seguinte tabela:

x	y	r	θ
1	0		
1	1		
0	1		
−1	1		
−1	0		
−1	−1		
0	−1		
1	−1		
0	0		

2.18 Desenvolva uma função bem estruturada para a qual é passada uma nota em valor numérico de 0 a 100 e que retorna uma nota em caracteres segundo o esquema:

Nota numérica	Critérios
A	90 ≤ valor numérico ≤ 100
B	80 ≤ valor numérico < 90
C	70 ≤ valor numérico < 80
D	60 ≤ valor numérico < 70
F	valor numérico < 60

2.19 Desenvolva uma função bem estruturada que determina (**a**) o fatorial; (**b**) o valor mínimo de um vetor; (**c**) a média dos valores de um vetor.

2.20 Desenvolva um programa bem estruturado para (**a**) determinar a raiz quadrada da soma de quadrados de elementos de uma matriz) e (**b**) normalize a matriz dividindo cada linha pelo máximo valor absoluto da linha, de modo que o elemento máximo em cada linha seja 1.



CAPÍTULO 3

Aproximações e Erros de Arredondamento

Como tantos dos métodos deste livro são simples na descrição e na aplicação, seria muito tentador, neste ponto, proceder diretamente ao corpo principal do texto e ensinar a usar essas técnicas. Entretanto, entender o conceito de erro é tão importante para o uso efetivo dos métodos numéricos que os próximos dois capítulos serão devotados a esse tópico.

A importância do erro foi introduzida na discussão sobre o pára-quedista em queda livre no Capítulo 1, quando a velocidade de um pára-quedista em queda livre foi determinado por métodos tanto analíticos quanto numéricos. Embora a técnica numérica tenha produzido estimativas que estavam próximas da solução analítica exata, havia uma discrepância, ou *erro*, porque o método numérico envolvia uma aproximação. Na realidade, tivemos sorte nesse caso, porque estava disponível uma solução analítica que permitiu calcular o erro exatamente. Para muitos problemas aplicados de engenharia, não é possível obter soluções analíticas. Portanto, não se pode calcular exatamente os erros associados aos métodos numéricos. Nesses casos, é preciso contentar-se com uma aproximação ou estimativa dos erros.

Tais erros são característicos da maioria das técnicas descritas neste livro. Essa afirmação pode, em princípio, parecer contrária ao que normalmente se julga engenharia segura. Os alunos e os engenheiros práticos constantemente tentam limitar os erros em seu trabalho. Quando se fazem provas ou lições de casa, o erro é penalizado, e não recompensado. Na prática profissional, os erros podem ser custosos e, algumas vezes, catastróficos. Se uma estrutura ou um mecanismo falha, vidas podem ser perdidas.

A perfeição é uma meta louvável, mas é raramente atingida. Por exemplo, embora o modelo desenvolvido a partir da segunda lei de Newton seja uma aproximação excelente, ele nunca prediria exatamente, na prática, a queda do pára-quedista. Uma variedade de fatores, como vento e pequenas variações na resistência do ar, causaria desvios do valor previsto. Se esses desvios forem sistematicamente para cima ou para baixo, então é necessário desenvolver um novo modelo. Entretanto, se eles forem distribuídos aleatoriamente e agrupados bem junto em torno da previsão, então os desvios poderão ser considerados desprezíveis, e o modelo, adequado. As aproximações numéricas também introduzem discrepâncias similares na análise. Novamente, a questão é: quanto erro está presente nos cálculos, e ele é tolerável?

Este capítulo e o próximo cobrem os tópicos básicos relacionados à identificação, quantificação e minimização desses erros. Neste capítulo, informações gerais referentes à quantificação do erro são revisadas nas primeiras seções. Isto é seguido por uma seção sobre uma das duas principais formas de erro numérico: o erro de arredondamento. O *erro de arredondamento* é conseqüência do fato de os computadores poderem representar apenas quantidades com números finitos de algoritmos. A seguir, o Capítulo 4 trata da outra forma importante de erro: o erro de truncamento. O *erro de truncamento* é a discrepância introduzida pelo fato de que os métodos numéricos podem empregar aproximações para representar operações e quantidades matemáticas exatas. Finalmente, serão discutidos brevemente os erros que não são ligados diretamente com os próprios métodos numéricos e que incluem enganos, erros na formulação ou no modelo, e incerteza nos dados.

3.1 ALGARISMOS SIGNIFICATIVOS

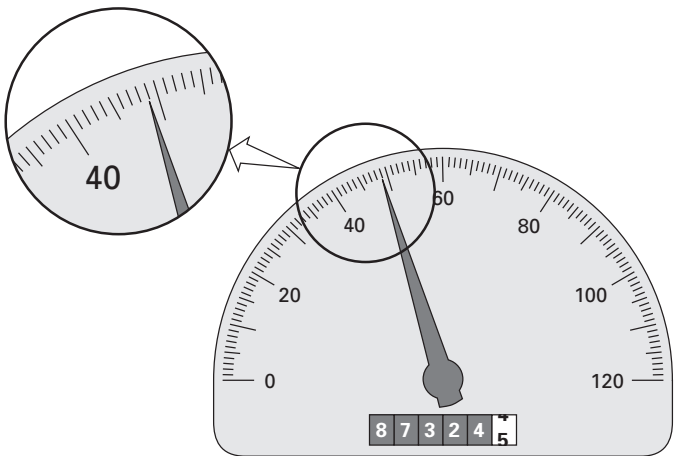
Este livro trata exaustivamente das aproximações ligadas com a manipulação dos números. Conseqüentemente, antes de discutir os erros associados aos métodos numéricos, é útil revisar os conceitos básicos ligados à representação aproximada dos próprios números.

Sempre que se emprega um número em um cálculo, é preciso estar certo de que ele pode ser usado com confiança. Por exemplo, a Figura 3.1 exibe um velocímetro e um hodômetro de um automóvel. A inspeção visual de um velocímetro indica que o carro está movendo-se entre 48 e 49 km/h. Como o indicador está mais alto do que o ponto médio entre os marcadores de escala, pode-se dizer com segurança que o carro está movendo-se a aproximadamente 49 km/h. O resultado é confiável porque, se dois ou mais indivíduos razoáveis lerem essa medida, eles chegarão à mesma conclusão. Entretanto, se a velocidade fosse estimada até uma casa decimal, uma pessoa poderia dizer 48,8, enquanto a outra poderia dizer 48,9 km/h. Mas, por causa dos limites desse instrumento, apenas os dois primeiros algarismos podem ser usados com confiança. Estimativas do terceiro algarismo (ou algarismos superiores) devem ser vistas como aproximações. Seria absurdo afirmar, com base nesse velocímetro, que o automóvel está movendo-se a 48,8642138 km/h. Em contraste, o hodômetro fornece até seis algarismos corretos. A partir da Figura 3.1, pode-se concluir que o carro percorreu um pouco menos que 87.324,5 km durante seu período de uso. Nesse caso, o sétimo algarismo (e os próximos) é incerto.

O conceito de um algarismo significativo foi desenvolvido para designar formalmente a confiabilidade de um valor numérico. Os *algarismos significativos* de um número são aqueles que podem ser usados com confiança. Eles correspondem ao número de algarismos corretos mais um algarismo estimado. Por exemplo, o velocímetro e o hodômetro na Figura 3.1 fornecem leituras de três e sete algarismos significativos, respectivamente. Para o velocímetro, os dois algarismos corretos são 48. Convenciona-se tomar o algarismo estimado como a metade da menor divisão de escala no aparelho de medida. Portanto, a leitura do velocímetro consistiria em três algarismos significativos: 48,5. De forma similar, o hodômetro forneceria uma leitura com sete algarismos significativos, ou seja, 87.324,45.

Embora usualmente seja um procedimento imediato avaliar os algarismos significativos de um número, alguns casos podem causar confusão. Por exemplo, zeros não são sempre algarismos significativos porque eles podem ser necessários apenas para localizar a vírgula decimal. Os números 0,00001845, 0,0001845 e 0,001845 têm quatro algarismos significativos. Analogamente, quando zeros à direita são usados em números grandes, não é claro quantos, ou se algum, destes zeros são significativos. Por exemplo, o valor nominal do número 45.300 pode ter três, quatro ou cinco algarismos significativos, dependendo de os zeros serem conhecidos com confiança. Tais incertezas podem ser resolvidas usando-se a notação científica, onde $4{,}53 \times 10^4$, $4{,}530 \times 10^4$, $4{,}5300 \times 10^4$ designam que o número é conhecido com três, quatro ou cinco algarismos significativos, respectivamente.

FIGURA 3.1
Um velocímetro e um hodômetro de automóvel ilustram o conceito de algarismos significativos.



O conceito de algarismos significativos tem duas implicações importantes no estudo de métodos numéricos:

1. Como introduzido no problema do pára-quedista em queda livre, os métodos numéricos fornecem resultados aproximados. É necessário, portanto, desenvolver critérios para especificar quanta confiança se tem no resultado aproximado. Uma forma de fazer isso é em termos de algarismos significativos. Por exemplo, pode-se decidir que a aproximação é aceitável se ela for correta até quatro algarismos significativos.
2. Embora quantidades como π, e ou $\sqrt{7}$ representem quantidades específicas, elas não podem ser expressas exatamente por um número limitado de algarismos. Por exemplo,

   $$\pi = 3{,}14159265358979323846264 3\ldots$$

   ad infinitum. Como os computadores mantêm apenas um número finito de algarismos significativos, tais números jamais podem ser representados exatamente. A omissão dos algarismos significativos remanescentes é chamada de erro de arredondamento.

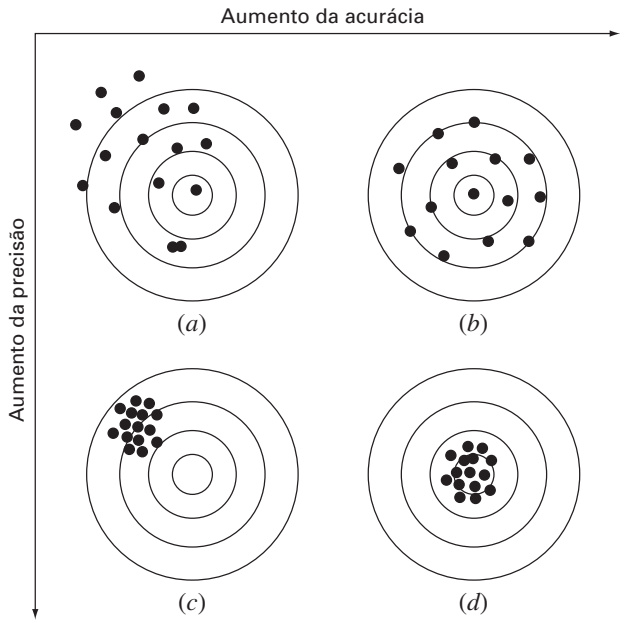
Tanto o erro de arredondamento quanto o uso de algarismos significativos para expressar confiança em um resultado numérico serão explorados em detalhes nas seções subseqüentes. Além disso, o conceito de algarismos significativos terá relevância na definição de acurácia e precisão na próxima seção.

3.2 ACURÁCIA E PRECISÃO

Os erros associados tanto aos cálculos quanto às medidas podem ser caracterizados com relação a sua acurácia e precisão. A *acurácia* se refere a quão próximo o valor calculado ou medido está do valor verdadeiro. A *precisão* se refere a quão próximos os valores individuais calculados ou medidos estão uns dos outros.

Esses conceitos podem ser ilustrados graficamente usando-se uma analogia da prática de tiro ao alvo. Os buracos de bala em cada alvo na Figura 3.2 podem ser pensa-

FIGURA 3.2
Um exemplo do tiro ao alvo ilustrando os conceitos de acurácia e precisão. (*a*) Inacurado e impreciso; (*b*) acurados e impreciso; (*c*) inacurado e preciso; (*d*) acurado e preciso.



dos como previsões de uma técnica numérica, enquanto a mosca representa a verdade. *Inacurácia* (também chamada *viés*) é definida como um desvio sistemático da verdade. Embora os tiros na Figura 3.2*c* estejam agrupados mais juntos do que aqueles na Figura 3.2*a*, os dois casos são igualmente inacurados porque ambos estão centrados no quadrante superior esquerdo do alvo. As *imprecisões* (também chamadas *incertezas*), por outro lado, referem-se à intensidade do espalhamento. Portanto, embora as Figuras 3.2*b* e *d* sejam igualmente exatas (isto é, centradas na mosca), a última é mais precisa porque os tiros estão agrupados mais juntos.

Os métodos numéricos deveriam ser suficientemente acurados ou sem viés para satisfazer os requisitos de um problema de engenharia particular. Eles também deveriam ser suficientemente precisos para permitirem projetos adequados de engenharia. Neste livro, será usado o termo coletivo *erro* para representar tanto a inacurácia quanto a imprecisão de predições. Com esses conceitos como base, podem-se discutir os fatores que contribuem para o erro dos cálculos numéricos.

3.3 DEFINIÇÕES DE ERRO

Os erros numéricos são causados pelo uso de aproximações para representar operações e quantidades matemáticas exatas. Eles incluem *erros de truncamento,* que resultam quando são feitas aproximações para representar procedimentos matemáticos exatos, e *erros de arredondamento,* que aparecem quando números com uma quantidade limitada de algoritmos significativos são usados para representar números exatos. Para ambos os tipos, a relação entre o resultado exato ou verdadeiro e a aproximação pode ser formulada como

$$\text{Valor verdadeiro} = \text{aproximação} + \text{erro} \tag{3.1}$$

Rearranjando a Equação (3.1), encontra-se que o erro numérico é igual à discrepância entre o valor verdadeiro e a aproximação, ou seja,

$$E_t = \text{valor verdadeiro} - \text{aproximação} \tag{3.2}$$

onde E_t é usado para designar o valor exato do erro. O subscrito t é incluído para designar que esse é o erro "verdadeiro"(*true*, em inglês) em contraste com outros casos, como será descrito em breve, em que uma estimativa "aproximada" do erro deve ser empregada.

Um defeito dessa definição é que ela não leva em conta a ordem de grandeza do valor que está sendo examinado. Por exemplo, um erro de um centímetro é muito mais significativo quando se mede um rebite do que uma ponte. Uma forma de considerar o valor das quantidades que estão sendo calculadas é normalizar o erro com o valor verdadeiro, como em

$$\text{Erro relativo fracionário verdadeiro} = \frac{\text{erro verdadeiro}}{\text{valor verdadeiro}}$$

onde, como especificado pela Equação (3.2), erro = valor verdadeiro − aproximação. O erro relativo também pode ser multiplicado por 100% para expressá-lo como

$$\varepsilon_t = \frac{\text{erro verdadeiro}}{\text{valor verdadeiro}} 100\% \tag{3.3}$$

onde ε_t designa o erro relativo porcentual verdadeiro.

EXEMPLO 3.1 Cálculo de Erros

Enunciado do Problema. Suponha que você tenha a tarefa de medir os comprimentos de uma ponte e de um rebite e que conseguiu 9.999 e 9 cm, respectivamente. Se os valores verdadeiros forem 10.000 e 10 cm, respectivamente, calcule (a) o erro verdadeiro e (b) o erro relativo porcentual verdadeiro para cada caso.

Solução.

(a) O erro na medida da ponte é [Equação (3.2)]

$$E_t = 10.000 - 9.999 = 1 \text{ cm}$$

e para o rebite é

$$E_t = 10 - 9 = 1 \text{ cm}$$

(b) O erro relativo porcentual para a ponte é [Equação (3.3)]

$$\varepsilon_t = \frac{1}{10.000}100\% = 0{,}01\%$$

e para o rebite é

$$\varepsilon_t = \frac{1}{10}100\% = 10\%$$

Portanto, embora ambas as medidas tenham um erro de 1 cm, o erro relativo para o rebite é muito maior. A conclusão seria que foi feito um trabalho adequado na medição da ponte, enquanto que a estimativa para o rebite deixa algo a desejar.

Observe que, para as Equações (3.2) e (3.3), E e ε têm o subscrito t para indicar que o erro está normalizado com relação ao valor verdadeiro. No Exemplo 3.1, esse valor foi fornecido. Entretanto, nas situações reais, tais informações raramente estão disponíveis. Para os métodos numéricos, o valor verdadeiro será conhecido apenas ao se lidar com funções que podem ser resolvidas analiticamente. Este será o caso típico quando se investigar o comportamento teórico de uma técnica particular para sistemas simples. Entretanto, nas aplicações do mundo real, obviamente não se conhecerá *a priori* a resposta verdadeira. Para tais situações, uma alternativa é normalizar o erro usando a melhor estimativa disponível do valor verdadeiro, isto é, a aproximação propriamente dita, como em

$$\varepsilon_a = \frac{\text{erro aproximado}}{\text{aproximação}}100\% \tag{3.4}$$

onde o subscrito a significa que o erro foi normalizado por um valor aproximado. Observe também que, para aplicações do mundo real, a Equação (3.2) não pode ser usada para calcular o termo do erro para a Equação (3.4). Um dos desafios dos métodos numéricos é determinar estimativas de erro na ausência de conhecimento relativo ao valor verdadeiro. Por exemplo, certos métodos numéricos usam uma *abordagem iterativa* para calcular as respostas. Em tais abordagens, uma aproximação atual é feita com base em uma aproximação prévia. Esse processo é realizado repetidamente, ou iterativamente, para se calcular com sucesso (é o que se espera) aproximações cada vez melhores. Para tais casos, o erro é freqüentemente estimado como a diferença entre as aproximações prévia e corrente. Assim, o erro relativo porcentual é determinado de acordo com

$$\varepsilon_a = \frac{\text{aproximação atual} - \text{aproximação prévia}}{\text{aproximação atual}}100\% \tag{3.5}$$

Essa e outras abordagens para expressar erros serão mais bem discutidas em capítulos subseqüentes.

Os sinais nas Equações (3.2) até (3.5) podem ser tanto positivos quanto negativos. Se a aproximação for maior do que o valor verdadeiro (ou a aproximação prévia for maior do que a aproximação atual), o erro é negativo; se a aproximação for menor do que o valor verdadeiro, o erro é positivo. Além disso, para as Equações (3.3) a (3.5), o denominador pode ser menor do que zero, o que também pode levar a um erro negativo. Freqüentemente, ao se realizar cálculos, não há preocupação com o sinal do erro, mas interesse em saber se o valor absoluto porcentual é menor que uma tolerância porcentual pré-especifi-

cada ε_s. Portanto, em geral, é útil usar o valor absoluto das Equações (3.2) até (3.5). Para tais casos, os cálculos são repetidos até que

$$|\varepsilon_a| < \varepsilon_s \tag{3.6}$$

Se essa relação for válida, supõe-se que o resultado está dentro do nível aceitável pré-especificado ε_s. Observe que, no resto deste texto, serão utilizados quase que exclusivamente valores absolutos quando forem usados erros relativos.

Também é conveniente relacionar esses erros ao número de algarismos significativos na aproximação. Pode ser mostrado (Scarborough, 1966) que, se o seguinte critério for satisfeito, é possível ter certeza de que o resultado é correto até *pelo menos n* algarismos significativos.

$$\varepsilon_s = (0{,}5 \times 10^{2-n})\ \% \tag{3.7}$$

EXEMPLO 3.2 Estimativas de Erros para Métodos Iterativos

Enunciado do Problema. Em matemática, as funções, em geral, podem ser representadas por séries infinitas. Por exemplo, a função exponencial pode ser calculada usando-se

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} \tag{E3.2.1}$$

Portanto, conforme mais termos forem adicionados em seqüência, a aproximação se torna uma estimativa cada vez melhor do valor verdadeiro de e^x. A Equação (E3.2.1) é chamada de *expansão em série de Maclaurin*.

Começando com a versão mais simples, $e^x = 1$, some um termo de cada vez para estimar $e^{0,5}$. Depois que cada termo for adicionado, calcule o erro verdadeiro e o erro relativo porcentual aproximado com as Equações (3.3) e (3.5), respectivamente. Observe que o valor verdadeiro é $e^{0,5} = 1{,}648721\ldots$. Adicione termos até que o valor absoluto do erro estimado aproximado ε_a esteja dentro do critério de erro pré-especificado ε_s que garanta três algarismos significativos.

Solução. Inicialmente, a Equação (3.7) pode ser usada para determinar o critério de erro que garante o resultado correto até pelo menos três algarismos significativos:

$$\varepsilon_s = (0{,}5 \times 10^{2-3})\% = 0{,}05\%$$

Portanto, serão adicionados termos à série até que ε_a caia abaixo desse nível.

A primeira estimativa é simplesmente igual à Equação (E3.2.1) com um único termo. Portanto, a primeira estimativa é igual a 1. A segunda estimativa é então gerada somando-se o segundo termo, como em

$$e^x = 1 + x$$

ou, para $x = 0{,}5$,

$$e^{0,5} = 1 + 0{,}5 = 1{,}5$$

Isso representa um erro relativo porcentual verdadeiro de [Equação (3.3)]

$$\varepsilon_t = \frac{1{,}648721 - 1{,}5}{1{,}648721}100\% = 9{,}02\%$$

A Equação (3.5) pode ser usada para determinar uma estimativa aproximada do erro, como em

$$\varepsilon_a = \frac{1{,}5 - 1}{1{,}5}100\% = 33{,}3\%$$

Como ε_a não é menor do que o valor pedido de ε_s, os cálculos continuariam com a soma de um outro termo, x^2 / 2!, e repetindo-se os cálculos de erro. O processo deve ser continuado até que $\varepsilon_a < \varepsilon_s$. Todos os cálculos podem ser resumidos em

Termos	Resultados	ε_t (%)	ε_a (%)
1	1	39,3	
2	1,5	9,02	33,3
3	1,625	1,44	7,69
4	1,645833333	0,175	1,27
5	1,648437500	0,0172	0,158
6	1,648697917	0,00142	0,0158

Portanto, depois que seis termos forem incluídos, o erro aproximado cai abaixo de ε_s = 0,05% e os cálculos param. Entretanto, observe que, em vez de três algarismos significativos, os resultados são exatos até cinco! Isso acontece porque, nesse caso, tanto a Equação (3.5) quanto a (3.7) são conservadoras — ou seja, garantem que o resultado é pelo menos tão bom quanto elas especificaram. Embora, como discutido no Capítulo 6, esse não seja sempre o caso para a Equação (3.5), será verdade a maior parte das vezes.

Com as definições precedentes como base, é possível avançar para dois tipos de erros ligados diretamente com os métodos numéricos: os erros de arredondamento e os erros de truncamento.

3.4 ERROS DE ARREDONDAMENTO

Como mencionado anteriormente, os erros de arredondamento se originam do fato de que o computador mantém apenas um número fixo de algarismos significativos durante os cálculos. Números como π, e ou $\sqrt{7}$ não podem ser expressos por um número fixo de algarismos significativos. Portanto, eles não podem ser representados exatamente por um computador. Além disso, como os computadores usam uma representação na base 2, não podem representar precisamente certos números exatos na base 10. A discrepância introduzida por essa omissão de algarismos significativos é chamada de *erro de arredondamento.*

3.4.1 Representação dos Números no Computador

Os erros numéricos de arredondamento estão diretamente relacionados à maneira como os números são armazenados no computador. A unidade fundamental na qual a informação é representada é chamada de *palavra*, entidade que consiste em uma seqüência de dígitos binários (*b*inary dig*its*), ou *bits*. Os números são tipicamente armazenados em uma ou mais palavras. Para entender como isso é feito, primeiro é necessário revisar algum material relacionado aos sistemas numéricos.

Sistemas Numéricos. Um *sistema numérico* é meramente uma convenção para representar quantidades. Como são 10 dedos na mão e 10 dedos no pé, o sistema numérico com o qual há mais familiaridade é o sistema *decimal* ou o sistema numérico na *base* 10. A base é o número usado como referência para construir o sistema. O sistema na base 10 usa os 10 algarismos — 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 — para representar os números. Por si só, esses algarismos são satisfatórios para contar de 0 a 9.

Para quantidades maiores, combinações desses algarismos básicos devem ser usadas, com a posição ou o *valor do lugar* especificando a ordem de grandeza. O algarismo mais à direita em um número completo representa um número entre 0 e 9. O segundo algarismo a partir da direita representa um múltiplo de 10. O terceiro algarismo a partir da direita representa um múltiplo de 100 e assim por diante. Por exemplo, no número 86.409 há oito grupos de 10.000, seis grupos de 1.000, quatro grupos de 100, zero grupos de 10 e nove unidades a mais, ou

$$(8 \times 10^4) + (6 \times 10^3) + (4 \times 10^2) + (0 \times 10^1) + (9 \times 10^0) = 86.409$$

A Figura 3.3*a* fornece uma representação visual de como o número é formulado no sistema na base 10. Esse tipo de representação é chamada *notação posicional*.

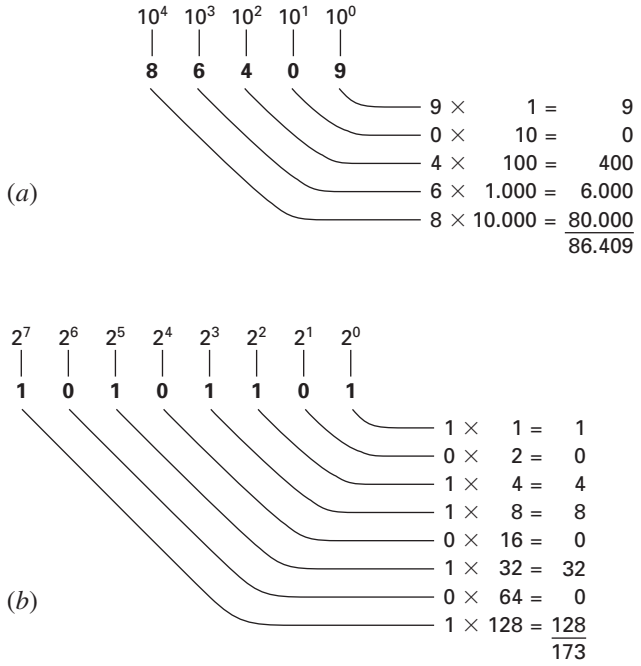
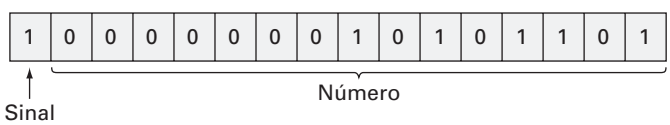


FIGURA 3.3
Como os sistemas (*a*) decimal (base 10) e (*b*) binário (base 2) funcionam. Em (*b*), o número binário 10101101 é equivalente ao número decimal 173.

Como o sistema decimal é tão familiar, em geral não se percebe que existem alternativas. Por exemplo, se os seres humanos tivessem oito dedos na mão e oito dedos no pé, sem dúvida teriam desenvolvido um sistema *octogonal* ou uma representação *na base 8*. Nesse mesmo sentido, o amigo computador é como um animal de dois dedos que é limitado a dois estados — ou 0 ou 1. Isso está relacionado ao fato de que as unidades lógicas primárias dos computadores digitais são componentes elétricas ligadas ou desligadas. Portanto, os números em um computador são representados com um sistema *binário* ou *na base 2*. Da mesma forma como no sistema decimal, as quantidades podem ser representadas usando-se a notação posicional. Por exemplo, o número binário 11 é equivalente a $(1 \times 2^1) + (1 \times 2^0) = 2 + 1 = 3$ no sistema decimal. A Figura 3.3*b* ilustra um exemplo mais complicado.

Representação Inteira. Agora que foi revisado como o número na base 10 pode ser representado na forma binária, é simples conceber como os inteiros são representados em um computador. A abordagem mais direta, chamada de *método dos valores com sinal*, utiliza o primeiro bit de uma palavra para indicar o sinal, com um 0 para positivo e um 1 para negativo. Os bits restantes são usados para armazenar o número. Por exemplo, o valor inteiro −173 seria armazenado em um computador de 16 bits, como na Figura 3.4.

FIGURA 3.4
A representação de um inteiro decimal −173 em um computador de 16 bits usando-se o método do valor com sinal.



EXEMPLO 3.3 Intervalo de Inteiros

Enunciado do Problema. Determine o intervalo dos inteiros na base 10 que pode ser representado em um computador de 16 bits.

Solução. Dos 16 bits, o primeiro bit representa o sinal. Os restantes 15 bits podem conter números binários de 0 a 111111111111111. O limite superior pode ser convertido para um inteiro decimal, como

$$(1 \times 2^{14}) + (1 \times 2^{13}) + \cdots + (1 \times 2^{1}) + (1 \times 2^{0})$$

que é igual a 32.767 (observe que essa expressão pode ser calculada simplesmente como $2^{15} - 1$). Portanto, uma palavra no computador de 16 bits pode armazenar inteiros decimais variando de −32.767 a 32.767. Além disso, como o zero já foi definido como 0000000000000000, é redundante usar o número 1000000000000000 para definir um "menos zero". Portanto, ele é normalmente usado para representar um número negativo adicional: −32.768, e o intervalo é de −32.768 a 32.767.

Observe que esse método descrito do valor com sinal não é usado para representar inteiros nos computadores convencionais. Uma abordagem preferida, chamada de técnica do *complemento de 2*, incorpora diretamente o sinal no valor absoluto do número em vez de fornecer um bit separado para representar mais ou menos (ver Chapra e Canale, 1994). Entretanto, o Exemplo 3.3 ainda serve para ilustrar como todos os computadores digitais são limitados em sua capacidade de representar inteiros. Isto é, os números acima ou abaixo de um intervalo não podem ser representados. Uma limitação mais séria é encontrada no armazenamento e na manipulação de quantidades fracionárias como descrito a seguir.

Representação em Ponto Flutuante. As quantidades fracionárias são representadas tipicamente em computadores usando-se a forma de ponto flutuante. Nessa abordagem, o número é expresso como uma parte fracionária, chamada *mantissa* ou *significando*, e uma parte inteira, chamada de *expoente* ou *característica*, como em

$$m \cdot b^{e}$$

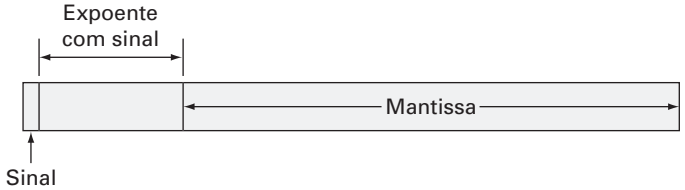
onde m é a mantissa, b é a base do sistema numérico que está sendo usado, e e é o expoente. Por exemplo, o número 156,78 poderia ser representado como $0{,}15678 \times 10^{3}$ em um sistema de ponto flutuante na base 10.

A Figura 3.5 mostra uma forma na qual um número em ponto flutuante poderia ser armazenado em uma palavra. O primeiro bit fica reservado para o sinal; a próxima série de bits, para o expoente com sinal; e os últimos bits, para a mantissa.

Observe que a mantissa está usualmente *normalizada* se ela tiver o algarismo dominante nulo. Por exemplo, suponha que a quantidade 1/34 = 0,029411765… seja armazenada em um sistema na base 10 em ponto flutuante que permita que apenas quatro casas decimais sejam armazenadas. Então, a quantidade 1/34 seria armazenada como

$$0{,}0294 \times 10^{0}$$

FIGURA 3.5
A maneira como um número em ponto flutuante é armazenado em uma palavra.



Entretanto, nesse processo, a inclusão do zero inútil à direita da vírgula nas força a abandonar o algarismo 1 na quinta casa decimal. O número pode ser normalizado para se remover o zero dominante pela multiplicação da mantissa por 10 e pela diminuição do expoente por 1 para fornecer

$$0{,}2941 \times 10^{-1}$$

Assim, manteve-se um algarismo significativo adicional quando o número foi armazenado.

A conseqüência da normalização é que o valor absoluto de m é limitado. Isto é,

$$\frac{1}{b} \leq m < 1 \tag{3.8}$$

onde b é a base. Por exemplo, para um sistema na base 10, m iria variar entre 0,1 e 1, e para um sistema na base 2, entre 0,5 e 1.

A representação em ponto flutuante permite que tanto frações quanto números muito grandes sejam expressos em um computador. Contudo, ela tem algumas desvantagens. Por exemplo, os números em ponto flutuante ocupam mais espaço e levam mais tempo para processar do que os números inteiros. Mais significativo, porém, é que seu uso introduz uma fonte de erros, já que a mantissa mantém apenas um número finito de algarismos significativos. Portanto, foi introduzido um erro de arredondamento.

EXEMPLO 3.4 Conjunto Hipotético de Números em Ponto Flutuante

Enunciado do Problema. Crie um conjunto de números hipotéticos em ponto flutuante para uma máquina que armazena informação usando palavras de 7 bits. Use o primeiro bit para o sinal do número, os próximos três para o sinal e o módulo do expoente, e os três últimos para o módulo da mantissa (Figura 3.6).

Solução. O menor número positivo possível é mostrado na Figura 3.6. O 0 inicial indica que a quantidade é positiva. O 1 na segunda posição mostra que o expoente tem um sinal negativo. Os 1 na terceira e na quarta posição dão o valor máximo do expoente como sendo

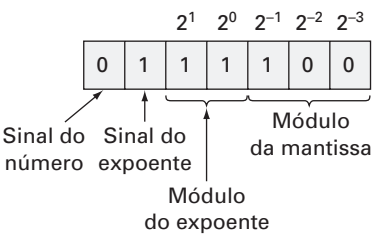
$$1 \times 2^1 + 1 \times 2^0 = 3$$

Portanto, o expoente será -3. Finalmente, a mantissa é especificada pelo 100 nas últimas três posições, o que coincide com

$$1 \times 2^{-1} + 0 \times 2^{-2} + 0 \times 2^{-3} = 0{,}5$$

Embora uma mantissa menor seja possível (isto é, 000, 001, 010, 011), o valor 100 é usado por causa do limite imposto pela normalização [Equação (3.8)]. Desse modo, o menor número positivo possível para esse sistema é $+0{,}5 \times 2^{-3}$, que é igual a 0,0625 no sistema na base 10. Os próximos números mais altos são obtidos aumentando-se a mantissa, como em

FIGURA 3.6
O menor número positivo em ponto flutuante possível para o Exemplo 3.4.



$$0111101 = (1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{-3} = (0{,}078125)_{10}$$

$$0111110 = (1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3}) \times 2^{-3} = (0{,}093750)_{10}$$

$$0111111 = (1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{-3} = (0{,}109375)_{10}$$

Observe que os equivalentes na base 10 são igualmente espaçados, com um intervalo de 0,015625.

Nesse ponto, para continuar a aumentar, é preciso diminuir o expoente para 10, o que dá um valor de

$$1 \times 2^1 + 0 \times 2^0 = 2$$

A mantissa é diminuída para o seu menor valor, 100. Portanto, o próximo número é

$$0110100 = (1 \times 2^{-1} + 0 \times 2^{-2} + 0 \times 2^{-3}) \times 2^{-2} = (0{,}125000)_{10}$$

Isso ainda representa um salto de $0{,}125000 - 0{,}109375 = 0{,}015625$. No entanto, agora quando os números mais altos forem gerados aumentando-se a mantissa, o espaçamento terá um comprimento de 0,03125,

$$0110101 = (1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{-2} = (0{,}156250)_{10}$$

$$0110110 = (1 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3}) \times 2^{-2} = (0{,}187500)_{10}$$

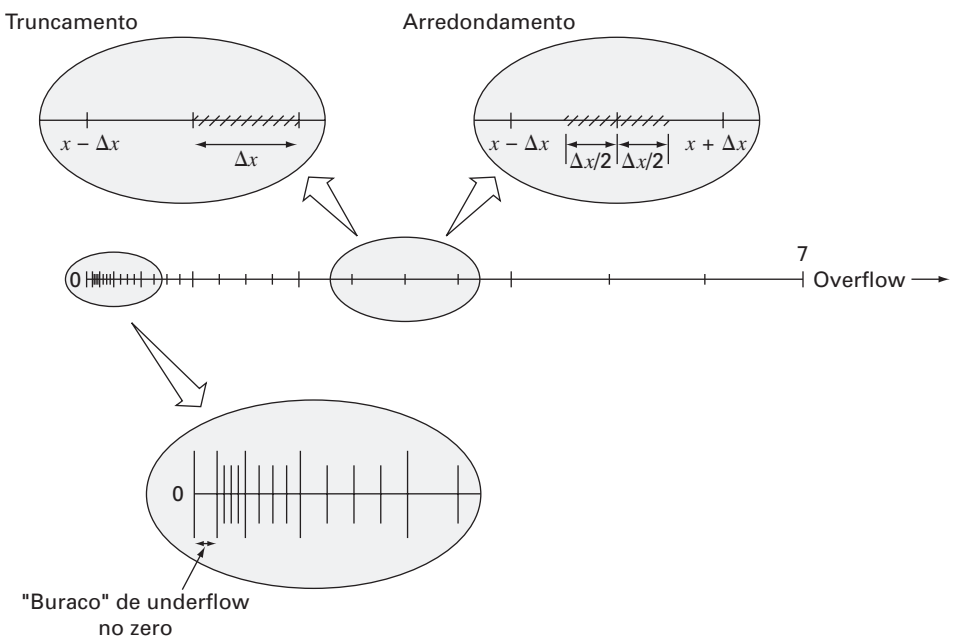
$$0110111 = (1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{-2} = (0{,}218750)_{10}$$

Esse padrão é repetido conforme quantidades maiores forem representadas até que o número máximo seja atingido,

$$0011111 = (1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}) \times 2^{3} = (7)_{10}$$

O conjunto numérico final é mostrado graficamente na Figura 3.7.

FIGURA 3.7
O sistema numérico hipotético desenvolvido no Exemplo 3.4. Cada valor é indicado por um traço. Apenas os números positivos são mostrados. Um conjunto idêntico também se estenderia na direção negativa.



A Figura 3.7 mostra diversos aspectos da representação em ponto flutuante que são significativos com relação aos erros de arredondamento do computador:

1. *Existe um intervalo limitado de quantidades que podem ser representadas*. Exatamente como no caso dos inteiros, há números positivos e negativos grandes que não podem ser representados. Tentativas de usar números fora do intervalo aceitável vão resultar no que é chamado de um *erro de overflow*. Entretanto, além das quantidades grandes, a representação em ponto flutuante tem a limitação adicional de que números muito pequenos também não podem ser representados. Isso é ilustrado pelo "buraco" causado pelo *underflow* entre zero e o primeiro número positivo na Figura 3.7. Deve ser observado que esse buraco se torna maior por causa da restrição de normalização da Equação (3.8).
2. *Existe apenas um número finito de quantidades que podem ser representadas dentro do intervalo*. Portanto, o grau de precisão é limitado. Obviamente, os números irracionais não podem ser representados exatamente. Além disso, os números racionais que não coincidem exatamente com muitos valores no conjunto também não podem ser representados precisamente. Os erros introduzidos pela aproximação em ambos os casos são chamados de erros de *quantização*. A aproximação propriamente dita é feita de uma das duas maneiras: truncando ou arredondando. Por exemplo, suponha que o valor de $\pi = 3{,}14159265358\ldots$ deva ser armazenado em um sistema numérico na base 10 com sete algarismos significativos. Um método de aproximação seria simplesmente omitir ou truncar o oitavo termo e os termos mais altos, como em $\pi = 3{,}141592$, com a introdução de um erro associado de [Equação (3.2)].

   $E_t = 0{,}00000065...$

   Essa técnica de manter apenas os termos significativos é chamada de "truncamento" no jargão da computação. Preste atenção para não confundir esse termo com os erros de truncamento discutidos no Capítulo 4. Observe que, para o sistema numérico na base 2 na Figura 3.7, truncar significa que qualquer quantidade que caia dentro de um intervalo de comprimento Δx será armazenada como a quantidade na extremidade inicial do intervalo. Assim, o limitante superior do erro para o truncamento é Δx. Adicionalmente, um viés é introduzido porque todos os erros são positivos. A deficiência do truncamento decorre de que termos mais altos na representação decimal completa não têm impacto na versão aproximada. Por exemplo, no exemplo de π, o primeiro algarismo descartado é 6. Portanto, o último dígito mantido deveria ser arredondado para cima para fornecer 3,141593. Tal *arredondamento* reduz o erro para

   $E_t = 0{,}00000035...$

   Conseqüentemente, o arredondamento fornece um erro absoluto menor do que o truncamento. Observe que, para o sistema numérico na base 2 na Figura 3.7, arredondar significa que qualquer quantidade caindo dentro de um intervalo de comprimento Δx será representada pelo número permitido mais próximo. Então, o limitante superior do erro para o arredondamento é $\Delta x/2$. Adicionalmente, nenhum viés é introduzido porque alguns erros são positivos e alguns são negativos. Alguns computadores usam arredondamento. Entretanto, isso sobrecarrega a carga computacional e, conseqüentemente, muitas máquinas usam simplesmente o truncamento. Essa abordagem é justificada pela hipótese de que o número de algarismos significativos é suficientemente grande para que o erro de arredondamento resultante seja usualmente desprezível.
3. *O intervalo entre os números, Δx, aumenta quando o módulo dos números cresce*. É essa característica, claro, que permite que a representação em ponto flutuante preserve os algarismos significativos. Mas ela também significa que o erro de quantização será proporcional ao módulo do número que está sendo representado. Para números em ponto flutuante normalizados, essa proporcionalidade pode ser expressa, para os casos em que é empregado o truncamento, como

$$\frac{|\Delta x|}{|x|} \leq \mathscr{E} \tag{3.9}$$

e, para os casos nos quais o arredondamento é usado, como

$$\frac{|\Delta x|}{|x|} \leq \frac{\mathscr{E}}{2} \tag{3.10}$$

onde $\mathscr{E}$ é chamado de *épsilon da máquina*, que pode ser calculado como

$$\mathscr{E} = b^{1-t} \tag{3.11}$$

onde b é a base numérica e t é o número de algarismos significativos na mantissa. Observe que as desigualdades nas Equações (3.9) e (3.10) significam que esses são limitantes para os erros. Isto é, eles especificam os piores casos.

EXEMPLO 3.5 Épsilon da Máquina

Enunciado do Problema. Determine o épsilon da máquina e verifique sua efetividade na caracterização dos erros do sistema numérico do Exemplo 3.4. Suponha que seja usado truncamento.

Solução. O sistema em ponto flutuante hipotético do Exemplo 3.4 usa valores de base $b = 2$ e o número de bits da mantissa $t = 3$. Logo, o épsilon da máquina seria, [Equação (3.11)]

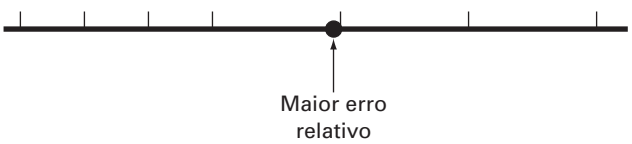
$$\mathscr{E} = 2^{1-3} = 0{,}25$$

Conseqüentemente, o erro de quantização relativo deveria ser limitado por 0,25 para o truncamento. Os maiores erros relativos deveriam ocorrer para aquelas quantidades que caem imediatamente abaixo do limitante superior do primeiro intervalo entre os números igualmente espaçados sucessivos (Figura 3.8). Aqueles números caindo nos maiores intervalos a seguir teriam o mesmo valor de Δx mas um valor de x maior, portanto, teriam um erro relativo menor. Um exemplo de um erro máximo seria um valor caindo imediatamente abaixo do limitante superior do intervalo entre $(0{,}125000)_{10}$ e $(0{,}156250)_{10}$. Para este caso, o erro seria menor que

$$\frac{0{,}03125}{0{,}125000} = 0{,}25$$

Portanto, o erro é o previsto pela Equação (3.9).

FIGURA 3.8
O maior erro de quantização ocorrerá para aqueles valores caindo imediatamente abaixo do limitante superior do primeiro de uma série de intervalos igualmente espaçados.




```
epsilon = 1
DO
  IF (epsilon+1 ≤ 1)EXIT
  epsilon = epsilon/2
END DO
epsilon = 2 × epsilon
```

FIGURA 3.9
Um pseudocódigo para determinar o épsilon da máquina para um computador binário.

A dependência do módulo dos erros de quantização tem algumas aplicações práticas nos métodos numéricos. A maioria delas diz respeito à operação comumente usada de testar se dois números são iguais. Isso ocorre tanto quando se testa a convergência de quantidades como também nos mecanismos de parada dos processos iterativos (lembre-se do Exemplo 3.2). Para esses casos, deveria ficar claro que, em vez de testar se duas quantidades são iguais, seria aconselhável testar se a diferença delas é menor do que uma pequena tolerância aceitável. Além disso, também deveria ficar evidente que deveriam ser comparadas diferenças normalizadas em vez de absolutas, particularmente ao se lidar

com números de módulo grande. Além disso, o épsilon da máquina pode ser usado na formulação de critérios de parada ou de convergência. Isso garante que os programas são robustos — ou seja, eles não dependem do computador no qual estão sendo implementados. A Figura 3.9 lista um pseudocódigo para determinar automaticamente o épsilon da máquina de um computador binário.

Precisão Estendida. Deve-se notar neste ponto que, embora os erros de arredondamento possam ser importantes em contextos tais como testes de convergência, o número de algarismos significativos implementados na maioria dos computadores permite que a maior parte dos cálculos em engenharia seja feita com precisão mais do que aceitável. Por exemplo, o sistema numérico hipotético na Figura 3.7 é um exagero grosseiro que foi usado para propósitos ilustrativos. Os computadores comerciais usam palavras muito maiores e, conseqüentemente, permitem que os números sejam expressos com precisão mais do que adequada: computadores que utilizam o formato IEEE permitem que sejam usados 24 bits para a mantissa, o que se traduz em cerca de sete algarismos significativos[1] na base 10 com um intervalo de cerca de 10^{-38} a 10^{39}.

Reconhecido isso, existem ainda casos nos quais os erros de arredondamento se tornam críticos. Por essa razão, a maior parte dos computadores permite que seja especificada a precisão estendida. A mais comum delas é a dupla precisão, na qual o número de palavras usadas para armazenar um número em ponto flutuante é dobrado, o que fornece cerca de 15 a 16 algarismos decimais de precisão com um intervalo de aproximadamente 10^{-308} a 10^{308}.

Em muitos casos, o uso de quantidades em dupla precisão diminui enormemente o efeito dos erros de arredondamento. Entretanto, essas soluções têm um custo, pois elas também exigem muito mais memória e tempo de execução. A diferença no tempo de execução para pequenos cálculos pode parecer insignificante. Porém, conforme os programas se tornam maiores e mais complicados, o tempo adicional de execução pode tornar-se considerável e ter um impacto negativo na eficiência da ferramenta para resolver o problema. Portanto, a precisão estendida não deveria ser usada frivolamente. Em vez disso, deveria ser utilizada seletivamente onde fornecerá um benefício máximo a um custo mínimo em termos de tempo e execução. Nas próximas seções, será abordado com mais detalhes como os erros de arredondamento afetam os cálculos, o que proporcionará uma base de entendimento para guiá-lo no uso do recurso de dupla precisão.

Antes de prosseguir, é preciso observar que alguns pacotes de software comumente usados (por exemplo, Excel, Mathcad) usualmente utilizam dupla precisão para representar quantidades numéricas. Portanto, as pessoas que desenvolveram esses pacotes decidiram que minimizar os erros de arredondamento tinha precedência sobre qualquer perda de velocidade causada pelo uso da precisão estendida. Outros, como o MATLAB, permitem que o usuário utilize precisão estendida, se quiser.

3.4.2 Manipulações Aritméticas de Números no Computador

Paralelamente às limitações de um sistema numérico no computador, as manipulações aritméticas reais envolvendo esses números também podem resultar em erros de arredondamento. Na seção seguinte, inicialmente será ilustrado como as operações aritméticas comuns afetam os erros de arredondamento. A seguir, serão investigadas algumas manipulações particulares especialmente suscetíveis a erros de arredondamento.

Operações Aritméticas Comuns. Por causa de sua familiaridade, será usado um sistema numérico na base 10 normalizado para ilustrar os efeitos dos erros de arredondamento em adição, subtração, multiplicação e divisão simples. Outras bases numéricas se comportariam de uma forma análoga. Para simplificar a discussão, um computador decimal hipotético com uma mantissa de 4 algarismos e um expoente de 1 algarismo será empregado, e, além disso será usado o truncamento. O arredondamento levaria a erros similares, embora menos dramáticos.

[1] Observe que apenas 23 bits são realmente usados para armazenar a mantissa. Entretanto, por causa da normalização, o primeiro bit da mantissa é sempre 1 e, portanto, não é armazenado. Logo, o primeiro bit junto com os 23 bits armazenados dá um total de 24 bits de precisão para a mantissa.

Quando dois números em ponto flutuante são somados, a mantissa com o número de menor expoente é modificada de forma que os expoentes sejam os mesmos, o que tem o efeito de alinhar os pontos decimais. Por exemplo, suponha que você queira somar $0{,}1557 \cdot 10^1 + 0{,}4381 + 10^{-1}$. A vírgula da mantissa do segundo número é deslocada para a esquerda um número de posições igual à diferença dos expoentes $[1 - (-1) = 2]$, como em

$$0{,}4381 \cdot 10^{-1} \rightarrow 0{,}004381 \cdot 10^1$$

Agora, os números podem ser somados,

$$\begin{array}{l} 0{,}1557 \quad \cdot 10^1 \\ \underline{0{,}004381 \cdot 10^1} \\ 0{,}160081 \cdot 10^1 \end{array}$$

e o resultado, truncado para $0{,}1600 \cdot 10^1$. Observe como os dois últimos algarismos do segundo número que foram transladados para a direita foram essencialmente perdidos nos cálculos.

A subtração é feita de forma idêntica à soma, exceto pelo fato de que o sinal do subtraendo é trocado. Por exemplo, suponha que estejamos subtraindo 26,86 de 36,41. Isto é,

$$\begin{array}{r} 0{,}3641 \cdot 10^2 \\ \underline{-\ 0{,}2686 \cdot 10^2} \\ 0{,}0955 \cdot 10^2 \end{array}$$

Para esse caso, o resultado não está normalizado e é necessário transladar a vírgula uma posição para a direita para fornecer $0{,}9550 \cdot 10^1 = 9{,}550$. Observe que o zero adicionado no final da mantissa não é significativo, mas é simplesmente um apêndice para preencher um espaço vazio criado pela translação. Resultados ainda mais dramáticos seriam obtidos quando os números fossem muito próximos, como

$$\begin{array}{r} 0{,}7642 \cdot 10^3 \\ \underline{-\ 0{,}7641 \cdot 10^3} \\ 0{,}0001 \cdot 10^3 \end{array}$$

que deveria ser convertido para $0{,}1000 \cdot 10^0 = 0{,}1000$. Logo, nesse caso, três zeros não-significativos são adicionados. Isso introduz um erro computacional substancial porque as manipulações subseqüentes se comportariam como se esses zeros fossem significativos. Como se verá em uma seção posterior, a perda de significado durante a subtração de números quase iguais está entre as maiores fontes de erros de arredondamento nos métodos numéricos.

A multiplicação e a divisão são um pouco mais simples que a adição e a subtração. Os expoentes são somados e as mantissas, multiplicadas. Como a multiplicação de duas mantissas de n algarismos fornecerá o resultado com 2 n algarismos, a maior parte dos computadores tem resultados intermediários em um registro de comprimento duplo. Por exemplo,

$$0{,}1363 \cdot 10^3 \times 0{,}6423 \cdot 10^{-1} = 0{,}08754549 \cdot 10^2$$

Se, como nesse caso, um zero dominante for introduzido, o resultado é normalizado,

$$0{,}08754549 \cdot 10^2 \rightarrow 0{,}8754549 \cdot 10^1$$

e truncado para fornecer

$$0{,}8754 \cdot 10^1$$

A divisão é feita de uma forma parecida, mas as mantissas são divididas e os expoentes são subtraídos. A seguir os resultados são normalizados e truncados.

Cálculos Grandes. Alguns métodos necessitam de um número extremamente grande de manipulações aritméticas para chegarem ao seu resultado final. Além disso, esses cálculos são freqüentemente interdependentes. Isto é, os últimos cálculos dependem dos resultados dos primeiros. Conseqüentemente, mesmo que os erros de arredondamento individuais sejam pequenos, o efeito cumulativo sobre um grande número de cálculos pode ser significativo.

EXEMPLO 3.6 Número Grande de Cálculos Interdependentes

Enunciado do Problema. Investigue o efeito do erro de arredondamento em um número grande de cálculos interdependentes. Desenvolva um programa para somar um número 100.000 vezes. Some o número 1 em precisão simples, e 0,00001 em precisão simples e dupla.

Solução. A Figura 3.10 mostra um programa em Fortran 90 que faz essa soma. Enquanto a soma em precisão simples de 1 fornece o resultado esperado, a soma de 0,00001 em precisão simples dá uma discrepância grande. Esse erro é reduzido significativamente quando 0,00001 é somado em dupla precisão.

Os erros de quantização são a fonte dessas discrepâncias. Como o inteiro 1 pode ser representado exatamente no computador, ele pode ser somado exatamente. Em contraste, 0,00001 não pode ser representado exatamente e é quantizado por um valor ligeiramente diferente de seu valor verdadeiro. Embora essa discrepância muito pequena fosse desprezível para um cálculo pequeno, ela se acumula depois de somas repetidas. O problema ainda ocorre em dupla precisão, mas é atenuado porque o erro de quantização é muito menor.

FIGURA 3.10
Um programa em Fortran 90 para somar um número 10^5 vezes. O caso da soma do número 1 em precisão simples e do número 10^{-5} em precisão simples e em precisão dupla.

```
PROGRAM fig0310
IMPLICIT none
INTEGER::i
REAL::sum1, sum2, x1, x2
DOUBLE PRECISION::sum3, x3
sum1=0.
sum2=0.
sum3=0.
x1=1.
x2=1.e-5
x3=1.d-5
DO i=1,100000
  sum1=sum1+x1
  sum2=sum2+x2
  sum3=sum3+x3
END DO
PRINT *, sum1
PRINT *, sum2
PRINT *, sum3
END
output:
100000.000000
       1.000990
  9.999999999980838E-001
```

Observe que o tipo de erro ilustrado no exemplo anterior é, de uma certa forma, atípico, no sentido de que todos os erros nas operações repetidas têm o mesmo sinal. Na maioria dos casos, os erros em cálculos longos alternam os sinais de uma forma aleatória e, portanto, em geral se cancelam. Entretanto, também há exemplos em que tais erros não se cancelam, mas, na realidade, levam a um resultado final espúrio. As seções seguintes são destinadas a fornecer uma percepção das formas pelas quais isso pode acontecer.

Adicionando um Número Grande e um Número Pequeno. Suponha que seja somado um número pequeno, 0,0010, a um número grande, 4.000, usando-se um computa-

dor hipotético com 4 algarismos na mantissa e 1 algarismo no expoente. O número pequeno é modificado de forma que o seu expoente coincida com o grande,

$$\begin{array}{ll} 0{,}4000 & \cdot 10^4 \\ \underline{0{,}0000001} & \underline{\cdot 10^4} \\ 0{,}4000001 & \cdot 10^4 \end{array}$$

que é truncado para $0{,}4000 \cdot 10^4$. Portanto, a soma poderia muito bem não ter sido feita!

Esse tipo de erro pode ocorrer em cálculos de uma série infinita. Os termos iniciais de tais séries são em geral relativamente grandes em comparação com os termos posteriores. Portanto, depois que uns poucos termos foram somados, a situação é a de somar quantidades muito pequenas a uma quantidade grande.

Uma forma de atenuar esse tipo de erro é somar as séries na ordem inversa — isto é, em uma ordem crescente em vez de decrescente. Dessa forma, cada novo termo será de valor absoluto comparável à soma acumulada (ver Problema 3.4).

Cancelamento na Subtração. Esse termo se refere ao erro de arredondamento induzido quando se subtraem dois números em ponto flutuante quase iguais.

Um exemplo comum em que isso pode ocorrer envolve encontrar as raízes de uma equação quadrática ou uma parábola usando a fórmula quadrática,

$$\begin{matrix} x_1 \\ x_2 \end{matrix} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{3.12}$$

Para os casos nos quais $b^2 \gg 4ac$, a diferença no numerador vai ser muito pequena, e a dupla precisão pode aliviar o problema. Além disso, uma formulação alternativa pode ser usada para minimizar o cancelamento na subtração,

$$\begin{matrix} x_1 \\ x_2 \end{matrix} = \frac{-2c}{b \pm \sqrt{b^2 - 4ac}} \tag{3.13}$$

Uma ilustração do problema e o uso dessa fórmula alternativa são fornecidos no próximo exemplo.

EXEMPLO 3.7 Cancelamento na Subtração

Enunciado do Problema. Calcule os valores das raízes de uma equação quadrática com $a = 1$, $b = 3000.001$ e $c = 3$. Compare os valores calculados com as raízes verdadeiras $x_1 = -0{,}001$ e $x_2 = -3000$.

Solução. A Figura 3.11 mostra um programa em Fortran 90 que calcula as raízes x_1 e x_2 com base na fórmula quadrática [Equação (3.12)]. Observe que são dadas tanto a versão com precisão simples quanto a com dupla precisão. Enquanto os resultados para x_2 são adequados, os erros percentuais relativos para x_1 são ruins na versão com precisão simples, $\varepsilon_t = 2{,}4\%$. Esse nível poderia ser inadequado para muitas aplicações em problemas de engenharia. Tal resultado é particularmente surpreendente, porque se utiliza uma fórmula analítica para obter a solução!

A perda de significado ocorre na linha de ambos os programas na qual dois números relativamente grandes são subtraídos. Problemas similares não ocorrem quando os mesmos números são somados.

Com base nessa discussão, é possível chegar à conclusão geral de que a fórmula quadrática será suscetível a cancelamento na subtração sempre que $b^2 \gg 4ac$. Uma forma de contornar esse problema é usar dupla precisão. Uma outra é reescrever a fórmula quadrática no formato da Equação (3.13). Como na saída do programa, ambas as opções dão um erro muito menor porque o cancelamento na subtração é minimizado ou evitado.

Observe que, como no exemplo anterior, muitas vezes o cancelamento na subtração pode ser contornado usando uma transformação. Entretanto, o único remédio geral é usar precisão estendida.

```
PROGRAM fig0311
IMPLICIT none
REAL::a,b,c,d,x1,x2,x1r
DOUBLE PRECISION::aa,bb,cc,dd,x11,x22
a = 1.
b = 3000.001
c = 3.
d = SQRT(b * b - 4. * a * c)
x1 = (-b + d) / (2. * a)
x2 = (-b - d) / (2. * a)
PRINT *, 'Resultado em precisão simples'
PRINT '(1x,a10,f20.14)', 'x1 = ', x1
PRINT '(1x,a10,f10.4)', 'x2 = ', x2
PRINT *
aa = 1.
bb = 3000.001
cc = 3.
dd = SQRT(bb * bb - 4. * aa * cc)
x11 = (-bb + dd) / (2. * aa)
x22 = (-bb - dd) / (2. * aa)
PRINT *, 'Resultado em dupla precisão:'
PRINT '(1x,a10,f20.14)', 'x1 = ', x11
PRINT '(1x,a10,f10.4)', 'x2 = ', x22
PRINT *
PRINT *, 'Fórmula modificada para a
primeira raiz:'
x1r = -2. * c / (b + d)
PRINT '(1x,a10,f20.14)', 'x1 = ', x1r
END

OUTPUT:
Resultado em precisão simples:
     x1 =     -.00097656250000
     x2 = -3000.0000
Resultado em dupla precisão:
     x1 =     -.00100000000771
     x2 = -3000.0000
Fórmula modificada para a primeira raiz:
     x1 =     -.00100000000000
```

FIGURA 3.11
Um programa em Fortran 90 para determinar as raízes de uma equação quadrática.

Borrão. O borrão ocorre sempre que termos individuais em uma soma são maiores do que a soma propriamente dita. Como no próximo exemplo, um caso no qual isso acontece é em uma série com sinais mistos.

EXEMPLO 3.8 Cálculo de e^x usando Série Infinita

Enunciado do Problema. A função exponencial $y = e^x$ é dada pela série infinita

$$y = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots$$

Calcule esta função para $x = 10$ e $x = -10$ e preste atenção aos problemas dos erros de arredondamento.

Solução. A Figura 3.12*a* fornece um programa em Fortran 90 que usa a série infinita para calcular e^x. A variável *i* é o número de termos na série, *termo* é o valor do termo atual somado à série, e *soma* é o valor acumulado da série. A variável *teste* é o valor acumulado da série antes de adicionar *termo*. A soma da série é interrompida quando o computador não puder detectar a diferença entre *teste* e *soma*.

A Figura 3.12*b* mostra o resultado de rodar o programa para $x = 10$. Observe que esse caso é completamente satisfatório. O resultado final é atingido com 31 termos com o valor da série idêntico ao valor da função de biblioteca até 7 algarismos significativos.

A Figura 3.12*c* mostra resultados similares para $x = -10$. Entretanto, para esse caso, os resultados calculados pela série não têm nem o mesmo sinal que o resultado verdadeiro. Na realidade, o resultado negativo está abrindo uma questão séria, porque e^x nunca pode ser menor do que zero. O problema aqui é causado pelos erros de arredondamento. Observe que vários dos termos que compõem a soma são muito maiores do que o resultado final da soma. Além disso, ao contrário do caso anterior, os termos individuais variam de sinal. Portanto, na realidade estão sendo somados e subtraídos números grandes (cada um com algum pequeno erro) e um grande significado é posto nas diferenças — isto é, no cancelamento na subtração. Logo, pode-se ver que o culpado nesse exemplo do borrão é, na verdade, o cancelamento na subtração. Para tais casos, é apropriado procurar alguma outra estratégia computacional — por exemplo,

tentar calcular $y = e^{-10}$ como $y = (e^{-1})^{10}$. Além de tal reformulação, o único recurso geral é a precisão estendida.

Produtos Internos. Deve ficar claro da última seção que algumas séries infinitas são particularmente suscetíveis a erros de arredondamento. Felizmente, o cálculo de séries não é uma das operações mais comuns nos métodos numéricos. Uma manipulação muito mais comum é o cálculo do produto escalar, como em

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

FIGURA 3.12
(*a*) Um programa em Fortran 90 para calcular e^x usando uma série infinita. (*b*) Cálculo de e^x. (*c*) Cálculo de e^{-x}.

(*a*) Programa

```
PROGRAM fig0312
IMPLICIT none
REAL::termo, teste, soma, x
INTEGER::i
i = 0
termo = 1.
soma = 1.
teste = 0.
PRINT *, 'x = '
READ *, x
PRINT *, 'i', 'termo', 'soma'
DO
  IF (soma.EQ.teste) EXIT
  PRINT *, i, termo, soma
  i = i + 1
  termo = termo*x/i
  teste = soma
  soma = soma+termo
END DO
PRINT *, 'valor exato =',exp(x)
END
```

(*b*) Cálculo de e^{10}

```
x=
10
i       termo             soma
0        1,000000         1,000000
1       10,000000        11,000000
2       50,000000        61,000000
3      166,666700       227,666700
4      416,666700       644,333400
5      833,333400      1477,667000
                  .
                  .
                  .
27   9,183693E-02     22026,420000
28   3,279890E-02     22026,450000
29   1,130997E-02     22026,460000
30   3,769989E-03     22026,470000
31   1,216126E-03     22026,470000
valor exato =   22026,460000
```

(*c*) Cálculo de e^{-10}

```
x=
-10
i       termo             soma
0        1,000000         1,000000
1      -10,000000        -9,000000
2       50,000000        41,000000
3     -166,666700      -125,666700
4      416,666700       291,000000
5     -833,333400      -542,333400
                  .
                  .
                  .
41  -2,989312E-09     8,137590E-05
42   7,117410E-10     8,137661E-05
43  -1,655212E-10     8,137644E-05
44   3,761845E-11     8,137648E-05
45  -8,359655E-12     8,137647E-05
valor exato =   4,539993E-05
```

Essa operação é muito comum, particularmente na solução de equações algébricas lineares simultâneas. Tais somas são suscetíveis a erro de arredondamento. Conseqüentemente, em geral é desejável calculá-las com precisão estendida.

Embora as seções anteriores devessem fornecer regras empíricas para atenuar o erro de arredondamento, elas não disponibilizam uma forma direta além da tentativa e erro para realmente determinar o efeito de tais erros em um cálculo. No próximo capítulo, será introduzida a série de Taylor, que vai fornecer uma abordagem matemática para estimar tais efeitos.

PROBLEMAS

3.1 Converta os seguintes números na base 2 para a base 10: (**a**)101101, (**b**) 101,101 e (**c**) 0,01101.

3.2 Faça seu próprio programa baseado na Figura 3.9 e use-o para determinar o épsilon da máquina do seu computador.

3.3 De forma parecida àquela na Figura 3.9, escreva um programa curto para determinar o menor número, x_{min}, usado no computador que você vai utilizar junto com este livro. Observe que o seu computador será incapaz de distinguir de forma confiável entre zero e uma quantidade que seja menor que esse número.

3.4 A série infinita

$$f(n) = \sum_{i=1}^{n} \frac{1}{i^4}$$

converge para um valor $f(n) = \pi^4/90$ quando n tende a infinito. Escreva um programa em precisão simples para calcular $f(n)$ para $n = 10.000$, calculando a soma de $i = 1$ a 10.000. A seguir, repita os cálculos, mas na ordem inversa — isto é, de $i = 10.000$ até 1 usando incrementos de -1. Em cada caso, calcule o erro relativo percentual verdadeiro. Explique os resultados.

3.5 Calcule e^{-5} usando as duas abordagens

$$e^{-x} = 1 - x + \frac{x^2}{2} - \frac{x^3}{3!} + \cdots$$

e

$$e^{-x} = \frac{1}{e^x} = \frac{1}{1 + x + \dfrac{x^2}{2} + \dfrac{x^3}{3!} + \cdots}$$

e compare com o valor verdadeiro de $6{,}737947 \times 10^{-3}$. Use 20 termos para calcular cada série e calcule os erros relativos aproximados verdadeiros quando cada termo for adicionado.

3.6 A derivada de $f(x) = 1/(1 - 3x^2)^2$ é dada por

$$\frac{6x}{(1 - 3x^2)^2}$$

Você espera ter dificuldades calculando essa função em $x = 0{,}577$? Tente isso usando aritmética com 3 e 4 algarismos significativos, com truncamento.

3.7 **(a)** Calcule o polinômio

$$y = x^3 - 7x^2 + 8x - 0{,}35$$

em $x = 1{,}37$. Use aritmética com 3 algarismos significativos e truncamento. Calcule o erro relativo percentual.

(b) Repita (**a**), mas expresse y como

$$y = ((x - 7)x + 8)x - 0{,}35$$

Calcule o erro e compare com a parte (**a**).

3.8 Calcule a memória de acesso randômica (RAM) em megabytes necessária para armazenar uma tabela multidimensional que tem $20 \times 40 \times 120$. Essa tabela está em dupla precisão e cada valor requer uma palavra de 64 bytes. Lembre que uma palavra de 64 bytes = 8 bytes e 1 kilobyte = 2^{10} bytes. Suponha que o índice começa em 1.

3.9 Determine o número de termos necessários para aproximar $\cos x$ até 8 algarismos significativos usando a aproximação por série de Maclaurin

$$\cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots$$

Calcule a aproximação usando um valor de $x = 0{,}3\pi$. Escreva um programa para determinar o seu resultado.

3.10 Use aritmética com 5 algarismos significativos com truncamento para determinar as raízes da seguinte equação com as Equações (3.12) e (3.13)

$$x^2 - 5000.002x + 10$$

Calcule os erros relativos percentuais para o seu resultado.

3.11 Como pode o épsilon da máquina ser empregado para formular um critério de parada ε_s para os seus programas? Dê um exemplo.

CAPÍTULO 4

Erros de Truncamento e Séries de Taylor

Os *erros de truncamento* são aqueles que resultam do uso de uma aproximação no lugar de um procedimento matemático exato. Por exemplo, no Capítulo 1 aproximamos a derivada da velocidade de um pára-quedista em queda livre por uma equação de diferenças divididas da forma [Equação (1.11)]

$$\frac{dv}{dt} \cong \frac{\Delta v}{\Delta t} = \frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i} \tag{4.1}$$

Um erro de truncamento foi introduzido na solução numérica porque a equação de diferenças apenas aproxima o valor verdadeiro da derivada (lembre-se da Figura 1.4). Para obter uma percepção das propriedades de tais erros, será abordada agora uma formulação matemática que é amplamente usada nos métodos numéricos para expressar uma função de forma aproximada — a série de Taylor.[1]

4.1 A SÉRIE DE TAYLOR

O teorema de Taylor (Quadro 4.1) e sua fórmula associada, a série de Taylor, são de grande valia no estudo dos métodos numéricos. Em essência, a *série de Taylor* fornece um meio para prever o valor da função em um ponto em termos do valor da função e suas derivadas em um outro ponto. Em particular, o teorema afirma que qualquer função lisa pode ser aproximada por um polinômio.

Uma maneira útil de ganhar intuição sobre a série de Taylor é construí-la termo a termo. Por exemplo, o primeiro termo na série é

$$f(x_{i+1}) \cong f(x_i) \tag{4.2}$$

Essa relação, chamada de *aproximação de ordem zero*, indica que o valor de f no novo ponto é o mesmo que seu valor no ponto antigo. Este resultado faz sentido do ponto de vista intuitivo, pois se x_i e x_{i+1} estiverem perto um do outro, é provável que o novo valor seja parecido com o anterior.

A Equação (4.2) fornece uma estimativa perfeita se a função que estiver sendo aproximada for de fato constante. Entretanto, se a função variar de alguma forma no intervalo, termos adicionais da série de Taylor são necessários para fornecer uma estimativa melhor. Por exemplo, a *aproximação de primeira ordem* é deduzida adicionando-se mais um termo para obter

$$f(x_{i+1}) \cong f(x_i) + f'(x_i)(x_{i+1} - x_i) \tag{4.3}$$

O termo adicional de primeira ordem consiste em uma inclinação $f'(x_i)$ multiplicada pela distância entre x_i e x_{i+1}. Assim, a expressão agora tem a forma de uma reta e é capaz de prever um aumento ou uma diminuição entre x_i e x_{i+1}.

Embora a Equação (4.3) possa prever uma variação, ela é exata apenas para retas ou

[1] N.R.T.: O autor não faz distinção de nomenclatura entre a série de Taylor e um polinômio de Taylor. O sentido fica claro a partir do contexto.

Quadro 4.1 Teorema de Taylor

Teorema de Taylor

Se a função f e suas primeiras $n + 1$ derivadas forem contínuas em um intervalo contendo a e x, então o valor da função em x é dado por

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f^{(3)}(a)}{3!}(x-a)^3 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n \quad \text{(Q4.1.1)}$$

onde o resto R_n é definido por

$$R_n = \int_a^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\,dt \quad \text{(Q4.1.2)}$$

onde t é uma variável muda. A Equação (Q4.1.1) é chamada de *série de Taylor ou fórmula de Taylor*. Se o resto for omitido, o lado direito da Equação (Q4.1.1) é a aproximação em polinômio de Taylor de $f(x)$. Em essência, o teorema afirma que qualquer função lisa pode ser aproximada por um polinômio.

A Equação (Q4.1.2) é apenas uma das maneiras, chamada de *forma integral*, nas quais o resto pode ser expresso. Uma formulação alternativa pode ser deduzida com base no teorema do valor médio para a integral.

Primeiro Teorema da Média para Integrais

Se uma função g for contínua (portanto, integrável) em um intervalo contendo a e x, então existe um ponto ξ entre a e x tal que

$$\int_a^x g(t)\,dt = g(\xi)(x-a) \quad \text{(Q4.13)}$$

Em outras palavras, esse teorema afirma que a integral pode ser representada por um valor médio para a função $g(\xi)$ vezes o comprimento do intervalo $x - a$. Como a média dever ocorrer entre os valores mínimo e máximo no intervalo, existe um ponto $x = \xi$ no qual a função assume o valor médio.

O primeiro teorema é, na realidade, um caso especial do segundo teorema do valor médio para integrais.

Segundo Teorema da Média para Integrais

Se as funções g e h forem contínuas (logo, integráveis) em um intervalo contendo a e x, e se h não mudar de sinal no intervalo, então existe um ponto ξ entre a e x tal que

$$\int_a^x g(t)h(t)\,dt = g(\xi)\int_a^x h(t)\,dt \quad \text{(Q4.1.4)}$$

Portanto, a Equação (Q4.1.3) é equivalente à Equação (Q4.1.4) com $h(t) = 1$.

O segundo teorema pode ser aplicado à Equação (Q4.1.2) com

$$g(t) = f^{(n+1)}(t) \qquad h(t) = \frac{(x-t)^n}{n!}$$

Quando t varia de a a x, $h(t)$ é contínua e não muda de sinal. Portanto, se $f^{(n+1)}(t)$ for contínua, então o teorema do valor médio para a integração é válido e

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$

Essa equação é conhecida como a *forma derivada* ou *de Lagrange* do resto.

tendências *lineares*. Portanto, um termo de *segunda ordem* é adicionado à série para capturar alguma da curvatura que a função possa apresentar:

$$f(x_{i+1}) \cong f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2!}(x_{i+1} - x_i)^2 \quad (4.4)$$

De forma similar, termos adicionais podem ser incluídos para se obter a expansão completa em série de Taylor:

$$f(x_{i+1}) = f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(x_i)}{2!}(x_{i+1} - x_i)^2 + \frac{f^{(3)}(x_i)}{3!}(x_{i+1} - x_i)^3 + \cdots + \frac{f^{(n)}(x_i)}{n!}(x_{i+1} - x_i)^n + R_n \quad (4.5)$$

Observe que, como a Equação (4.5) é uma série infinita, o sinal de igual substitui o sinal de aproximação usado nas Equações (4.2) a (4.4). O resto é incluído para representar todos os termos de $n + 1$ até infinito:

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x_{i+1} - x_i)^{n+1} \quad (4.6)$$

onde o subscrito n indica que este é o resto para a aproximação de ordem n e ξ é um valor de x que está em algum ponto entre x_i e x_{i+1}. Essa introdução do ξ é tão importante que uma seção inteira (Seção 4.1.1) será dedicada à sua dedução. Por agora, é suficiente reconhecer que existe um tal valor que fornece uma determinação exata do erro.

Em geral, é conveniente simplificar a série de Taylor definindo um tamanho do

passo $h = x_{i+1} - x_i$ e expressando a Equação (4.5) como

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f^{(3)}(x_i)}{3!}h^3 + \cdots + \frac{f^{(n)}(x_i)}{n!}h^n + R_n \tag{4.7}$$

onde o resto agora é

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1} \tag{4.8}$$

EXEMPLO 4.1 Aproximação de um Polinômio por Série de Taylor

Enunciado do Problema. Use expansões em séries de Taylor de ordem zero até ordem quatro para aproximar a função

$$f(x) = -0{,}1x^4 - 0{,}15x^3 - 0{,}5x^2 - 0{,}25x + 1{,}2$$

a partir de $x_i = 0$ com $h = 1$. Isto é, faça uma previsão do valor da função em $x_{i+1} = 1$.

Solução. Como se trata de uma função conhecida, pode-se calcular valores para $f(x)$ entre 0 e 1. Os resultados (Figura 4.1) indicam que a função começa em $f(0) = 1{,}2$ e então se curva para baixo até $f(1) = 0{,}2$. Logo, o valor verdadeiro que se tenta prever é 0,2.

A aproximação em série de Taylor com $n = 0$ é [Equação (4.2)]

$$f(x_{i+1}) \simeq 1{,}2$$

Então, como na Figura 4.1, a aproximação de ordem zero é uma constante. Usando essa formulação, obtém-se um erro de truncamento [lembre-se da Equação (3.2)] de

$$E_t = 0{,}2 - 1{,}2 = -1{,}0$$

em $x = 1$.

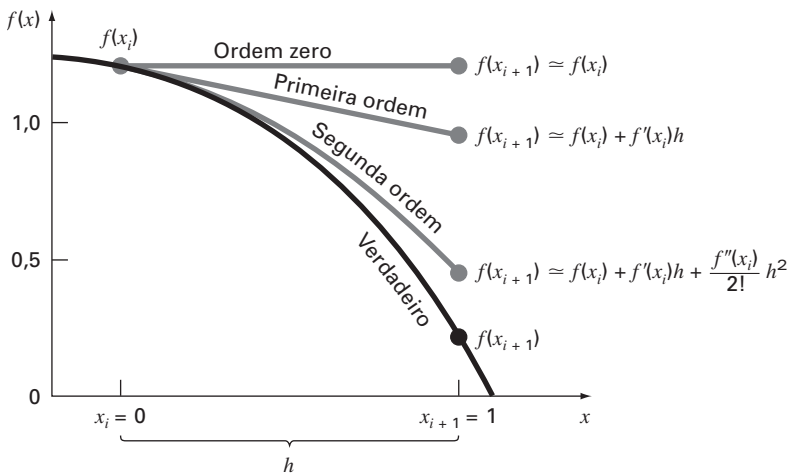
Para $n = 1$, a primeira derivada deve ser determinada e calculada em $x = 0$:

$$f'(0) = -0{,}4(0{,}0)^3 - 0{,}45(0{,}0)^2 - 1{,}0(0{,}0) - 0{,}25 = -0{,}25$$

Portanto, a aproximação de primeira ordem é [Equação (4.3)]

$$f(x_{i+1}) \simeq 1{,}2 - 0{,}25h$$

FIGURA 4.1
A aproximação de $f(x) = -0{,}1x^4 - 0{,}15x^3 - 0{,}5x^2 - 0{,}25x + 1{,}2$ em $x = 1$ por expansões em séries de Taylor de ordem zero, de primeira ordem e de segunda ordem.



que pode ser usada para calcular $f(1) = 0{,}95$. Conseqüentemente, a aproximação começa a capturar a trajetória voltada para baixo da função na forma de uma reta inclinada (Figura 4.1), o que resulta em uma redução do erro de truncamento para

$$E_t = 0{,}2 - 0{,}95 = -0{,}75$$

Para $n = 2$, a segunda derivada é calculada em $x = 0$:

$$f''(0) = -1{,}2(0{,}0)^2 - 0{,}9(0{,}0) - 1{,}0 = -1{,}0$$

Logo, de acordo com a Equação (4.4),

$$f(x_{i+1}) \simeq 1{,}2 - 0{,}25h - 0{,}5h^2$$

e substituindo $h = 1$, $f(1) = 0{,}45$. A inclusão da segunda derivada agora adicionou alguma curvatura para baixo resultando em uma estimativa melhor, como visto na Figura 4.1. O erro de truncamento foi reduzido ainda mais, para $0{,}2 - 0{,}45 = -0{,}25$.

Termos adicionais melhorariam a aproximação ainda mais. De fato, a inclusão da terceira e da quarta derivadas resulta exatamente na mesma equação do começo:

$$f(x) = 1{,}2 - 0{,}25h - 0{,}5h^2 - 0{,}15h^3 - 0{,}1h^4$$

onde o termo de resto é

$$R_4 = \frac{f^{(5)}(\xi)}{5!}h^5 = 0$$

porque a quinta derivada de um polinômio de quarta ordem é nula. Conseqüentemente, a expansão em série de Taylor até a quarta derivada fornece uma estimativa exata em $x_{i+1} = 1$:

$$f(1) = 1{,}2 - 0{,}25(1) - 0{,}5(1)^2 - 0{,}15(1)^3 - 0{,}1(1)^4 = 0{,}2$$

Em geral, a expansão em série de Taylor de ordem n será exata para um polinômio de grau n. Para outras funções diferenciáveis (portanto, contínuas), como exponenciais e funções senoidais, um número finito de termos não resultará em uma estimativa exata. Cada termo adicional irá contribuir com alguma melhora, ainda que pequena, para a aproximação. Esse comportamento será ilustrado no Exemplo 4.2. Apenas se um número infinito de termos for somado a série fornecerá um valor exato.[2]

Embora o que foi dito seja verdadeiro, o valor prático da expansão em série de Taylor é que, na maioria dos casos, a inclusão de apenas poucos termos resultará em uma aproximação que será suficientemente próxima do valor verdadeiro para fins práticos. A avaliação de quantos termos são necessários para ficar "suficientemente próximo" é baseada no resto da expansão. Lembre-se de que o resto tem a forma geral da Equação (4.8). Essa relação tem dois grandes defeitos. Primeiro, ξ não é conhecido exatamente, mas simplesmente está em algum lugar entre x_i e x_{i+1}. Segundo, para se calcular a Equação (4.8), é preciso determinar a $(n + 1)$-ésima derivada de $f(x)$. Para fazer isso, é necessário conhecer $f(x)$. Entretanto, se $f(x)$ fosse conhecida, não seria preciso fazer a expansão em série de Taylor no contexto atual!

Apesar desse dilema, a Equação (4.8) ainda é útil para se ganhar intuição sobre os erros de truncamento, porque, *de fato*, há controle sobre o termo h na equação. Em outras palavras, é possível escolher quão longe de x se quer calcular $f(x)$ e controlar o número de termos incluídos na expansão. Conseqüentemente, a Equação (4.8) em geral é expressa como

$$R_n = O(h^{n+1})$$

onde a nomenclatura $O(h^{n+1})$ significa que o erro de truncamento é da ordem de h^{n+1}. Ou seja, o erro é proporcional ao tamanho do passo h elevado à (n + 1)-ésima potência. Embora essa aproximação não tenha nenhuma implicação no valor da derivada que multiplica h^{n+1},

[2] N.R.T.: Na verdade, isto só é válido para as funções cujas séries de Taylor convergem para ele, ou seja, para funções reais analíticas. As funções deficiárias que ocorrem nas aplicações, em geral têm essa propriedade.

ela é extremamente útil para julgar o erro comparativo de métodos numéricos baseados em expansões em séries de Taylor. Por exemplo, se o erro for $O(h)$, dividir o tamanho do passo por dois fará com que o erro seja dividido por dois. Por outro lado, se o erro for da ordem de $O(h^2)$, dividir o tamanho do passo por dois fará com que o erro seja dividido por quatro.

Em geral, pode-se supor que o erro de truncamento diminui com a adição de termos na série de Taylor. Em muitos casos, se h for suficientemente pequeno, o primeiro e alguns outros termos de baixa ordem geralmente dão conta de uma percentagem desproporcionalmente alta dos erros. Assim, apenas poucos termos serão necessários para se obter uma estimativa adequada. Essa propriedade é ilustrada no exemplo a seguir.

EXEMPLO 4.2

Uso da Expansão em Série de Taylor para Aproximar uma Função com um Número Infinito de Derivadas

Enunciado do Problema. Use expansões em séries de Taylor com $n = 0$ até 6 para aproximar $f(x) = \cos x$ em $x_{i+1} = \pi/3$ com base no valor de $f(x)$ e suas derivadas em $x_i = \pi/4$. Observe que isso significa que $h = \pi/3 - \pi/4 = \pi/12$.

Solução. Como visto no Exemplo 4.1, o conhecimento da função verdadeira significa que é possível determinar o valor correto de $f(\pi/3) = 0{,}5$.

A aproximação de ordem 0 é [Equação (4.3)]

$$f\left(\frac{\pi}{3}\right) \cong \cos\left(\frac{\pi}{4}\right) = 0{,}707106781$$

o que representa um erro porcentual relativo de

$$\varepsilon_t = \frac{0{,}5 - 0{,}707106781}{0{,}5} 100\% = -41{,}4\%$$

Para a aproximação de primeira ordem, adiciona-se o termo da primeira derivada, $f'(x) = -\text{sen } x$:

$$f\left(\frac{\pi}{3}\right) \cong \cos\left(\frac{\pi}{4}\right) - \text{ sen}\left(\frac{\pi}{4}\right)\left(\frac{\pi}{12}\right) = 0{,}521986659$$

que tem $\varepsilon_t = -4{,}40\%$.

Para a aproximação de segunda ordem, adiciona-se o termo da segunda derivada, $f''(x) = -\cos x$:

$$f\left(\frac{\pi}{3}\right) \cong \cos\left(\frac{\pi}{4}\right) - \text{ sen}\left(\frac{\pi}{4}\right)\left(\frac{\pi}{12}\right) - \frac{\cos(\pi/4)}{2}\left(\frac{\pi}{12}\right)^2 = 0{,}497754491$$

com $\varepsilon_t = 0{,}449\%$. Logo, a inclusão de um termo adicional resultou em uma melhora na estimativa.

Pode-se continuar o processo e listar os resultados, como na Tabela 4.1. Observe que as derivadas nunca se anulam, como aconteceu com o polinômio do Exemplo 4.1. Portanto, cada termo adicional resulta em alguma melhora na estimativa. Entretanto, observe

TABELA 4.1 Aproximação em série de Taylor de $f(x) = \cos x$ em $x_{i+1} = \pi/3$ usando o ponto base $\pi/4$. São mostrados os valores para diversas ordens (n) de aproximação.

Ordem n	$f^{(n)}(x)$	$f(\pi/3)$	ε_t
0	$\cos x$	0,707106781	−41,4
1	$-\text{sen } x$	0,521986659	−4,4
2	$-\cos x$	0,497754491	0,449
3	$\text{sen } x$	0,499869147	$2{,}62 \times 10^{-2}$
4	$\cos x$	0,500007551	$-1{,}51 \times 10^{-3}$
5	$-\text{sen } x$	0,500000304	$-6{,}08 \times 10^{-5}$
6	$-\cos x$	0,499999988	$2{,}44 \times 10^{-6}$

também que a maior parte da melhora acontece nos termos iniciais. Nesse caso, quando se tiver somado o termo de terceira ordem, o erro terá sido reduzido a $2{,}62 \times 10^{-2}$ %, o que significa que foram obtidos 99,973% do valor verdadeiro. Conseqüentemente, embora a adição de mais termos reduza mais ainda o erro, a melhora se torna desprezível.

4.1.1 O Resto na Expansão em Série de Taylor

Antes de ilustrar como a série de Taylor é realmente usada para estimar erros numéricos, é preciso explicar por que o argumento ξ foi incluído na Equação (4.8). Uma dedução matemática é apresentada no Quadro 4.1. Agora será desenvolvida uma exposição alternativa baseada em uma interpretação um pouco mais visual. Então, esse caso específico poderá ser estendido à formulação mais geral.

Suponha que a expansão em série de Taylor [Equação(4.7)] seja truncada depois do termo de ordem zero para se obter

$$f(x_{i+1}) \cong f(x_i)$$

Uma descrição visual dessa previsão de ordem zero é mostrada na Figura 4.2. O resto — ou erro — dessa previsão, que também está mostrado na figura, consiste na série infinita dos termos truncados:

$$R_0 = f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f^{(3)}(x_i)}{3!}h^3 + \cdots$$

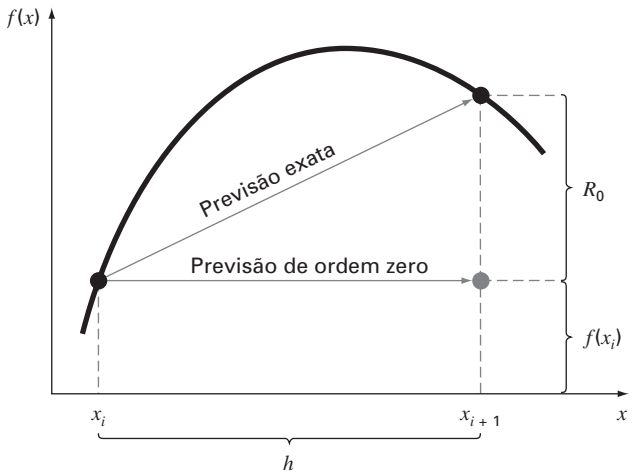
É óbvia a inconveniência de lidar com o resto nessa forma de série infinita. Uma simplificação seria truncar o próprio resto, como em

$$R_0 \cong f'(x_i)h \tag{4.9}$$

Embora, como afirmado na seção anterior, as derivadas de ordem mais baixa em geral sejam responsáveis por parte maior do resto do que os termos de ordem mais alta, esse resultado ainda não é exato por causa dos termos de segunda ordem e de ordens mais altas desprezados. Essa "inexatidão" é representada pelo símbolo de igualdade aproximada ($\cong$) usado na Equação (4.9).

Uma simplificação alternativa que transforma a aproximação em uma equivalência é baseada em uma observação gráfica. Como na Figura 4.3, o teorema do valor médio para derivadas afirma que, se uma função $f(x)$ e sua primeira derivada forem contínuas em um intervalo da forma x_i a x_{i+1}, então existe pelo menos um ponto na função que tem uma

FIGURA 4.2
Descrição gráfica da previsão de ordem zero da série de Taylor e do resto.



inclinação, denotada por $f'(\xi)$, que é paralela à reta ligando $f(x_i)$ e $f(x_{i+1})$. O parâmetro ξ designa o valor de x onde essa inclinação ocorre (Figura 4.3). Uma ilustração física desse teorema é que, se você viajar entre dois pontos com uma dada velocidade média, existirá pelo menos um instante durante o percurso no qual você estará movendo-se na velocidade média.

Usando esse teorema, é simples perceber que, como ilustrado pela Figura 4.3, a inclinação $f'(\xi)$ é igual à elevação R_0 dividida pelo passo h, ou

$$f'(\xi) = \frac{R_0}{h}$$

que pode ser reorganizado para fornecer

$$R_0 = f'(\xi)h \tag{4.10}$$

Assim, deduz-se a versão para ordem zero da Equação (4.8). As versões em ordem superior são meramente uma extensão lógica do raciocínio usado para deduzir a Equação (4.10). A versão de primeira ordem é

$$R_1 = \frac{f''(\xi)}{2!}h^2 \tag{4.11}$$

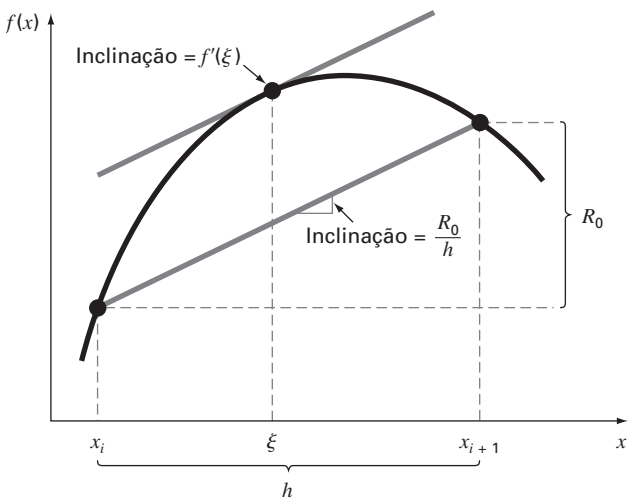
Para esse caso, o valor de ξ deve-se adequar ao valor de x correspondendo à segunda derivada que torna a Equação (4.11) exata. Versões similares de ordem superior podem ser deduzidas a partir da Equação (4.8).

4.1.2 Usando a Série de Taylor para Fazer uma Estimativa do Erro de Truncamento

Embora a série de Taylor vá ser extremamente útil na estimativa dos erros de truncamento em todo este livro, pode não ser claro para você como a expansão pode realmente ser aplicada no método numérico. De fato, já fizemos isso no exemplo do pára-quedista em queda livre. Lembre que o objetivo dos Exemplos 1.1 e 1.2 era prever a velocidade como uma função do tempo. Isto é, o interesse era determinar $v(t)$. Como especificado pela Equação (4.5), $v(t)$ pode ser expandida em uma série de Taylor:

$$v(t_{i+1}) = v(t_i) + v'(t_i)(t_{i+1} - t_i) + \frac{v''(t_i)}{2!}(t_{i+1} - t_i)^2 + \cdots + R_n \tag{4.12}$$

FIGURA 4.3
Descrição gráfica do teorema do valor médio para derivadas.



Agora, trunca-se a série depois do termo da primeira derivada:

$$v(t_{i+1}) = v(t_i) + v'(t_i)(t_{i+1} - t_i) + R_1 \tag{4.13}$$

A Equação (4.13) pode ser reescrita como

$$v'(t_i) = \underbrace{\frac{v(t_{i+1}) - v(t_i)}{t_{i+1} - t_i}}_{\substack{\text{Aproximação de} \\ \text{primeira ordem}}} - \underbrace{\frac{R_1}{t_{i+1} - t_i}}_{\substack{\text{Erro de} \\ \text{truncamento}}} \tag{4.14}$$

A primeira parte da Equação (4.14) é exatamente a mesma relação que foi usada para aproximar a derivada no Exemplo 1.2 [Equação (1.11)]. Entretanto, com a abordagem com a série de Taylor, obteve-se agora uma estimativa para o erro de truncamento associado com essa aproximação da derivada. Usando as Equações (4.6) e (4.14), obtém-se

$$\frac{R_1}{t_{i+1} - t_i} = \frac{v''(\xi)}{2!}(t_{i+1} - t_i) \tag{4.15}$$

ou

$$\frac{R_1}{t_{i+1} - t_i} = O(t_{i+1} - t_i) \tag{4.16}$$

Assim, a estimativa da derivada [Equação (1.11) ou a primeira parte da Equação (4.14)] tem um erro de truncamento da ordem de $t_{i+1} - t_i$. Em outras palavras, o erro na aproximação da derivada deveria ser proporcional ao tamanho do passo. Conseqüentemente, se o tamanho do passo for dividido por dois, deve-se esperar que o erro na derivada seja dividido por dois.

EXEMPLO 4.3

O Efeito da Não-Linearidade e do Tamanho do Passo na Aproximação em Série de Taylor

Enunciado do problema. A Figura 4.4 é o gráfico da função

$$f(x) = x^m \tag{E4.3.1}$$

para $m = 1, 2, 3,$ e 4 no intervalo de $x = 1$ a 2. Observe que para $m = 1$ a função é linear, e conforme m aumenta, mais curvatura ou não-linearidade é introduzida na função. Use a série de Taylor de primeira ordem para aproximar essa função para vários valores do expoente m e do tamanho do passo h.

Solução. A Equação (E4.3.1) pode ser aproximada por uma expansão em série de Taylor de primeira ordem, como em

$$f(x_{i+1}) = f(x_i) + mx_i^{m-1}h \tag{E4.3.2}$$

que tem um resto

$$R_1 = \frac{f''(x_i)}{2!}h^2 + \frac{f^{(3)}(x_i)}{3!}h^3 + \frac{f^{(4)}(x_i)}{4!}h^4 + \cdots$$

Primeiro, pode-se examinar como a aproximação se comporta à medida que m aumenta — isto é, à medida que a função se torna mais não-linear. Para $m = 1$, o valor real da função em $x = 2$ é 2.
A série de Taylor fornece

$$f(2) = 1 + 1(1) = 2$$

e

$$R_1 = 0$$

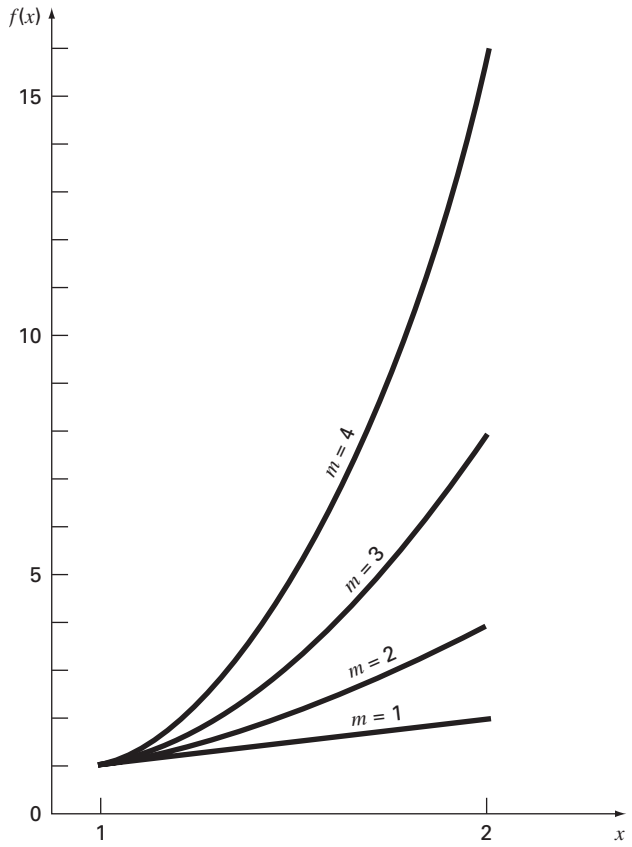


FIGURA 4.4
Gráfico da função $f(x) = x^m$ para $m = 1, 2, 3$ e 4. Observe que a função se torna mais não-linear à medida que m aumenta.

O resto é zero porque a segunda derivada e as derivadas de ordem mais alta de uma função linear são nulas. Assim, como esperado, a expansão em série de Taylor de primeira ordem é perfeita quando a função envolvida for linear.

Para $m = 2$, o valor real é $f(2) = 2^2 = 4$. A aproximação em série de Taylor de primeira ordem é

$$f(2) = 1 + 2(1) = 3$$

e

$$R_1 = \tfrac{2}{2}(1)^2 + 0 + 0 + \cdots = 1$$

Assim, como a função é uma parábola, a aproximação por uma reta resulta em uma discrepância. Observe que o resto é determinado exatamente.

Para $m = 3$, o valor real é $f(2) = 2^3 = 8$. A aproximação em série de Taylor é

$$f(2) = 1 + 3(1)^2(1) = 4$$

e

$$R_1 = \tfrac{6}{2}(1)^2 + \tfrac{6}{6}(1)^3 + 0 + 0 + \cdots = 4$$

Novamente, existe uma discrepância que pode ser determinada exatamente a partir da série de Taylor.

Para $m = 4$, o valor real é $f(2) = 2^4 = 16$. A aproximação em série de Taylor é

$$f(2) = 1 + 4(1)^3(1) = 5$$

e

$$R_1 = \tfrac{12}{2}(1)^2 + \tfrac{24}{6}(1)^3 + \tfrac{24}{24}(1)^4 + 0 + 0 + \cdots = 11$$

Com base nesses quatro casos, observa-se que R_1 cresce à medida que a função se torna mais não-linear. Além disso, R_1 representa exatamente essa discrepância, porque a Equação (E4.3.1) é um simples monômio com um número finito de derivadas não-nulas. Isso permite a determinação completa do resto da série de Taylor.

A seguir, examinaremos a Equação (E4.3.2) para o caso $m = 4$ e observaremos como R_1 varia à medida que o tamanho do passo h varia. Para $m = 4$, a Equação (E4.3.2) é

$$f(x + h) = f(x) + 4x_i^3 h$$

Se $x = 1, f(1) = 1$ e essa equação pode ser expressa por

$$f(1 + h) = 1 + 4h$$

com um resto de

$$R_1 = 6h^2 + 4h^3 + h^4$$

Isso leva à conclusão de que a discrepância irá diminuir conforme h for reduzido. Além disso, para valores suficientemente pequenos de h, o erro deve tornar-se proporcional a h^2. Ou seja, se h for dividido por dois, o erro será dividido por quatro. Esse comportamento é confirmado pela Tabela 4.2 e pela Figura 4.5.

Assim, conclui-se que o erro da aproximação por polinômio de Taylor de primeira ordem diminui quando m se aproxima de 1 e quando h diminui. Intuitivamente, isso significa que a série de Taylor se torna mais precisa quando a função que está sendo aproximada se torna mais parecida com uma reta no intervalo de interesse — o que pode ser conseguido ou reduzindo-se o tamanho do intervalo ou "retificando-se" a função pela redução de m. Obviamente, a última opção, em geral, não está disponível no mundo real, porque as funções que analisamos são determinadas tipicamente pelo contexto do problema físico. Conseqüentemente, não há controle de sua falta de linearidade, e o único recurso é reduzir o tamanho do passo ou incluir termos adicionais na expansão em série de Taylor.

TABELA 4.2 Comparação do valor exato da função $f(x) = x^4$ com a aproximação em série de Taylor de primeira ordem. Tanto a função como a aproximação estão calculadas em $x + h$, onde $x = 1$.

h	Verdadeiro	Aproximação de Primeira Ordem	R_1
1	16	5	11
0,5	5,0625	3	2,0625
0,25	2,441406	2	0,441406
0,125	1,601807	1,5	0,101807
0,0625	1,274429	1,25	0,024429
0,03125	1,130982	1,125	0,005982
0,015625	1,063980	1,0625	0,001480

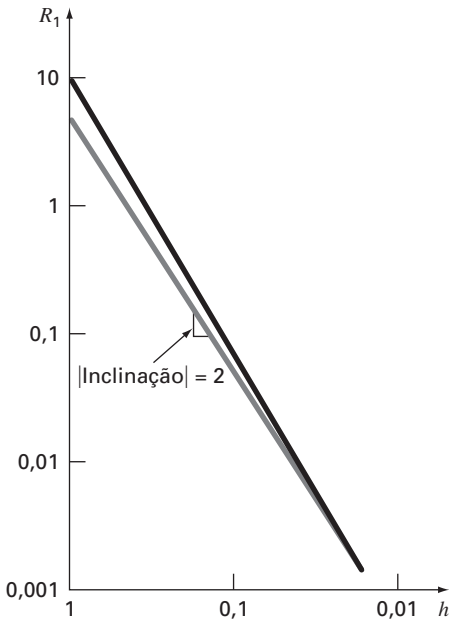


FIGURA 4.5
Gráfico log-log do resto R_1 da aproximação em série de Taylor de primeira ordem da função $f(x) = x^4$ em função do tamanho do passo h. Uma reta com coeficiente angular 2 também é mostrada para indicar que, conforme h decresce, o erro se torna proporcional a h^2.

4.1.3 Derivação Numérica

A Equação (4.14) tem uma designação formal nos métodos numéricos — ela é chamada uma *diferença dividida finita*. Ela pode ser representada em geral por

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} + O(x_{i+1} - x_i) \tag{4.17}$$

ou

$$f'(x_i) = \frac{\Delta f_i}{h} + O(h) \tag{4.18}$$

onde Δf_i é conhecida como *primeira diferença progressiva* e h é chamado de tamanho do passo, isto é, o comprimento do intervalo no qual a aproximação é feita. Ela é chamada diferença "progressiva" porque utiliza os dados em i e $i + 1$ para estimar a derivada (Figura 4.6*a*). O termo $\Delta fi/n$ todo é chamado de *primeira diferença dividida finita*.

Essa diferença dividida progressiva é apenas uma das muitas que podem ser deduzidas a partir da série de Taylor para aproximar derivadas numericamente. Por exemplo, aproximações com diferenças regressivas e centradas da primeira derivada podem ser deduzidas de forma similar à dedução da Equação (4.14). A primeira utiliza valores em x_{i-1} e x_i (Figura 4.6*b*), enquanto a última utiliza valores que são igualmente espaçados em torno do ponto no qual a derivada será estimada (Figura 4.6*c*). Aproximações mais acuradas da primeira derivada podem ser deduzidas incluindo-se termos de ordem mais alta na série de Taylor. Finalmente, todas essas versões também podem ser deduzidas para derivadas de segunda e terceira ordem ou superiores. A seção a seguir fornece um breve resumo ilustrando como alguns desses casos são deduzidos.

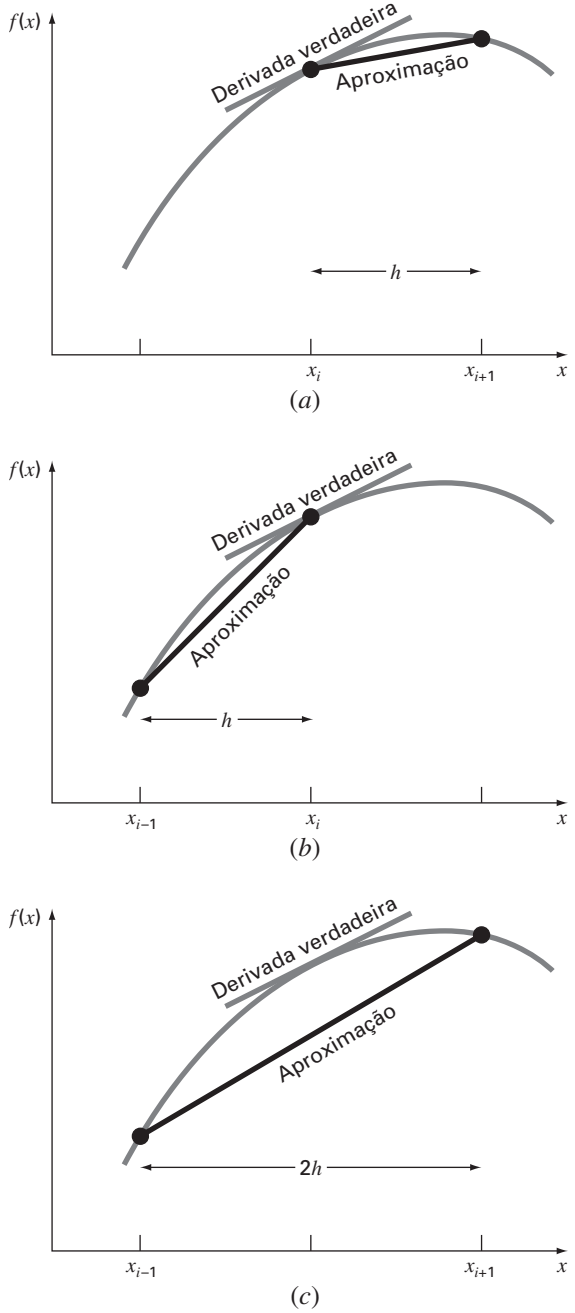
Aproximação da Primeira Derivada por Diferença Regressiva. A série de Taylor pode ser expandida regressivamente para calcular um valor anterior com base no valor atual, como em

$$f(x_{i-1}) = f(x_i) - f'(x_i)h + \frac{f''(x_i)}{2!}h^2 - \cdots \tag{4.19}$$

Truncando essa equação depois da primeira derivada e reorganizando, obtém-se

$$f'(x_i) \cong \frac{f(x_i) - f(x_{i-1})}{h} = \frac{\nabla f_1}{h} \tag{4.20}$$

FIGURA 4.6
Descrição gráfica das aproximações da primeira derivada por diferenças divididas finitas (*a*) progressiva, (*b*) regressiva e (*c*) centrada.



onde o erro é $O(h)$ e ∇f_i é conhecido como primeira diferença regressiva. Consulte a Figura 4.6*b* para uma representação gráfica.

Aproximação da Primeira Derivada por Diferença Centrada. Uma terceira forma de aproximar a primeira derivada é subtrair a Equação (4.19) da expansão em série de Taylor progressiva:

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \cdots \tag{4.21}$$

para obter

$$f(x_{i+1}) = f(x_{i-1}) + 2f'(x_i)h + \frac{2f^{(3)}(x_i)}{3!}h^3 + \cdots$$

que pode ser reescrita como

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} - \frac{f^{(3)}(x_i)}{6}h^2 - \cdots$$

ou

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} - O(h^2) \tag{4.22}$$

A Equação (4.22) é uma representação da primeira derivada como *diferença centrada*. Observe que o erro de truncamento é da ordem de h^2, em contraste com as aproximações progressiva e regressiva, em que era da ordem de h. Conseqüentemente, a análise da série de Taylor fornece a informação prática de que a diferença centrada é uma representação mais acurada da derivada (Figura 4.6*c*). Por exemplo, se o tamanho do passo fosse dividido por dois usando-se uma diferença progressiva ou regressiva, aproximadamente o erro de truncamento seria dividido por dois, enquanto que, para a diferença centrada, o erro seria dividido por quatro.

EXEMPLO 4.4 Aproximação de Derivadas por Diferenças Divididas Finitas

Enunciado do Problema. Use aproximações por diferenças progressiva e regressiva de $O(h)$ e uma aproximação por diferença centrada de $O(h^2)$ para fazer uma estimativa da primeira derivada de

$$f(x) = -0{,}1x^4 - 0{,}15x^3 - 0{,}5x^2 - 0{,}25x + 1{,}2$$

em $x = 0{,}5$ usando um tamanho de passo $h = 0{,}5$. Repita os cálculos usando $h = 0{,}25$. Observe que a derivada pode ser calculada diretamente por

$$f'(x) = -0{,}4x^3 - 0{,}45x^2 - 1{,}0x - 0{,}25$$

e pode ser usada para calcular o valor verdadeiro como sendo $f'(0{,}5) = -0{,}9125$.

Solução. Para $h = 0{,}5$, a função pode ser empregada para determinar

$$\begin{aligned} x_{i-1} &= 0 & f(x_{i-1}) &= 1{,}2 \\ x_i &= 0{,}5 & f(x_i) &= 0{,}925 \\ x_{i+1} &= 1{,}0 & f(x_{i+1}) &= 0{,}2 \end{aligned}$$

Esses valores podem ser usados para calcular a diferença dividida progressiva [Equação (4.17)],

$$f'(0{,}5) \cong \frac{0{,}2 - 0{,}925}{0{,}5} = -1{,}45 \qquad |\varepsilon_t| = 58{,}9\%$$

a diferença dividida regressiva [Equação (4.20)],

$$f'(0,5) \cong \frac{0,925 - 1,2}{0,5} = -0,55 \qquad |\varepsilon_t| = 39,7\%$$

e a diferença dividida centrada [Equação (4.22)],

$$f'(0,5) \cong \frac{0,2 - 1,2}{1,0} = -1,0 \qquad |\varepsilon_t| = 9,6\%$$

Para $h = 0,25$,

$$\begin{aligned} x_{i-1} &= 0,25 \qquad f(x_{i-1}) = 1,10351563 \\ x_i &= 0,5 \qquad f(x_i) = 0,925 \\ x_{i+1} &= 0,75 \qquad f(x_{i+1}) = 0,63632813 \end{aligned}$$

que podem ser usadas para se calcular a diferença dividida progressiva,

$$f'(0,5) \cong \frac{0,63632813 - 0,925}{0,25} = -1,155 \qquad |\varepsilon_t| = 26,5\%$$

a diferença dividida regressiva,

$$f'(0,5) \cong \frac{0,925 - 1,10351563}{0,25} = -0,714 \qquad |\varepsilon_t| = 21,7\%$$

e a diferença dividida centrada,

$$f'(0,5) \cong \frac{0,63632813 - 1,10351563}{0,5} = -0,934 \qquad |\varepsilon_t| = 2,4\%$$

Para ambos os tamanhos do passo, a aproximação por diferença centrada é mais acurada que as aproximações progressiva e regressiva. Além disso, como previsto pelo teorema de Taylor, dividir o tamanho do passo por dois aproximadamente divide o erro por dois nas diferenças progressiva e regressiva e divide o erro por quatro na diferença centrada.

Aproximações por Diferenças Finitas de Derivadas Superiores. Além das primeiras derivadas, a expansão em série de Taylor pode ser usada para deduzir estimativas numéricas das derivadas mais altas. Para fazer isso, escreve-se uma expansão em série de Taylor progressiva para $f(x_{i+2})$ em termos de $f(x_i)$:

$$f(x_{i+2}) = f(x_i) + f'(x_i)(2h) + \frac{f''(x_i)}{2!}(2h)^2 + \cdots \tag{4.23}$$

A Equação (4.21) pode ser multiplicada por 2 e subtraída da Equação (4.23) para fornecer

$$f(x_{i+2}) - 2f(x_{i+1}) = -f(x_i) + f''(x_i)h^2 + \cdots$$

que pode ser reescrita como

$$f''(x_i) = \frac{f(x_{i+2}) - 2f(x_{i+1}) + f(x_i)}{h^2} + O(h) \tag{4.24}$$

Essa relação é chamada *segunda diferença dividida finita progressiva*. Manipulações similares podem ser usadas para deduzir uma versão regressiva

$$f''(x_i) = \frac{f(x_i) - 2f(x_{i-1}) + f(x_{i-2})}{h^2} + O(h)$$

e uma versão centrada

$$f''(x_i) = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1})}{h^2} + O(h^2)$$

Como no caso das aproximações da primeira derivada, o caso centrado é mais acurado. Observe também que a versão centrada pode ser expressa alternativamente por

$$f''(x_i) \cong \frac{\dfrac{f(x_{i+1}) - f(x_i)}{h} - \dfrac{f(x_i) - f(x_{i-1})}{h}}{h}$$

Assim, da mesma forma como a derivada segunda é uma derivada da derivada, a aproximação pela segunda diferença dividida é a diferença de duas diferenças divididas.

O tópico de derivação numérica será retomado no Capítulo 23. Ele foi introduzido neste ponto porque é um exemplo muito bom de por que a série de Taylor é importante nos métodos numéricos. Além disso, muitas das fórmulas introduzidas nesta seção serão usadas antes do Capítulo 23.

4.2 PROPAGAÇÃO DE ERROS

O propósito desta seção é estudar como os erros nos números podem propagar-se através de funções matemáticas. Por exemplo, ao se multiplicar dois números que têm erros, gostaríamos de fazer uma estimativa do erro do produto.

4.2.1 Funções de uma Única Variável

Suponha que tenhamos uma função $f(x)$ que depende de uma única variável independente x, e que $\tilde{x}$ seja uma aproximação de x. Então, gostaríamos de estimar o efeito da discrepância entre x e $\tilde{x}$ no valor da função. Isto é, gostaríamos de estimar

$$\Delta f(\tilde{x}) = |f(x) - f(\tilde{x})|$$

O problema de calcular $\Delta f(\tilde{x})$ é que $f(x)$ é desconhecida, já que x é desconhecido. Pode-se superar essa dificuldade se $\tilde{x}$ estiver próximo de x e $f(\tilde{x})$ for contínua e diferenciável. Se tais condições forem válidas, a série de Taylor pode ser usada para calcular $f(x)$ próximo de $f(\tilde{x})$, como em

$$f(x) = f(\tilde{x}) + f'(\tilde{x})(x - \tilde{x}) + \frac{f''(\tilde{x})}{2}(x - \tilde{x})^2 + \cdots$$

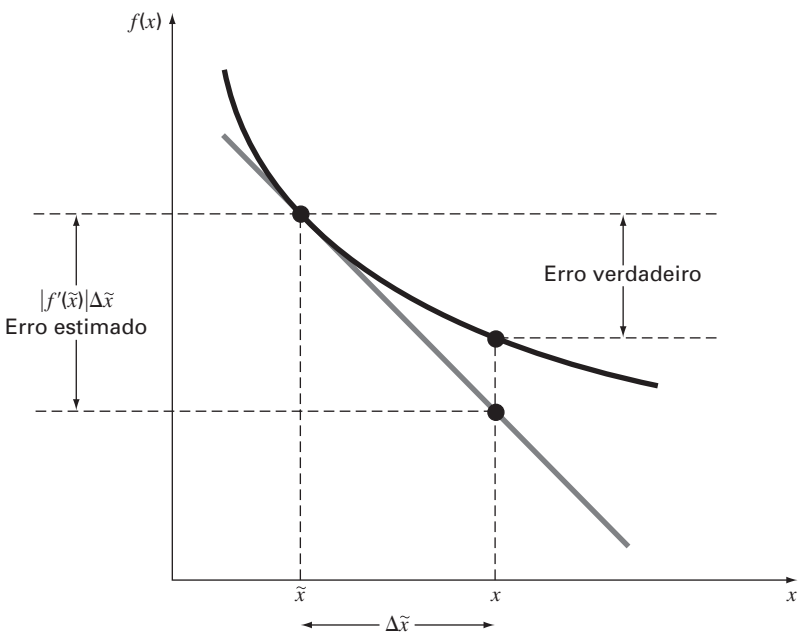


FIGURA 4.7
Descrição gráfica da propagação de erro de primeira ordem.

Desprezando os termos de grau maior ou igual a dois e reorganizando, obtemos

$$f(x) - f(\tilde{x}) \cong f'(\tilde{x})(x - \tilde{x})$$

ou

$$\Delta f(\tilde{x}) = |f'(\tilde{x})|(x - \tilde{x}) \tag{4.25}$$

onde $\Delta f(\tilde{x}) = |f(x) - f(\tilde{x})|$ representa uma estimativa do erro da função e $\Delta\tilde{x} = |x - \tilde{x}|$ representa uma estimativa do erro de x. A Equação (4.25) fornece recursos para aproximar o erro em $f(x)$ dadas a derivada da função e uma estimativa do erro da variável independente. A Figura 4.7 é uma ilustração gráfica da operação.

EXEMPLO 4.5 Propagação de Erro em uma Função de uma Única Variável

Enunciado do Problema. Dado um valor de $\tilde{x} = 2{,}5$ com um erro de $\Delta\tilde{x} = 0{,}01$, faça uma estimativa do erro resultante na função $f(x) = x^3$.

Solução. Usando a Equação (4.25),

$$\Delta f(\tilde{x}) \cong 3(2{,}5)^2(0{,}01) = 0{,}1875$$

Como $f(2{,}5) = 15{,}625$, prevê-se que

$$f(2{,}5) = 15{,}625 \pm 0{,}1875$$

ou que o valor verdadeiro deve ficar entre 15,4375 e 15,8125. De fato, se x for realmente 2,49, a função pode ser calculada como sendo 15,4382, e se x for 2,51, ela será 15,8132. Para este caso, a análise de erro de primeira ordem fornece uma estimativa bastante próxima do erro verdadeiro.

4.2.2 Funções de Mais de Uma Variável

A abordagem precedente pode ser generalizada para funções que dependem de mais de uma variável independente. Consegue-se isso com a versão para diversas variáveis da série de Taylor. Por exemplo, se tivemos uma função de duas variáveis independentes u e v, a série de Taylor pode ser escrita como

$$\begin{aligned} f(u_{i+1}, v_{i+1}) = f(u_i, v_i) &+ \frac{\partial f}{\partial u}(u_{i+1} - u_i) + \frac{\partial f}{\partial v}(v_{i+1} - v_i) \\ &+ \frac{1}{2!}\left[\frac{\partial^2 f}{\partial u^2}(u_{i+1} - u_i)^2 + 2\frac{\partial^2 f}{\partial u \partial v}(u_{i+1} - u_i)(v_{i+1} - v_i)\right. \\ &\left. + \frac{\partial^2 f}{\partial v^2}(v_{i+1} - v_i)^2\right] + \cdots \end{aligned} \tag{4.26}$$

onde todas as derivadas parciais são calculadas no ponto base i. Se todos os termos de segunda ordem e de ordem mais alta forem desconsiderados, a Equação (4.26) pode ser reescrita como

$$\Delta f(\tilde{u}, \tilde{v}) = \left|\frac{\partial f}{\partial u}\right| \Delta\tilde{u} + \left|\frac{\partial f}{\partial v}\right| \Delta\tilde{v}$$

onde $\Delta\tilde{u}$ e $\Delta\tilde{v}$ são estimativas dos erros em u e v, respectivamente.

Para n variáveis independentes $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n$ tendo erros $\Delta\tilde{x}_1, \Delta\tilde{x}_2, \ldots, \Delta x_n$, vale a seguinte relação geral:

$$\Delta f(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) \cong \left|\frac{\partial f}{\partial x_1}\right| \Delta\tilde{x}_1 + \left|\frac{\partial f}{\partial x_2}\right| \Delta\tilde{x}_2 + \cdots + \left|\frac{\partial f}{\partial x_n}\right| \Delta\tilde{x}_n \tag{4.27}$$

EXEMPLO 4.6 Propagação de Erros em uma Função de Várias Variáveis

Enunciado do Problema. A deflexão y no topo do mastro de um barco a vela é

$$y = \frac{FL^4}{8EI}$$

onde F é uma carga lateral uniforme (lb/pés), L é a altura (pés), E é o módulo de elasticidade (lb/pés^2) e I é o momento de inércia (pés^4). Faça uma estimativa do erro em y considerando os seguintes dados:

$$\tilde{F} = 50 \text{ lb/pés} \qquad \Delta\tilde{F} = 2 \text{ lb/pés}$$
$$\tilde{L} = 30 \text{ pés} \qquad \Delta\tilde{L} = 0{,}1 \text{ pés}$$
$$\tilde{E} = 1{,}5 \times 10^8 \text{ lb/pés}^2 \qquad \Delta\tilde{E} = 0{,}01 \times 10^8 \text{ lb/pés}^2$$
$$\tilde{I} = 0{,}06 \text{ pés}^4 \qquad \Delta\tilde{I} = 0{,}0006 \text{ pés}^4$$

Solução. Usando a Equação (4.27) obtém-se

$$\Delta y(\tilde{F}, \tilde{L}, \tilde{E}, \tilde{I}) = \left|\frac{\partial y}{\partial F}\right| \Delta\tilde{F} + \left|\frac{\partial y}{\partial L}\right| \Delta\tilde{L} + \left|\frac{\partial y}{\partial E}\right| \Delta\tilde{E} + \left|\frac{\partial y}{\partial I}\right| \Delta\tilde{I}$$

ou

$$\Delta y(\tilde{F}, \tilde{L}, \tilde{E}, \tilde{I}) \cong \frac{\tilde{L}^4}{8\tilde{E}\tilde{I}}\Delta\tilde{F} + \frac{\tilde{F}\tilde{L}^3}{2\tilde{E}\tilde{I}}\Delta\tilde{L} + \frac{\tilde{F}\tilde{L}^4}{8\tilde{E}^2\tilde{I}}\Delta\tilde{E} + \frac{\tilde{F}\tilde{L}^4}{8\tilde{E}\tilde{I}^2}\Delta\tilde{I}$$

Substituindo os valores apropriados, obtém-se

$$\Delta y = 0{,}0225 + 0{,}0075 + 0{,}00375 + 0{,}005625 = 0{,}039375$$

Portanto, $y = 0{,}5625 \pm 0{,}039375$. Em outras palavras, y está entre 0,523125 e 0,601875 pés. A validade dessas estimativas pode ser verificada substituindo-se os valores extremos das variáveis na equação para gerar um mínimo exato de

$$y_{\min} = \frac{48(29{,}9)^4}{8(1{,}51 \times 10^8)0{,}0606} = 0{,}52407$$

e

$$y_{\max} = \frac{52(30{,}1)^4}{8(1{,}49 \times 10^8)0{,}0594} = 0{,}60285$$

Logo, as estimativas de primeira ordem são razoavelmente próximas dos valores exatos.

A Equação (4.27) pode ser usada para definir relações de propagação de erros para operações matemáticas comuns. Os resultados estão resumidos na Tabela 4.3. Recomenda-se a dedução dessas fórmulas como exercícios para casa.

4.2.3 Estabilidade e Condicionamento

O *condicionamento* de um problema matemático se relaciona com sua sensibilidade a variações em seus valores de entrada. Diz-se que um cálculo é *numericamente instável* se as incertezas dos valores de entrada são brutalmente aumentadas pelo método numérico.

Essas idéias podem ser estudadas usando-se a série de Taylor de primeira ordem

$$f(x) = f(\tilde{x}) + f'(\tilde{x})(x - \tilde{x})$$

Essa relação pode ser empregada para se estimar o *erro relativo* de $f(x)$ como em

$$\frac{f(x) - f(\tilde{x})}{f(x)} \cong \frac{f'(\tilde{x})(x - \tilde{x})}{f(\tilde{x})}$$

TABELA 4.3 Limitantes estimados dos erros associados com as operações matemáticas usuais utilizando números inexatos $\tilde{u}$ e $\tilde{v}$.

Operação		Erro Estimado
Adição	$\Delta(\tilde{u}+\tilde{v})$	$\Delta\tilde{u}+\Delta\tilde{v}$
Subtração	$\Delta(\tilde{u}-\tilde{v})$	$\Delta\tilde{u}+\Delta\tilde{v}$
Multiplicação	$\Delta(\tilde{u}\times\tilde{v})$	$\lvert\tilde{u}\rvert\Delta\tilde{v}+\lvert\tilde{v}\rvert\Delta\tilde{u}$
Divisão	$\Delta\left(\frac{\tilde{u}}{\tilde{v}}\right)$	$\frac{\lvert\tilde{u}\rvert\Delta\tilde{v}+\lvert\tilde{v}\rvert\Delta\tilde{u}}{\lvert\tilde{v}\rvert^2}$

O *erro relativo* de x é dado por

$$\frac{x-\tilde{x}}{\tilde{x}}$$

Um *número de condicionamento* pode ser definido como a razão desses erros relativos:

$$\text{Número de condicionamento} = \frac{\tilde{x}\,f'(\tilde{x})}{f(\tilde{x})} \tag{4.28}$$

O número de condicionamento fornece uma medida da extensão pela qual uma incerteza em x é ampliada por $f(x)$. Um valor 1 diz que o erro relativo da função é idêntico ao erro relativo em x. Um valor maior que 1 diz que o erro relativo foi ampliado, enquanto um valor menor do que 1 diz que foi atenuado. Funções com valores muito grandes são ditas *mal condicionadas*. Qualquer combinação de fatores na Equação (4.28) que aumente o valor numérico do número de condicionamento tenderá a aumentar as incertezas no cálculo de $f(x)$.

EXEMPLO 4.7 Número de Condicionamento

Enunciado do Problema. Calcule e interprete o número de condicionamento para

$$f(x)=\text{tg}x \qquad \text{para } \tilde{x}=\frac{\pi}{2}+0{,}1\left(\frac{\pi}{2}\right)$$

$$f(x)=\text{tg}x \qquad \text{para } \tilde{x}=\frac{\pi}{2}+0{,}01\left(\frac{\pi}{2}\right)$$

Solução. O número de condicionamento é calculado como

$$\text{Número de condicionamento} = \frac{\tilde{x}(1/\cos^2 x)}{\text{tg}\tilde{x}}$$

Para $\tilde{x}=\pi/2+0{,}1(\pi/2)$,

$$\text{Número de condicionamento} = \frac{1{,}7279(40.\,86)}{-6{,}314} = -11{,}2$$

Logo, a função é mal condicionada. Para $\tilde{x}=\pi/2+0{,}01(\pi/2)$, a situação é ainda pior:

$$\text{Número de condicionamento} = \frac{1{,}5865(4053)}{-63{,}66} = -101$$

Nesse caso, a principal causa do mal condicionamento parece ser a derivada. Isso faz sentido porque, na vizinhança de $\pi/2$, a tangente se aproxima de valores infinitos tanto positivos quanto negativos.

4.3 ERRO NUMÉRICO TOTAL

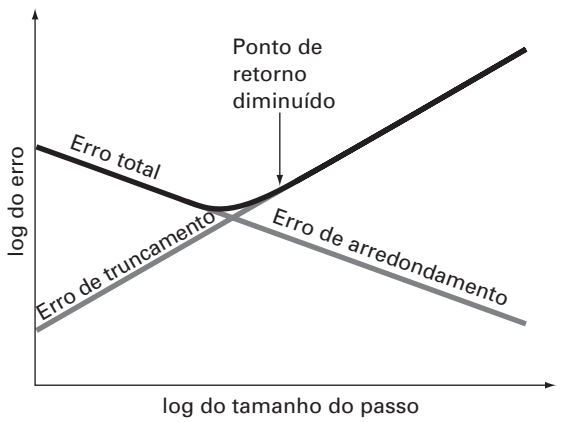
O *erro numérico total* é a soma dos erros de truncamento e de arredondamento. Em geral, a única maneira de minimizar os erros de arredondamento é aumentar o número de algarismos significativos do computador. Além disso, observa-se que os erros de arredondamento irão a*umentar* por causa de cancelamentos na subtração ou de um aumento do número de cálculos na análise. Em contraste, o Exemplo 4.4 mostrou que o erro de truncamento pode ser reduzido diminuindo o tamanho do passo. Como a diminuição do tamanho do passo pode levar a cancelamentos na subtração ou a um aumento nos cálculos, os erros de truncamento *diminuirão* enquanto os erros de arredondamento *aumentarão*. Então, encara-se o seguinte dilema: a estratégia para diminuir uma componente do erro total leva a um aumento da outra componente. Em um cálculo, pode-se, de modo concebível, diminuir o tamanho do passo para minimizar os erros de truncamento, apenas para descobrir que, fazendo isso, os erros de arredondamento começam a dominar a solução e o erro total cresce! Assim, o remédio se torna problema (Figura 4.8). Um desafio é determinar um tamanho de passo apropriado para um cálculo particular. Gostaríamos de escolher um tamanho de passo maior para diminuir o número de operações e os erros de arredondamento sem pagar o preço de grandes erros de truncamento. Se o erro total for como mostrado na Figura 4.8, o desafio é identificar o ponto de retorno diminuído onde os erros de arredondamento começam a neutralizar as vantagens da redução do tamanho do passo.

Nos casos reais, entretanto, tais situações são relativamente incomuns, porque a maioria dos computadores usa algarismos significativos suficientes para que os erros de arredondamento não predominem. Entretanto, elas de fato ocorrem algumas vezes e sugerem um tipo de "princípio da incerteza numérico" que coloca um limite absoluto na acurácia que pode ser obtida usando um certo método numérico computadorizado.

4.3.1 Controle dos Erros Numéricos

Na maioria dos casos práticos, não se sabe o erro exato associado com os métodos numéricos. A exceção, obviamente, é quando se obtém a solução exata que torna a aproximação numérica desnecessária. Então, na maior parte das aplicações da engenharia, é necessário contentar-se com alguma estimativa do erro nos cálculos.

FIGURA 4.8
Uma descrição gráfica do balanço entre erros de arredondamento e de truncamento que ocorrem algumas vezes na utilização de um método numérico. É mostrado o ponto de retorno diminuído, onde o erro de arredondamento começa a neutralizar as vantagens da redução do tamanho do passo.



Não existe uma abordagem geral e sistemática de obter estimativas dos erros numéricos para todos os problemas. Em muitos casos, as estimativas de erros são baseadas na experiência e no julgamento do engenheiro.

Embora a análise de erro seja até certo ponto uma arte, existem diversas diretrizes práticas de programação que aqui são sugeridas. Primeiro e principalmente, deve-se evitar subtrair dois números quase iguais. Perda de significado quase sempre ocorre quando isso é feito. Algumas vezes, o profissional pode reorganizar ou reformular o problema para evitar o cancelamento da subtração. Se isso não for possível, ele pode usar aritmética com precisão estendida. Além disso, quando somando ou subtraindo números, é melhor ordená-los e trabalhar com os números menores primeiro, pois tal conduta evita perda de significado.

Além dessas sugestões computacionais, pode-se tentar prever os erros numéricos totais usando formulações teóricas. A série de Taylor é a ferramenta básica na análise tanto dos erros de arredondamento quanto dos de truncamento. Foram apresentados diversos exemplos neste capítulo. A previsão do erro numérico total é muito complicada, mesmo para problemas de tamanho moderado, e tende a ser pessimista. Logo, usualmente é tentada apenas para tarefas em pequena escala.

A tendência é ir à frente com os cálculos numéricos e tentar estimar a acurácia dos resultados. Isso, algumas vezes, pode ser feito vendo se os resultados satisfazem alguma condição ou equação como uma verificação. Ou pode ser possível substituir os resultados de volta na equação original para verificar se ela realmente é satisfeita.

Finalmente, é preciso estar preparado para fazer experiências numéricas para aumentar a percepção dos erros numéricos e de possíveis problemas mal condicionados. Tais experiências podem envolver a repetição de cálculos com tamanho de passo ou método diferentes e a comparação dos resultados. Pode-se empregar análise de sensibilidade para ver como a solução varia quando os parâmetros do modelo ou os valores de entrada variam. Pode-se tentar algoritmos numéricos diferentes que têm fundamentos teóricos distintos, são baseados em estratégias computacionais diferentes ou têm propriedades de convergência e características de estabilidade distintas.

Quando os resultados de cálculos numéricos são extremamente críticos e podem provocar a perda de vidas humanas ou ter ramificações econômicas graves, é apropriado tomar cuidados especiais. Isso pode envolver o uso de dois ou mais grupos independentes para resolver o mesmo problema de modo que os seus resultados possam ser comparados.

O papel dos erros será um tópico de preocupação e análise em todas as seções deste livro. Essas investigações serão feitas em seções específicas.

4.4 ENGANOS, ERROS DE FORMULAÇÃO E INCERTEZA NOS DADOS

Embora as fontes de erros a seguir não estejam diretamente ligadas com a maioria dos métodos deste livro, elas, algumas vezes, têm grande impacto no sucesso do esforço de modelagem. Assim, deve-se tê-las em mente ao aplicar técnicas numéricas no contexto de problemas do mundo real.

4.4.1 Enganos

Todos temos grande familiaridade com os erros grosseiros ou enganos. Nos primeiros anos dos computadores, resultados numéricos errados podiam às vezes ser atribuídos ao mau funcionamento do próprio computador. Hoje, essa fonte de erro é altamente improvável e a maioria dos enganos deve ser atribuída à imperfeição humana.

Os enganos podem ocorrer em qualquer estágio do processo de modelagem matemática e podem contribuir para todas as outras componentes do erro. Eles podem ser evitados apenas por conhecimento profundo dos princípios fundamentais e com o cuidado com o qual se aborda e projeta a solução para um problema.

Os enganos são, em geral, desconsiderados na discussão dos métodos numéricos, o que, sem dúvida, decorre do fato de que, por mais que se tente, os erros são em certa medida inevitáveis. Entretanto, existe um certo número de maneiras pelas quais sua ocor-

rência pode ser minimizada. Em particular, os bons hábitos de programação que foram esboçados no Capítulo 2 são extremamente úteis para diminuir os enganos na programação. Além disso, existem formas simples de verificar se um método numérico particular está funcionando apropriadamente. Por todo o livro serão discutidas maneiras de verificar os resultados dos cálculos numéricos.

4.4.2 Erros de Formulação

Os erros de formulação ou *do modelo* se relacionam com vieses que podem ser atribuídos a modelos matemáticos incompletos. Um exemplo de um erro de formulação desprezível é o fato de a segunda lei de Newton não levar em conta efeitos relativísticos. Isso não afeta a adequação da solução do Exemplo 1.1 porque esses erros são mínimos nas escalas de tempo e espaço associadas com o problema do pára-quedista em queda livre.

Contudo, suponha que a resistência do ar não seja linearmente proporcional à velocidade de queda, como na Equação (1.7), mas uma função do quadrado da velocidade. Se este fosse o caso, as soluções analítica e numérica obtidas no primeiro capítulo estariam erradas por causa do erro de formulação. Considerações adicionais dos erros de formulação estão incluídas em algumas das aplicações em engenharia no restante do livro. É necessário estar ciente desses problemas e perceber que, ao se trabalhar com um modelo precariamente concebido, nenhum método numérico fornecerá resultados adequados.

4.4.3 Incerteza nos Dados

Erros, algumas vezes, entram em uma análise por causa de incertezas nos dados físicos nos quais um modelo é baseado. Por exemplo, suponha que se queira testar o modelo do pára-quedista em queda livre fazendo com que um indivíduo pule repetidas vezes e então medindo sua velocidade depois de um intervalo de tempo especificado. Incertezas estariam sem dúvida associadas a tais medidas, já que o pára-quedista cairia mais rapidamente durante alguns saltos do que em outros. Esses erros podem exibir tanto falta de acurácia quanto imprecisão. Se os instrumentos consistentemente subestimarem ou superestimarem a velocidade, o profissional está lidando com um aparelho não-acurado ou com viés. Por outro lado, se as medidas forem aleatoriamente altas e baixas, trata-se de uma questão de precisão.

Os erros nas medidas podem ser quantificados resumindo os dados com uma ou mais estatísticas bem escolhidas que transfiram tanta informação quanto possível com relação às características específicas dos dados. Essas estatísticas descritivas são mais freqüentemente escolhidas para representar (1) a posição do centro da distribuição de dados e (2) o grau de dispersão dos dados. Como tal, elas fornecem uma medida do viés e da imprecisão, respectivamente. O tópico de caracterizar as incertezas nos dados será retomado na Parte Cinco.

Embora seja necessário estar ciente dos enganos, erros de formulação e incerteza nos dados, os métodos numéricos usados para construir modelos podem ser estudados, na maior parte, independentemente de tais erros. Portanto, na maior parte deste livro, a suposição é de que não foram cometidos erros grosseiros, o modelo é seguro, e as medidas estão livres de erro. Nessas condições, é possível estudar os erros numéricos sem fatores complicadores.

PROBLEMAS

4.1 A seguinte série infinita pode ser usada para aproximar e^x:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$$

(a) Demonstre que essa expansão em série de Maclaurin é um caso especial da expansão em série de Taylor [Equação (4.7)] com $x_i = 0$ e $h = x$.

(b) Use a série de Taylor para fazer uma estimativa de $f(x) = e^{-x}$ em $x_{i+1} = 1$ para $x_i = 0{,}2$. Use as versões de ordem zero, um, dois e três e calcule $|\varepsilon_t|$ para cada caso.

4.2 A expansão em série de Maclaurin para cos x é

$$\cos x = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots$$

Começando com a versão mais simples, cos $x = 1$, some termos um a um para fazer uma estimativa de $\cos(\pi/3)$. Depois que cada novo termo for somado, calcule os erros relativos percentuais verdadeiro e aproximado. Utilize sua calculadora de bolso para determinar o valor verdadeiro. Some termos até que o valor absoluto da estimativa aproximada de erro fique abaixo de um critério de erro de até dois algarismos significativos.

4.3 Faça os mesmos cálculos do Problema 4.2, mas use a expansão em série de Maclaurin para o sen x para fazer uma estimativa de sen(π/3),

$$\operatorname{sen} x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

4.4 Use a expansão em série de Taylor de ordem zero até três para prever $f(3)$ para

$$f(x) = 25x^3 - 6x^2 + 7x - 88$$

utilizando o ponto-base $x = 1$. Calcule o erro relativo porcentual ε_t para cada aproximação.

4.5 Use a expansão em série de Taylor de ordem zero até quatro para prever $f(2{,}5)$ para $f(x) = \ln x$ utilizando o ponto-base $x = 1$. Calcule o erro relativo porcentual ε_t para cada aproximação. Discuta o significado dos resultados.

4.6 Use aproximações por diferenças progressiva e regressiva de $O(h)$ e uma aproximação por diferença centrada de $O(h^2)$ para fazer uma estimativa da primeira derivada da função examinada no Problema 4.4. Calcule a derivada em $x = 2$ usando um tamanho de passo $h = 0{,}2$. Compare seus resultados com o valor verdadeiro da derivada. Interprete seus resultados com base no resto da expansão em série de Taylor.

4.7 Use uma aproximação por diferença centrada de $O(h^2)$ para fazer uma estimativa da segunda derivada da função examinada no Problema 4.4. Faça o cálculo em $x = 2$ utilizando tamanhos de passos de $h = 0{,}25$ e $0{,}125$. Compare suas estimativas com o valor verdadeiro da segunda derivada. Interprete seus resultados com base no termo de resto da expansão em série de Taylor.

4.8 Lembre-se de que a velocidade do pára-quedista em queda livre pode ser calculada por [Equação (1.10)]

$$v(t) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right)$$

Utilize uma análise de erro de primeira ordem para fazer uma estimativa de v em $t = 6$, se $g = 9{,}8$ e $m = 50$ mas $c = 12{,}5 \pm 1{,}5$.

4.9 Repita o Problema 4.8 com $g = 9{,}8$, $t = 6$, $c = 12{,}5 \pm 1{,}5$, e $m = 50 \pm 2$.

4.10 A lei de Stefan-Boltzmann pode ser utilizada para se fazer uma estimativa da taxa de radiação de energia H de uma superfície, como em

$$H = Ae\sigma T^4$$

onde H está em watts, A é a área da superfície (m^2), e é a emissividade que caracteriza as propriedades de emissão da superfície (adimensional), σ é uma constante universal chamada constante de Stefan-Boltzmann ($= 5{,}67 \times 10^{-8}$ W m^{-2} K^{-4}) e T é temperatura absoluta (K). Determine o erro de H para uma placa de aço com $A = 0{,}15$ m^2, $e = 0{,}90$ e $T = 650 \pm 20$. Compare seus resultados com o erro exato. Repita os cálculos com $T = 650 \pm 40$. Interprete seus resultados.

4.11 Repita o Problema 4.10 agora para uma esfera de cobre com raio igual a $= 0{,}15 \pm 0{,}01$ m, $e = 0{,}90 \pm 0{,}05$ e $T = 550 \pm 20$.

4.12 Calcule e interprete o número de condicionamento para

(a) $f(x) = \sqrt{|x-1|} + 1$ para $x = 1{,}00001$

(b) $f(x) = e^{-x}$ para $x = 10$

(c) $f(x) = \sqrt{x^2+1} - x$ para $x = 300$

(d) $f(x) = \dfrac{e^{-x} - 1}{x}$ para $x = 0{,}001$

(e) $f(x) = \dfrac{\operatorname{sen} x}{1 + \cos x}$ para $x = 1{,}0001\pi$

4.13 Use as idéias da Seção 4.2 para deduzir as relações da Tabela 4.3.

4.14 Demonstre que a Equação (4.4) é exata para todos os valores de x se $f(x) = ax^2 + bx + c$.

4.15 A fórmula de Manning para um canal retangular pode ser escrita como

$$Q = \frac{1}{n}\frac{(BH)^{5/3}}{(B+2H)^{2/3}}\sqrt{S}$$

onde Q é o escoamento (m^3/s), n é o coeficiente de rugosidade, B é a largura (m), H é a profundidade (m) e S é o declive. Você está aplicando esta fórmula para um córrego onde você sabe que a largura é 20 m e que a profundidade e 0,3 m. Infelizmente, você sabe a rugosidade e o declive com uma precisão apenas de $\pm$ 10%. Ou seja, você sabe que a rugosidade é de cerca de 0,03 com uma variação de 0,027 a 0,033 e o declive é 0,0003 com uma variação de 0,00027 a 0,00033. Utilize uma análise de erro de primeira ordem para determinar a sensibilidade da previsão do escoamento para cada um desses fatores. Qual deles você deveria tentar medir com mais precisão?

4.16 Se $|x| < 1$, é sabido que

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

Repita o problema 4.2 para esta série para $x = 0{,}1$.

4.17 Um míssil deixa o solo com velocidade inicial v_0 formando um ângulo ϕ_0 com a vertical, como mostrado na Figura P4.17. A altitude máxima desejada é αR onde R é o raio da Terra. As leis da mecânica podem ser usadas para mostrar que

$$\operatorname{sen}\phi_0 = (1+\alpha)\sqrt{1 - \frac{\alpha}{1+\alpha}\left(\frac{v_e}{v_0}\right)^2}$$

onde v_e é a velocidade de escape do míssil. Quer-se disparar o míssil e atingir a velocidade máxima do projeto com uma acurácia de $\pm 2\%$. Determine o intervalo de valores para ϕ_0 se $v_e/v_0 = 2$ e $\alpha = 0{,}25$.

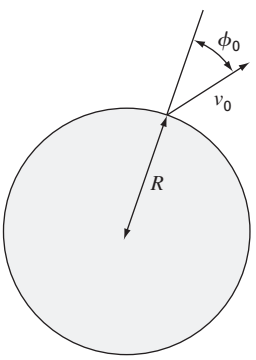


Figura P4.17

4.18 Para calcular as coordenadas espaciais de um planeta, é necessário calcular a função

$$f(x) = x - 1 - 0{,}5 \operatorname{sen} x$$

Tome o ponto-base $a = x_i = \pi/2$ no intervalo $[0, \pi]$. Determine a expansão em série de Taylor de ordem mais alta que resulta em um

erro máximo de 0,015 no intervalo especificado. O erro é igual ao valor absoluto da diferença entre a função dada e a expansão em série de Taylor específica. (*Sugestão*: resolva graficamente.)

4.19 Considere a função $f(x) = x^3 - 2x + 4$ no intervalo $[-2, 2]$ com $h = 0{,}25$. Use aproximações por diferenças finitas progressivas, regressivas e centradas para a primeira e a segunda derivadas de modo a ilustrar graficamente qual aproximação é mais acurada. Faça o gráfico de todas as três aproximações por diferenças finitas da primeira derivada junto com a derivada teórica, e faça o mesmo também para a segunda derivada.



EPÍLOGO: PARTE UM

PT1.4 PRÓS E CONTRAS

Os métodos numéricos são científicos no sentido que representam técnicas sistemáticas para resolver problemas matemáticos. Entretanto, existe um certo grau de arte, julgamento subjetivo e compromisso associado a seu uso efetivo na engenharia prática. Em cada problema, você pode ser confrontado por diversos métodos numéricos alternativos e muitos tipos diferentes de computadores. Portanto, a elegância e a eficiência das diferentes abordagens dos problemas são altamente individualistas e relacionadas com sua habilidade de escolher sensatamente entre as opções. Infelizmente, como em qualquer processo intuitivo, os fatores que influenciam essa escolha são difíceis de transmitir. Apenas pela experiência essas técnicas podem ser totalmente compreendidas e aperfeiçoadas. Entretanto, como elas desempenham um papel tão importante na implementação efetiva dos métodos, foi incluída esta seção como uma introdução aos prós e contras que você deve considerar quando for escolher um método numérico e as ferramentas para implementar esse método. Espera-se que a discussão a seguir influencie sua orientação quando abordar o material subseqüente. Além disso, espera-se que você consulte novamente este material quando for confrontado com escolhas e prós e contras no restante do livro.

1. *Tipo de Problema Matemático*. Como delineado anteriormente na Figura PT1.2, diversos tipos de problemas matemáticos são discutidos neste livro:
   (a) Raízes de equações.
   (b) Sistemas de equações algébricas lineares simultâneas.
   (c) Otimização.
   (d) Ajuste de curvas.
   (e) Integração numérica.
   (f) Equações diferenciais ordinárias.
   (g) Equações diferenciais parciais.

Os aspectos práticos dos métodos numéricos provavelmente lhe serão introduzidos ao enfrentar um problema em uma das áreas citadas. Os métodos numéricos serão necessários porque o problema não poderá ser resolvido eficientemente usando técnicas analíticas. Você deve estar ciente de que suas atividades profissionais irão eventualmente envolver problemas em todas as áreas acima. Assim, o estudo dos métodos numéricos e a escolha de equipamento de computação automática deveriam, no mínimo, considerar esses tipos básicos de problemas. Problemas mais avançados podem exigir recursos para manipular áreas tais como aproximação funcional, equações integrais etc. Essas áreas tipicamente exigem maior poder computacional e métodos avançados, não cobertos neste texto. Outras referências tais como Carnahan, Luther e Wilkes (1969); Hamming (1973); Ralstone Rabinowitz (1978) e Burden e Faires (1993) deverão ser consultados para problemas além do escopo deste livro. Além disso, no final de cada parte deste texto, vamos incluir um pequeno resumo e referências para métodos avançados para lhe fornecer os caminhos para prosseguir em outros estudos de métodos numéricos.

2. *Tipo, Disponibilidade, Precisão, Custo e Velocidade do Computador*. Você pode ter a opção de trabalhar com diversas ferramentas computacionais. Estas variam desde as calculadoras de bolso até grandes computadores centrais (*mainframe*). Obviamente, qualquer uma dessas ferramentas pode ser usada para implementar qualquer método numérico (incluindo simples papel e lápis). Em geral, não é uma questão de recursos fundamentais, mas sim de custo, conveniência, velocidade, confiança, repetitividade e precisão. Embora cada uma das ferramentas vá continuar a ter utilidade, os avanços rápidos recentes no desempenho dos computadores pessoais já tiveram um grande impacto na profissão de engenheiro. Esperamos que essa revolução se espalhe conforme os avanços tecnológicos continuem, porque os computadores pessoais oferecem um excelente compromisso na conveniência, custo, precisão, velocidade e capacidade de armazenamento. Além disso, eles podem ser prontamente aplicados à maioria dos problemas de engenharia práticos.
3. *Custo do Desenvolvimento de Programas versus Custos de Softwares versus Custo do Tempo de Execução*. Uma vez que o tipo de problemas matemáticos a serem resolvidos tenha sido identificado e que o sistema computacional tenha sido escolhido, será adequado considerar os custos de software e tempo de execução. O desenvolvimento de programas pode representar um esforço substancial em muitos projetos de engenharia e pode, assim, ter um custo significativo. Com relação a isso, é particularmente importante que você conheça muito bem os aspectos teóricos e práticos dos métodos numéricos relevantes. Além disso, você deveria estar familiarizado com os softwares desenvolvidos profissionalmente. Softwares de baixo custo estão amplamente disponíveis para a implementação de métodos numéricos que podem ser facilmente adaptados a uma grande variedade de problemas.
4. *Características do Método Numérico*. Quando os custos do hardware e software computacionais são altos, ou quando a disponibilidade do computador for limitada (por exemplo, em alguns sistemas de uso compartilhado), é vantajoso escolher cuidadosamente o método numérico mais adequado à situação. Por outro lado, se o problema ainda estiver no estágio de pesquisa e o acesso e o custo de computação não forem problema, pode ser mais apropriado escolher um método numérico que funcione sempre, mas que pode não ser o mais eficiente computacionalmente. Os métodos numéricos disponíveis para resolver qualquer tipo particular de problema envolvem os tipos de prós e contras que acabamos de discutir e ainda outros:
   (a) *Número de Aproximações ou Pontos Iniciais*. Alguns dos métodos numéricos para encontrar raízes de equações ou resolver equações diferenciais exigem que o usuário especifique aproximações ou pontos iniciais. Os métodos simples em geral exigem um valor, enquanto os métodos complicados podem exigir mais de um valor. A vantagem dos métodos complicados que são computacionalmente eficientes pode ser contrabalançada pela necessidade de pontos iniciais múltiplos. Você deve usar sua experiência e julgamento para avaliar os prós e contras para cada problema particular.
   (b) *Taxa de Convergência*. Certos métodos numéricos convergem mais rapidamente do que outros. Entretanto, essa convergência mais rápida pode exigir aproximações iniciais mais refinados e programação mais complexa do que um método com convergência mais lenta. Você deve usar seu julgamento na escolha de um método. Mais rápido nem sempre é melhor.
   (c) *Estabilidade*. Alguns métodos numéricos para encontrar raízes de equações ou soluções para sistemas de equações lineares podem divergir em vez de convergir para a resposta correta em certos problemas. Por que você toleraria essa possibilidade quando confrontado com problemas de projeto ou planejamento? A resposta é que esses métodos podem ser altamente eficientes quando funcionam. Portanto, prós e contras emergem novamente. Você precisa decidir se as necessidades do seu problema justificam o esforço necessário para aplicar um método que pode não convergir sempre.
   (d) *Acurácia e Precisão*. Alguns métodos numéricos são simplesmente mais acurados e precisos que outros. Bons exemplos são as diversas equações disponíveis para integração numérica. Em geral, o desempenho dos métodos com baixa

acurácia pode ser melhorado diminuindo o tamanho do passo ou aumentando o número de aplicações em um intervalo dado. É melhor usar um método com baixa acurácia com tamanho de passo pequeno ou um método com alta precisão e com tamanho de passo grande? Essa questão deve ser analisada caso a caso levando em consideração fatores adicionais tais como custo e facilidade de programação. Além disso, você também deve se preocupar com erros de arredondamento quando estiver usando aplicações múltiplas de métodos de baixa precisão e quando o número de operações se torna grande. Aqui, o número de algarismos significativos manipulados pelo computador pode ser o fator decisivo.

(e) *Abrangância das Aplicações*. Alguns métodos numéricos podem ser aplicados apenas a uma classe limitada de problemas ou a problemas que satisfaçam certas restrições matemáticas. Outros métodos não são afetados por tais limitações. Você deve avaliar se vale a pena seu esforço para desenvolver programas que usem técnicas que são apropriadas apenas para um número limitado de problemas. O fato de tais técnicas serem largamente usadas sugere que elas têm vantagens que em geral serão mais importantes que as desvantagens. Obviamente estão ocorrendo prós e contras.

(f) *Requisitos Especiais*. Algumas técnicas numéricas tentam aumentar a acurácia e a taxa de convergência usando informações adicionais ou especiais. Um exemplo seria usar valores teóricos ou estimados dos erros para melhorar a precisão. Entretanto, essas melhorias em geral não são alcançadas sem alguma inconveniência em termos de custos computacionais adicionais ou aumento na complexidade do programa.

(g) *Esforço de Programação Necessário*. Esforços para melhorar as taxas de convergência, estabilidade e acurácia podem ser criativos e engenhosos. Quando os aprimoramentos podem ser feitos sem aumentar a complexidade da programação, eles podem ser considerados elegantes e provavelmente serão imediatamente usados na profissão de engenheiro. Entretanto, se exigirem programas mais complicados, você estará se deparando novamente com uma situação de prós e contras que pode ou não favorecer o novo método.

É claro que a discussão acima com relação à escolha de métodos numéricos se reduz à discussão de custos e acurácia. Os custos são aqueles envolvidos com tempo de computação e desenvolvimento de programa. Acurácia adequada é uma questão de julgamento profissional e ética.

5. *Comportamento Matemático da Função, Equação ou Dados*. Na escolha de um método numérico particular, de um tipo de computador e de software, você deve considerar a complexidade de sua função, equação ou dados. Equações simples e dados lisos podem ser manipulados adequadamente por algoritmos numéricos simples e computadores baratos. O oposto é verdade para equações complicadas e dados exibindo descontinuidades.

6. *Facilidade de Aplicação (Amigável ao Usuário?).* Alguns métodos numéricos são fáceis de aplicar, outros são difíceis. Isso pode ser uma consideração na escolha de um método sobre outro. A mesma idéia se aplica em decisões relativas a custos de desenvolvimento de programa *versus* software desenvolvido profissionalmente. Pode exigir um esforço considerável converter um programa difícil em um que seja amigável ao usuário. Foram introduzidas maneiras de fazer isso no Capítulo 2 e estas serão elaboradas no transcorrer do livro.

7. *Manutenção*. Os programas para resolver problemas de engenharia necessitam de manutenção porque durante a aplicação invariavelmente ocorrem dificuldades. A manutenção pode exigir mudanças no código do programa ou expansão da documentação. Programas e algoritmos numéricos simples são mais fáceis de manter.

Os capítulos a seguir envolvem o desenvolvimento de diversos tipos de métodos numéricos para vários tipos de problemas matemáticos. Muitos métodos alternativos serão dados em cada capítulo. Esses diversos métodos (em vez de um único método escolhido pelos autores) serão apresentados porque não existe um único "melhor" método. Não

existe um melhor método porque existem muitos prós e contras que precisam ser considerados na aplicação do método a problemas práticos. Uma tabela que destaca os prós e contras envolvidos em cada método será encontrada no final de cada parte deste livro. Essa tabela deveria ajudá-lo a escolher o procedimento numérico para o contexto de seu problema particular.

PT1.5 RELAÇÕES E FÓRMULAS IMPORTANTES

A Tabela PT1.2 resume informações importantes que foram apresentadas na Parte Um. A tabela pode ser consultada para um acesso rápido a relações e fórmulas importantes. O epílogo de cada parte deste livro conterá tal resumo.

PT1.6 MÉTODOS AVANÇADOS E REFERÊNCIAS ADICIONAIS

O epílogo de cada parte deste livro também incluirá uma seção planejada para facilitar e encorajar estudos adicionais de métodos numéricos. Esta seção citará outros livros no assunto bem como material relacionado a métodos avançados.[1]

Para estender os fundamentos fornecidos na Parte Um, numerosos manuais sobre a programação de computadores estão disponíveis. Seria difícil fazer referência a todos os livros excelentes e manuais pertinentes a linguagens e computadores específicos. Além disso, você provavelmente já tem material de sua exposição anterior à programação. Entretanto, se esta for sua primeira experiência com computadores, Chapra e Canale (1994) fornecem uma introdução geral ao BASIC e Fortran. Seu professor e seus colegas também devem ser capazes de aconselhá-lo com relação a bons livros de referência para máquinas e linguagens disponíveis em sua escola.

Quanto à análise de erro, qualquer bom livro introdutório de cálculo incluirá material suplementar relativo a assuntos tais como expansão em série de Taylor. Os textos de Swokowski (1979), Thomas e Finney (1979) e Simmons (1985) fornecem discussões muito legíveis sobre esses assuntos. Além disso, Taylor (1982) apresenta uma boa introdução à análise de erros.

Finalmente, embora esperemos que nosso livro lhe sirva bem, é sempre bom consultar outras fontes quando se tenta dominar um novo assunto. Burden e Faires (1993); Ralston e Rabinowitz (1978); Hoffman (1992) e Carnahan, Luther e Wilkes (1969) fornecem discussões abrangentes da maioria dos métodos numéricos, incluindo alguns métodos avançados que estão além do escopo deste livro. Outros livros agradáveis no assunto são Gerald e Wheatley (1989); Rice (1983) e Cheney e Kincaid (1985). Além disso, Press et al. (1992) inclui códigos de computador para implementar diversos métodos.

[1]Aqui, os livros são citados apenas pelos autores; uma bibliografia completa será fornecida no final deste texto.

TABELA PT1.2 Resumo das informações importantes apresentadas na Parte Um.

Definições de Erro

Erro verdadeiro: $E_t = \text{erro verdadeiro} - \text{aproximação}$

Erro relativo porcentual verdadeiro: $\varepsilon_t = \dfrac{\text{erro verdadeiro} - \text{aproximação}}{\text{valor verdadeiro}}\ 100\%$

Erro relativo porcentual aproximado: $\varepsilon_a = \dfrac{\text{aproximação atual} - \text{aproximação anterior}}{\text{aproximação atual}}\ 100\%$

Critério de parada: Pare os cálculos quando $\varepsilon_a < \varepsilon_s$ onde ε_s é o erro relativo porcentual pedido

Série de Taylor

Expansão em série de Taylor:

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2!}h^2 + \frac{f'''(x_i)}{3!}h^3 + \cdots + \frac{f^{(n)}(x_i)}{n!}h^n + R_n$$

onde

Resto:

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1}$$

ou

$$R_n = O(h^{n+1})$$

Diferenciação Numérica

Primeira diferença dividida finita progressiva:

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} + O(h)$$

(Outras diferenças divididas estão resumidas nos Capítulos 4 e 23)

Propagação de erros

Para n variáveis independentes $x_1, x_2, \ldots, x_n$ tendo erros $\Delta\tilde{x}_1, \Delta\tilde{x}_2, \ldots, \Delta\tilde{x}_n$, o erro da função f pode ser estimado por

$$\Delta f = \left|\frac{\partial f}{\partial x_1}\right| \Delta\tilde{x}_1 + \left|\frac{\partial f}{\partial x_2}\right| \Delta\tilde{x}_2 + \cdots + \left|\frac{\partial f}{\partial x_n}\right| \Delta\tilde{x}_n$$

PARTE DOIS



RAÍZES DE EQUAÇÕES

PT2.1 MOTIVAÇÃO

Anos atrás, você aprendeu a usar a fórmula quadrática

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{PT2.1}$$

para resolver

$$f(x) = ax^2 + bx + c = 0 \tag{PT2.2}$$

Os valores calculados com a Equação (PT 2.1) são chamados de "raízes" da Equação (PT2.2). Eles representam os valores de x que fazem a Equação (PT2.2) igual a zero. Logo, pode-se definir uma raiz de uma equação como um valor de x que torna $f(x) = 0$. Por essa razão, as raízes são às vezes chamadas de *zeros* da equação.

Embora a fórmula quadrática seja muito cômoda para resolver a Equação (PT2.2), existem muitas outras funções para as quais as raízes não podem ser determinadas tão facilmente. Para esses casos, os métodos numéricos descritos nos Capítulos 5, 6 e 7 fornecerão maneiras de obter a resposta.

PT2.1.1 Métodos para Determinar as Raízes sem o Computador

Antes do aparecimento dos computadores digitais, existiam diversas maneiras de determinar as raízes de equações algébricas e transcendentais. Em alguns casos, as raízes podiam ser obtidas por métodos diretos, como foi feito com a Equação (PT2.1). Embora existissem equações como essa que podiam ser resolvidas diretamente, existiam muitas outras que não podiam. Por exemplo, mesmo uma função aparentemente simples como $f(x) = e^{-x} - x$ não pode ser resolvida analiticamente. Em tais casos, a única alternativa é uma técnica de solução aproximada.

Um método para obter uma solução aproximada é traçar o gráfico da função e determinar onde ele cruza o eixo dos x. Esse ponto, que representa um valor de x para o qual $f(x) = 0$, é a raiz. Técnicas gráficas são discutidas no início dos Capítulos 5 e 6.

Embora os métodos gráficos sejam úteis para obter estimativas grosseiras, eles são limitados por causa de sua falta de precisão. Uma abordagem alternativa é usar tentativa e erro. Essa "técnica" consiste em chutar um valor de x e calcular para ver se $f(x)$ é zero. Se não (como é quase sempre o caso), é feito um outro chute, e $f(x)$ é novamente calculado para determinar se o novo valor fornece uma estimativa melhor para a raiz. O processo é repetido até que seja feito um chute que resulte em um $f(x)$ que esteja próximo de zero.

Tais métodos tão casuais são obviamente ineficientes e inadequados para as necessidades da prática da engenharia. As técnicas descritas na Parte Dois representam alternativas que também são aproximadas, mas que empregam estratégias sistemáticas para mirar na raiz verdadeira. Como elaborado nas próximas páginas, a combinação desses métodos sistemáticos e computadores torna a solução da maioria dos problemas aplicados de raízes de equações uma tarefa simples e eficiente.

PT2.1.2 Raízes de Equações e a Prática da Engenharia

Embora apareçam em outros contextos de problemas, as raízes das equações em geral ocorrem na área de projetos de engenharia. A Tabela PT2.1 lista diversos princípios fundamentais que são rotineiramente usados no trabalho de projeto. Como introduzido no Capítulo 1, as equações ou modelos matemáticos deduzidos a partir desses princípios são usados para prever variáveis dependentes como função das variáveis independentes,

termos forçantes e parâmetros. Observe que, em cada caso, as variáveis dependentes refletem o estado ou o desempenho do sistema, enquanto os parâmetros representam suas propriedades ou composição.

Um exemplo de tal modelo é a equação, deduzida da segunda lei de Newton, usada no Capítulo 1 para a velocidade do pára-quedista:

$$v = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{PT2.3}$$

onde v é a variável dependente, o tempo t é a variável independente, a constante gravitacional g é o termo forçante e o coeficiente de arrasto c e a massa m são parâmetros. Se os parâmetros forem conhecidos, a Equação (PT2.3) pode ser usada para prever a velocidade do pára-quedista como uma função do tempo. Tais cálculos podem ser feitos diretamente porque v é expressa *explicitamente* como uma função do tempo. Isto é, ela está isolada em um lado do sinal de igual.

Entretanto, suponha que tenhamos que determinar o coeficiente de arrasto para que um pára-quedista de uma dada massa atinja uma certa velocidade em um determinado intervalo de tempo. Embora a Equação (PT2.3) forneça uma representação matemática da inter-relação entre as variáveis e os parâmetros do modelo, ela não pode ser resolvida explicitamente para determinar o coeficiente de arrasto. Tente fazer isso. Não há maneira de reorganizar a equação de forma que c esteja isolado em um lado do sinal de igual. Em tais casos, dizemos que c está *implícito*.

Isso representa um dilema real, porque muitos problemas de projeto de engenharia envolvem especificar as propriedades ou a composição do sistema (como representado por seus parâmetros) para garantir que funcionem da maneira desejada (como represen-

TABELA PT2.1 Princípios fundamentais usados em problemas e projetos de engenharia.

Princípio Fundamental	Variáveis Dependentes	Variáveis Independentes	Parâmetros
Balanço de calor	Temperatura	Tempo e posição	Propriedades térmicas do material e geometria do sistema
Balanço de massa	Concentração ou quantidade de massa	Tempo e posição	Comportamento químico do material, coeficientes de transferência de massa e geometria do sistema
Balanço de energia	Intensidade e direção das forças	Tempo e posição	Resistência do material, propriedades estruturais e geometria do sistema
Balanço de energia	Mudanças nos estados de energia cinética e potencial do sistema	Tempo e posição	Propriedades térmicas, massa do material e geometria do sistema
Leis de movimento de Newton	Aceleração, velocidade ou posição	Tempo e posição	Massa do material, geometria do sistema e parâmetros dissipativos tais como atrito ou arrasto
Leis de Kirchhoff	Correntes e voltagens em circuitos elétricos	Tempo	Propriedades elétricas de sistema tais como resistência, capacitância e indutância

tado por suas variáveis). Assim, esses problemas freqüentemente exigem a determinação de parâmetros implícitos.

A solução do dilema é fornecida pelos métodos numéricos para raízes de equações. Para resolver o problema usando métodos numéricos, é conveniente reescrever a Equação (PT2.3). Isso é feito subtraindo-se a variável dependente v de ambos os lados da equação para obter

$$f(c) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) - v \tag{PT2.4}$$

O valor de c que torna $f(c) = 0$ é, portanto, a raiz da equação. Esse valor também representa o coeficiente de arrasto que resolve o problema de engenharia.

A Parte Dois deste livro trata de uma variedade de métodos numéricos e gráficos para a determinação de raízes de relações como a Equação (PT2.4). Essas técnicas podem ser aplicadas a problemas de projeto de engenharia que são baseados nos princípios fundamentais delineados na Tabela PT2.1 e também a muitos outros problemas confrontados rotineiramente na prática da engenharia.

PT2.2 FUNDAMENTOS MATEMÁTICOS

Para a maioria das áreas de assuntos neste livro, em geral existem alguns pré-requisitos matemáticos necessários para ter sucesso em dominar o tópico. Por exemplo, os conceitos de estimativa do erro e de expansão em série de Taylor discutidos nos Capítulos 3 e 4 têm relevância direta em nossa discussão de raízes de equações. Adicionalmente, antes deste ponto já mencionamos os termos equações "algébricas" e "transcendentais". Poderia ser útil definir formalmente esses termos e discutir como eles se relacionam ao escopo desta parte do livro.

Por definição, uma função dada por $y = f(x)$ é algébrica se ela puder ser expressa na forma

$$f_n y^n + f_{n-1} y^{n-1} + \cdots + f_1 y + f_0 = 0 \tag{PT2.5}$$

onde f_i é um polinômio de grau n em x. Os *polinômios* são uma classe simples de funções algébricas que são representadas em geral por

$$f_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \tag{PT2.6}$$

onde n é o *grau* do polinômio e os a's são constantes. Alguns exemplos específicos são

$$f_2(x) = 1 - 2{,}37x + 7{,}5x^2 \tag{PT2.7}$$

e

$$f_6(x) = 5x^2 - x^3 + 7x^6 \tag{PT2.8}$$

Uma função *transcendental* é uma que não é algébrica. Incluem-se as funções trigonométricas, exponenciais, logarítmicas e outras funções menos familiares. Exemplos são

$$f(x) = \ln x^2 - 1 \tag{PT2.9}$$

e

$$f(x) = e^{-0{,}2x}\operatorname{sen}(3x - 0{,}5) \tag{PT2.10}$$

As raízes das equações podem ser reais ou complexas. Embora existam casos nos quais as raízes complexas de funções não polinomiais sejam de interesse, tais situações são menos comuns do que para polinômios. Como conseqüência, os métodos padrão para localizar raízes tipicamente se dividem em duas áreas de problemas um pouco relacionadas, mas primariamente distintas:

1. *A determinação de raízes reais de equações algébricas e transcendentais*. Essas técnicas são usualmente desenvolvidas para determinar o valor de uma única raiz com base no conhecimento prévio de sua localização aproximada.

2. *A determinação de todas as raízes reais e complexas de polinômios*. Esses métodos são desenvolvidos especificamente para polinômios. Determinam sistematicamente todas as raízes do polinômio em vez de determinar uma única raiz real a partir de uma posição aproximada.

Neste livro discutiremos ambos os casos. Os Capítulos 5 e 6 são dedicados à primeira categoria. O Capítulo 7 lida com polinômios.

PT2.3 ORIENTAÇÃO

Alguma orientação será útil antes de prosseguir para os métodos numéricos para a determinação de raízes de equações. O que segue é destinado a lhe dar uma visão geral do material na Parte Dois. Além disso, foram incluídos alguns objetivos para lhe ajudar a focalizar seus esforços quando estudar o material.

PT2.3.1 Escopo e Visão Geral

A Figura PT2.1 é uma representação esquemática da organização da Parte Dois. Examine essa figura cuidadosamente, começando na parte de cima e prosseguindo no sentido horário.

Depois dessa introdução, o Capítulo 5 é dedicado aos *métodos intervalares* para determinar uma raiz. Esses métodos começam com a descoberta de um intervalo que contenha a raiz e depois sistematicamente diminuem a largura do intervalo. Dois métodos específicos são discutidos: *bissecção* e *falsa posição.* Os métodos gráficos são usados para fornecer uma percepção visual das técnicas. As formulações de erros são deduzidas para ajudá-lo a determinar quanto esforço computacional é necessário para fazer uma estimativa da raiz até o nível de precisão pré-especificado.

O **Capítulo 6** cobre os *métodos abertos*. Esses métodos também envolvem iterações sistemáticas de tentativa e erro, mas não exigem aproximações iniciais que delimitem a raiz. Descobriremos que em geral estes métodos são computacionalmente mais eficientes do que os métodos intervalares, mas não funcionam sempre. Os métodos de *iteração de um ponto, Newton-Raphson* e da *secante* serão descritos. Os métodos gráficos serão usados para fornecer uma percepção geométrica dos casos nos quais os métodos abertos não funcionam. As fórmulas são deduzidas para fornecer uma idéia de quão rapidamente os métodos abertos atingem a raiz. Além disso, é explicada uma abordagem para estender o método de Newton-Raphson a *sistemas de equações não lineares.*

O **Capítulo 7** é dedicado a encontrar as *raízes de polinômios*. Depois da seção de fundamentos sobre polinômio, o uso dos métodos convencionais (em particular dos métodos abertos do Capítulo 6) é discutido. A seguir, dois métodos especiais para localizar raízes de polinômios são descritos: os métodos de Müller e Bairstow. O capítulo termina com informações relacionadas a encontrar as raízes com programas de biblioteca e pacotes de software.

O **Capítulo 8** estende os conceitos acima para problemas de engenharia reais. Os casos de engenharia estudados são usados para ilustrar a força e as fraquezas de cada método e para desenvolver a intuição na aplicação das técnicas na prática profissional. As aplicações também destacam os prós e contras (como discutido na Parte Um) associados a vários métodos.

Um epílogo é incluído no final da Parte Dois. Ele contém uma comparação detalhada dos métodos discutidos nos Capítulos 5, 6 e 7. Essa comparação inclui uma descrição dos prós e contras relacionados com o uso adequado de cada técnica. Essa seção também fornece um resumo de fórmulas importantes, juntamente com referências para alguns métodos numéricos que estão além do escopo deste texto.

PT 2.3.2 Metas e Objetivos

Objetivos de Estudo. Depois de completar a Parte Dois, você deveria ter informação suficiente para abordar com sucesso uma grande variedade de problemas de engenharia



FIGURA PT2.1
Esquema de organização do material na Parte Dois: Raízes de Equações.

relacionados com raízes de equações. Em geral, você deve ter dominado as técnicas, aprendido a avaliar sua confiabilidade e ser capaz de escolher o melhor método (ou métodos) para qualquer problema particular. Além dessas metas gerais, os conceitos específicos na Tabela PT2.2 devem estar assimilados para um entendimento completo do material da Parte Dois.

Objetivos Computacionais. O livro lhe oferece software e algoritmos computacionais simples para implementar as técnicas discutidas na Parte Dois. Todos têm utilidade como ferramentas computacionais.

Pseudocódigos para diversos métodos também são fornecidos diretamente no texto. Essa informação vai permitir-lhe expandir sua biblioteca de softwares, incluindo programas que são mais eficientes que o método da bissecção. Por exemplo, você pode querer ter seu próprio software para as técnicas da falsa posição, de Newton-Raphson e da secante, que são em geral mais eficientes que o método da bissecção.

TABELA PT2.2 Objetivos de estudo específicos da Parte Dois.

1. Entender a interpretação gráfica de uma raiz.
2. Saber a interpretação gráfica do método da falsa posição e por que em geral ele é superior ao método da bissecção.
3. Entender a diferença entre os métodos intervalares e os métodos abertos para localizar raízes.
4. Entender os conceitos de convergência e divergência; usar o método gráfico de duas curvas para fornecer uma descrição visual desses conceitos.
5. Saber por que os métodos intervalares sempre convergem, enquanto os métodos abertos podem às vezes divergir.
6. Perceber que a convergência dos métodos abertos é mais provável se a aproximação inicial estiver próximo da raiz verdadeira.
7. Entender os conceitos de convergência linear e quadrática e suas conseqüências na eficiência dos métodos de iteração de ponto fixo e de Newton-Raphson.
8. Saber a diferença fundamental entre os métodos da falsa posição e da secante e como ela se relaciona com a convergência.
9. Entender o problema causado pelas raízes múltiplas e as modificações disponíveis para abrandá-los.
10. Saber como estender a abordagem de Newton-Raphson para uma única equação para resolver sistemas de equações não-lineares.

Finalmente, pacotes como os softwares Excel, MATLAB e bibliotecas de programas têm recursos poderosos para localizar raízes. Você pode usar esta parte do livro para se familiarizar com esses recursos.

CAPÍTULO 5

Métodos Intervalares

Este capítulo sobre raízes de equações trata de métodos que exploram o fato de que uma função tipicamente muda de sinal na vizinhança de uma raiz. Essas técnicas (*bracketing methods*), que isolam a raiz em um intervalo, exigem duas estimativas iniciais para a raiz. Tais estimativas devem "delimitar" a — ou estar uma de cada lado da — raiz. Os métodos particulares descritos aqui usam estratégias diferentes para sistematicamente diminuir a largura do intervalo e, portanto, aproximar-se da resposta correta.

Como uma introdução a essas técnicas, serão discutidos brevemente métodos gráficos para descrever as funções e suas raízes. Além de sua utilidade para fornecer estimativas grosseiras, as técnicas gráficas também são úteis na visualização das propriedades das funções e do comportamento de diversos métodos numéricos.

5.1 MÉTODOS GRÁFICOS

Um método simples para obter uma estimativa da raiz da equação $f(x) = 0$ é fazer um gráfico da função e observar onde ela corta o eixo dos x. Esse ponto, que representa o valor de x para o qual $f(x) = 0$, fornece uma aproximação grosseira da raiz.

EXEMPLO 5.1 A Abordagem Gráfica

Enunciado do Problema. Use a abordagem gráfica para determinar o coeficiente de arrasto c necessário para que um pára-quedista de massa $m = 68{,}1$ kg tenha uma velocidade de 40 m/s depois de cair em queda livre por $t = 10$ s. *Observação*: a aceleração da gravidade é 9,8 m/s^2.

Solução. Esse problema pode ser resolvido determinando-se as raízes da Equação (PT2.4) usando os parâmetros $t = 10$, $g = 9{,}8$, $v = 40$ e $m = 68{,}1$:

$$f(c) = \frac{9{,}8(68{,}1)}{c}\left(1 - e^{-(c/68{,}1)10}\right) - 40$$

ou

$$f(c) = \frac{667{,}38}{c}\left(1 - e^{-0{,}146843c}\right) - 40 \qquad \text{(E5.1.1)}$$

Pode-se substituir vários valores de c no lado direito dessa equação para calcular

c	$f(c)$
4	34,115
8	17,653
12	6,067
16	−2,269
20	−8,401

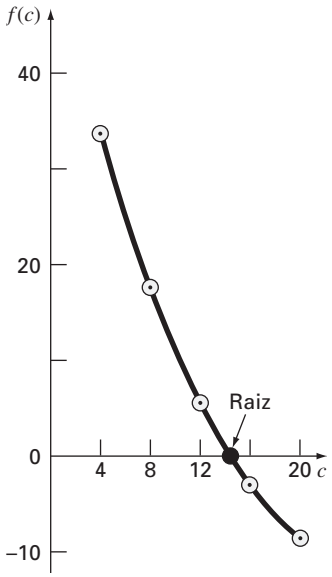


FIGURA 5.1
A abordagem gráfica para determinar as raízes de uma equação.

Esses pontos estão marcados na Figura 5.1. A curva resultante cruza o eixo c entre 12 e 16. Uma inspeção visual do gráfico fornece uma estimativa grosseira da raiz de 14,75. A validade da estimativa gráfica pode ser verificada pela substituição na Equação (E5.1.1), o que fornece

$$f(14{,}75) = \frac{667{,}38}{14{,}75}\left(1 - e^{-0{,}146843(14{,}75)}\right) - 40 = 0{,}059$$

que está próximo de zero. Isso também pode ser verificado pela substituição na Equação (PT2.4) juntamente com os parâmetros desse exemplo, o que dá

$$v = \frac{9{,}8(68{,}1)}{14{,}75}\left(1 - e^{-(14{,}75/68{,}1)10}\right) = 40{,}059$$

que está muito próximo da velocidade de queda pedida de 40 m/s.

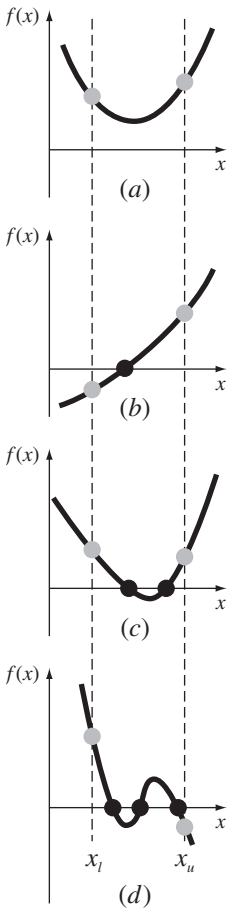


FIGURA 5.2
Ilustração de diversas formas gerais nas quais uma raiz pode ocorrer em um intervalo determinado por uma extremidade inferior x_l e uma extremidade superior x_u. As partes (*a*) e (*c*) indicam que, se $f(x_l)$ e $f(x_u)$ tiverem o mesmo sinal, ou não existirão raízes ou existirá um número par de raízes no intervalo. As partes (*b*) e (*d*) indicam que, se a função tiver sinais diferentes nas extremidades, existirá um número ímpar de raízes no intervalo.

As técnicas gráficas têm valor prático limitado porque não são precisas. Entretanto, os métodos gráficos podem ser usados para obter estimativas grosseiras das raízes. Essas estimativas são usadas como aproximações iniciais para os métodos numéricos discutidos neste e no próximo capítulo.

Além de fornecer estimativas grosseiras para as raízes, as interpretações gráficas são ferramentas importantes para se entender as propriedades das funções e antecipar as armadilhas dos métodos numéricos. Por exemplo, a Figura 5.2 mostra diversas maneiras pelas quais as raízes podem ocorrer (ou não) em um intervalo determinado por uma extremidade inferior x_l e por uma extremidade superior x_u. A Figura 5.2*b* descreve o caso no qual uma única raiz foi delimitada por valores positivos e negativos de $f(x)$. Entretanto, a Figura 5.2*d*, onde $f(x_l)$ e $f(x_u)$ também estão de lados opostos do eixo x, mostra três raízes ocorrendo nesse intervalo. Em geral, se $f(x_l)$ e $f(x_u)$ têm sinais opostos, existe um número ímpar de raízes no intervalo. Como indicado pelas Figuras 5.2*a* e *c*, se $f(x_l)$ e $f(x_u)$ têm ambos o mesmo sinal, ou não existe nenhuma raiz ou existe um número par de raízes entre os valores.

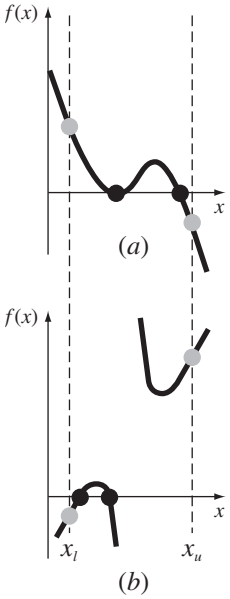


FIGURA 5.3
Ilustrações de algumas exceções aos casos gerais mostrados na Figura 5.2. (*a*) Raízes múltiplas que ocorrem quando a função é tangente ao eixo *x*. Nesse caso, embora as extremidades tenham sinais opostos, há um número par de intersecções com o eixo nesse intervalo. (*b*) Funções descontínuas nas quais as extremidades de sinais opostos cercam um número par de raízes. Estratégias especiais são necessárias para determinar as raízes nesses casos.

Essas generalizações são quase sempre verdadeiras, mas há casos em que elas não funcionam. Por exemplo, as funções que são tangentes ao eixo x (Figura 5.3*a*) ou que são funções descontínuas (Figura 5.3*b*) podem violar esses princípios. Um exemplo de uma função que é tangente ao eixo é a equação cúbica $f(x) = (x - 2)(x - 2)(x - 4)$. Observe que $x = 2$ torna dois termos nesse polinômio iguais a zero. Matematicamente, $x = 2$ é chamado de *raiz múltipla*. No final do Capítulo 6, serão apresentadas técnicas destinadas a localizar raízes múltiplas.

A existência de casos do tipo mostrados na Figura 5.3 torna difícil desenvolver algoritmos computacionais gerais que garantam a localização de todas as raízes em um intervalo. Entretanto, quando usados junto com abordagens gráficas, os métodos descritos nas seções seguintes são extremamente úteis para resolver problemas de raízes de equações confrontados rotineiramente por engenheiros e matemáticos aplicados.

EXEMPLO 5.2 Uso de Gráficos Computacionais para Localizar Raízes

Enunciado do Problema. Os gráficos computacionais podem tornar mais rápidos e melhores seus esforços para localizar as raízes de equações. A função

$$f(x) = \text{sen } 10x - \cos 3x$$

tem diversas raízes no intervalo de $x = 0$ a $x = 5$. Use gráficos computacionais para adquirir uma percepção do comportamento dessa função.

Solução. Os pacotes de software como Excel e MATLAB podem ser usados para gerar gráficos. A Figura 5.4*a* é um gráfico de $f(x)$ de $x = 0$ a $x = 5$. Esse gráfico sugere a presença de diversas raízes, incluindo uma possível raiz dupla perto de $x = 4{,}2$ onde $f(x)$ parece ser tangente ao eixo x. Uma figura mais detalhada do comportamento de $f(x)$ é obtida variando-se o domínio do gráfico de $x = 3$ a $x = 5$, como mostrado na Figura 5.4*b*. Finalmente, na Figura 5.4*c*, a escala vertical é estreitada ainda mais para $f(x) = -0{,}15$ a $f(x) = 0{,}15$ e a escala horizontal é estreitada para $x = 4{,}2$ a $x = 4{,}3$. Esse gráfico mostra claramente que não existe uma raiz dupla nessa região e que na realidade existem duas raízes distintas perto de $x = 4{,}23$ e de $x = 4{,}26$.

Os gráficos computacionais terão grande utilidade nos seus estudos de métodos numéricos. Esse recurso também terá muitas outras aplicações em suas outras disciplinas e também em suas atividades profissionais.

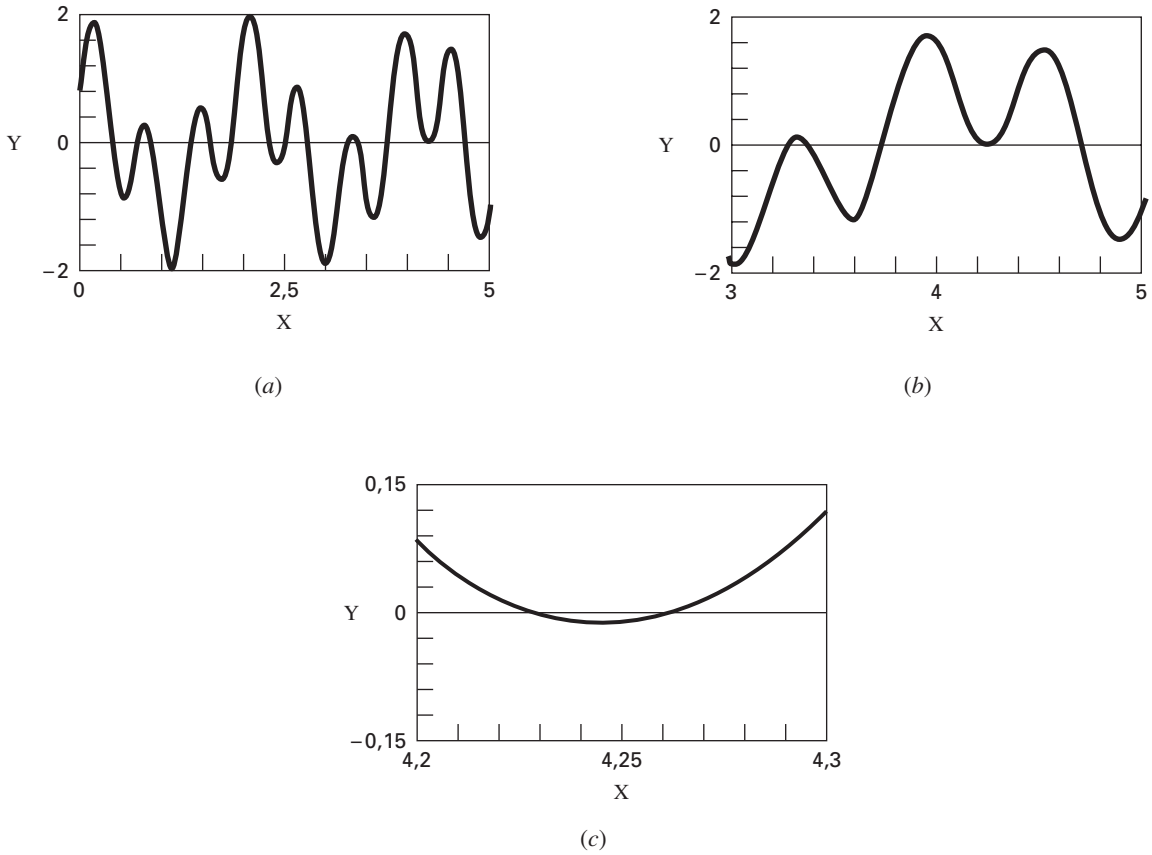


FIGURA 5.4
Uma ampliação progressiva de $f(x) = \text{sen } 10x + \cos 3x$ pelo computador. Esses recursos gráficos interativos permitem ao analista determinar que existem duas raízes distintas entre $x = 4{,}2$ e $x = 4{,}3$.

5.2 O MÉTODO DA BISSECÇÃO

Quando aplicou a técnica gráfica do Exemplo 5.1, você observou (Figura 5.1) que $f(x)$ mudou de sinal em lados opostos da raiz. Em geral, se $f(x)$ for real e contínua no intervalo de x_l a x_u e $f(x_l)$ e $f(x_u)$ tiverem sinais opostos, isto é,

$$f(x_l)\, f(x_u) < 0 \tag{5.1}$$

então existe pelo menos uma raiz real entre x_l e x_u.

Os *métodos de busca incrementais* tiram vantagem dessa observação localizando um intervalo no qual a função muda de sinal. Então, a posição da mudança de sinal (e, conseqüentemente, da raiz) é identificada mais precisamente dividindo-se o intervalo em diversos subintervalos. Procura-se em cada um desses subintervalos para localizar a mudança de sinal. O processo é repetido e a estimativa da raiz é refinada dividindo-se os subintervalos em incrementos menores. Voltaremos ao tópico geral de buscas incrementais na Seção 5.4.

O *método da bissecção*, que é alternativamente chamado de truncamento binário, divisão do intervalo na metade, ou método de Bolzano, é um tipo de método de busca incremental no qual o intervalo é sempre dividido na metade. Se uma função muda de sinal em um intervalo, calcula-se o valor da função em seu ponto médio. A posição da raiz é então determinada como sendo o ponto médio do subintervalo no qual a mudança de sinal ocorre. Esse processo é repetido para obter estimativas refinadas. Um algoritmo simples para os cálculos da bissecção é listado na Figura 5.5 e uma descrição gráfica do método é fornecida na Figura 5.6. O exemplo a seguir exibe todos os cálculos realmente envolvidos no método.

Passo 1: Escolha as aproximações inferior x_l e superior x_u para a raiz de modo que a função mude de sinal no intervalo. Isso pode ser verificado garantindo que $f(x_l)f(x_u) < 0$.

Passo 2: Uma estimativa da raiz é determinada por

$$x_r = \frac{x_l + x_u}{2}$$

Passo 3: Faça os seguintes cálculos para determinar em qual subintervalo a raiz está:

(*a*) se $f(x_l)f(x_r) < 0$, a raiz está no subintervalo inferior. Portanto, faça $x_u = x_r$ e volte ao passo 2.

(*b*) se $f(x_l)f(x_r) > 0$, a raiz está no subintervalo superior. Portanto, faça $x_l = x_r$ e volte ao passo 2.

(*c*) se $f(x_l)f(x_r) = 0$, a raiz é igual a x_r; pare os cálculos.

FIGURA 5.5

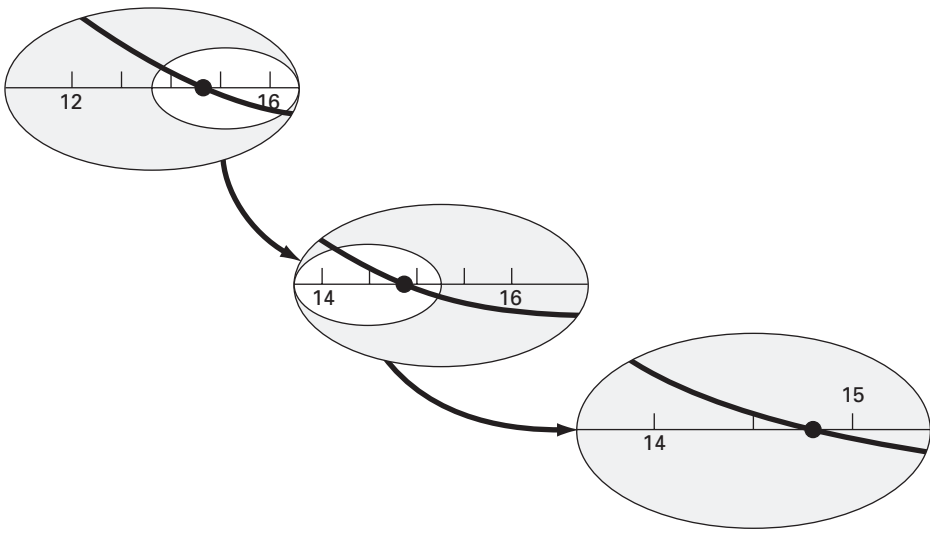


FIGURA 5.6
Uma descrição gráfica do método da bissecção. Esse gráfico mostra as primeiras três iterações do Exemplo 5.3.

EXEMPLO 5.3 Bissecção

Enunciado do Problema. Use bissecção para resolver o mesmo problema tratado graficamente no Exemplo 5.1.

Solução. O primeiro passo na bissecção é descobrir dois valores para a incógnita (no problema presente, c) que dê valores de $f(c)$ com sinais diferentes. A partir da Figura 5.1, vê-se que a função muda de sinal entre 12 e 16. Portanto, a estimativa inicial da raiz x_r é o ponto médio do intervalo

$$x_r = \frac{12 + 16}{2} = 14$$

Essa estimativa representa um erro relativo porcentual verdadeiro de $\varepsilon_t = 5{,}3\%$ (observe que o valor verdadeiro da raiz é 14,7802). A seguir, calcula-se o produto do valor da função no extremo inferior e no ponto médio:

$$f(12)\,f(14) = 6{,}067(1{,}569) = 9{,}517$$

que é maior do que zero e, portanto, não ocorre troca de sinal entre a extremidade inferior e o ponto médio. Conseqüentemente, a raiz deve estar localizada entre 14 e 16. Logo,

criamos um novo intervalo redefinindo a extremidade inferior como 14 e determinando uma estimativa revisada da raiz por

$$x_r = \frac{14 + 16}{2} = 15$$

o que representa um erro porcentual verdadeiro de $\varepsilon_t = 1{,}5\%$. O processo pode ser repetido para se obter estimativas refinadas. Por exemplo,

$$f(14)f(15) = 1{,}569(-0{,}425) = -0{,}666$$

Portanto, a raiz está entre 14 e 15. A extremidade superior é redefinida como 15, e a raiz estimada na terceira iteração é calculada por

$$x_r = \frac{14 + 15}{2} = 14{,}5$$

o que representa um erro relativo porcentual de $\varepsilon_t = 1{,}9\%$. O método pode ser repetido até que o resultado seja suficientemente acurado para satisfazer suas necessidades.

No exemplo anterior, pode-se notar que o erro verdadeiro não decresce com cada iteração. Entretanto, o intervalo no qual a raiz está localizada é dividido na metade em cada passo do processo. Como discutiremos na próxima seção, a largura do intervalo fornece uma estimativa exata do limitante superior do erro para o método da bissecção.

5.2.1 Critério de Parada e Estimativa do Erro

Terminamos o Exemplo 5.3 com a afirmação de que se poderia continuar o método para obter uma estimativa refinada da raiz. Agora é preciso desenvolver um critério objetivo para decidir quando parar o método.

Uma sugestão inicial poderia ser parar os cálculos quando o erro verdadeiro cair abaixo de algum nível pré-especificado. Como ilustração, no Exemplo 5.3, o erro relativo caiu de 5,3 para 1,9% durante os cálculos. A decisão pode ser parar quando o erro cair abaixo de 0,1%. Essa estratégia é falha porque os erros estimados no exemplo foram baseados no conhecimento da raiz verdadeira da função — o que não seria o caso em situações reais, pois não haveria motivo para usar o método se a raiz já fosse conhecida.

Portanto, é necessária uma estimativa de erro que não dependa do conhecimento prévio dessa raiz. Como visto anteriormente na Seção 3.3, um erro relativo porcentual aproximado ε_a pode ser calculado por [lembre-se da Equação (3.5)]

$$\varepsilon_a = \left| \frac{x_r^{\text{novo}} - x_r^{\text{velho}}}{x_r^{\text{novo}}} \right| 100\% \tag{5.2}$$

onde x_r^{novo} é a raiz da iteração atual e x_r^{velho} é a raiz da iteração prévia. O valor absoluto é usado porque, em geral, estamos interessados no módulo de ε_a em vez de no seu sinal. Quando ε_a se torna menor do que um critério de parada pré-especificado ε_s, param-se os cálculos.

EXEMPLO 5.4 Estimativa de Erro para a Bissecção

Enunciado do Problema. Continue o Exemplo 5.3 até que o erro aproximado fique abaixo do critério de parada $\varepsilon_s = 0{,}5\%$. Use a Equação (5.2) para calcular os erros.

Solução. Os resultados das primeiras duas iterações no Exemplo 5.3 foram 14 e 15. A substituição desses valores na Equação (5.2) fornece

$$|\varepsilon_a| = \left| \frac{15 - 14}{15} \right| 100\% = 6{,}667\%$$

Lembre que o erro relativo porcentual verdadeiro para a estimativa da raiz de 15 era 1,5%. Logo, ε_a é maior do que ε_t. Esse comportamento se manifesta nas outras iterações:

Iteração	x_l	x_u	x_r	ε_a (%)	ε_t (%)
1	12	16	14		5,279
2	14	16	15	6,667	1,487
3	14	15	14,5	3,448	1,896
4	14,5	15	14,75	1,695	0,204
5	14,75	15	14,875	0,840	0,641
6	14,75	14,875	14,8125	0,422	0,219

Assim, depois de seis iterações, ε_a finalmente fica abaixo de $\varepsilon_s = 0{,}5\%$, e os cálculos podem parar.

Esses resultados estão resumidos na Figura 5.7. A natureza "irregular" do erro verdadeiro se deve ao fato de que, na bissecção, a raiz verdadeira pode estar em qualquer lugar dentro do intervalo que a cerca. Os erros verdadeiro e aproximado estão bem distantes quando acontece de o intervalo estar centrado na raiz verdadeira. Eles estão próximos quando a raiz verdadeira está próxima de qualquer das extremidades.

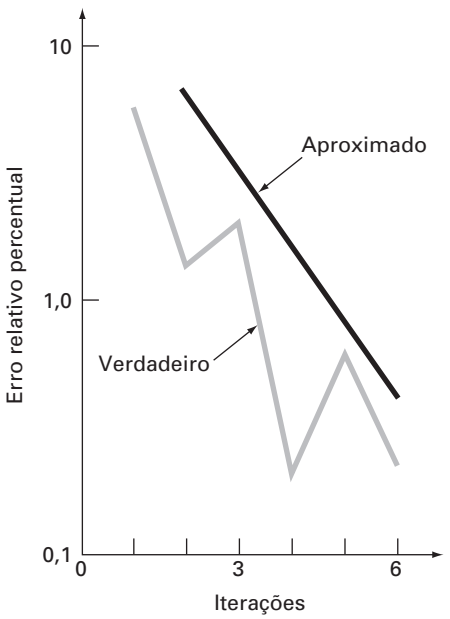
Embora o erro aproximado não forneça uma estimativa exata do erro verdadeiro, a Figura 5.7 sugere que ε_a captura a tendência geral para baixo de ε_t. Além disso, o gráfico mostra a característica extremamente atrativa de que ε_a é sempre maior do que ε_t. Logo, quando ε_a cai abaixo de ε_s, podemos parar os cálculos com a confiança de que a raiz é conhecida pelo menos tão acuradamente quanto o nível aceitável pré-especificado.

Embora seja sempre perigoso tirar conclusões gerais de um único exemplo, pode ser demonstrado que ε_a será sempre maior do que ε_t para o método da bissecção. Isso ocorre porque, cada vez que uma raiz aproximada é localizada usando a bissecção com $x_r = (x_l + x_u)/2$, sabemos que a raiz verdadeira está em algum ponto dentro de um intervalo de comprimento $(x_u - x_l)/2 = \Delta x/2$. Logo, a raiz deve estar a $\pm\Delta x/2$ de nossa estimativa (Figura 5.8). Por exemplo, quando o Exemplo 5.3 foi parado, poderíamos fazer a afirmação definitiva de que

$$x_r = 14{,}5 \pm 0{,}5$$

Como $\Delta x/2 = x_r^{\text{novo}} - x_r^{\text{velho}}$ (Figura 5.9), a Equação (5.2) fornece um limitante superior exato para o erro verdadeiro. Para que esse limitante fosse excedido, a raiz

FIGURA 5.7
Erros para o método da bissecção. Os erros verdadeiro e estimado estão mostrados em função do número de iterações.



verdadeira deveria estar fora do intervalo que a cerca, o que, por definição, nunca poderia ocorrer no método da bissecção. Como ilustrado no exemplo seguinte (Exemplo 5.7), outras técnicas de localizar raízes não se comportam sempre tão bem. Embora a bissecção seja em geral mais lenta do que outros métodos, a simplicidade de sua análise de erro é certamente um aspecto positivo que poderia torná-la atrativa para certas aplicações de engenharia.

Antes de prosseguir para programas de computador para a bissecção, deve-se observar que as relações (Figura 5.9)

$$x_r^{\text{novo}} - x_r^{\text{velho}} = \frac{x_u - x_l}{2}$$

e

$$x_r^{\text{novo}} = \frac{x_l + x_u}{2}$$

podem ser substituídas na Equação (5.2) para desenvolver uma formulação alternativa para o erro relativo porcentual

$$\varepsilon_a = \left| \frac{x_u - x_l}{x_u + x_l} \right| 100\% \tag{5.3}$$

Essa equação fornece resultados idênticos à Equação (5.2) para a bissecção. Além disso, ela permite calcular uma estimativa de erro com base nas aproximações iniciais — isto é, na primeira iteração. Por exemplo, na primeira iteração do Exemplo 5.2, um erro aproximado pode ser calculado como

$$\varepsilon_a = \left| \frac{16 - 12}{16 + 12} \right| 100\% = 14{,}29\%$$

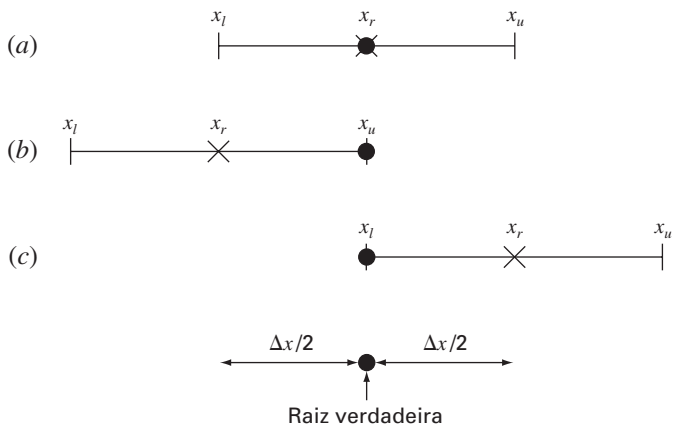


FIGURA 5.8
Três maneiras pelas quais o intervalo pode delimitar a raiz. Em (*a*) o valor verdadeiro está no centro do intervalo, ao passo que em (*b*) e em (*c*) o valor verdadeiro está próximo à extremidade. Observe que a discrepância entre o valor verdadeiro e o ponto médio do intervalo nunca excede a metade do comprimento do intervalo ou $\Delta x/2$.

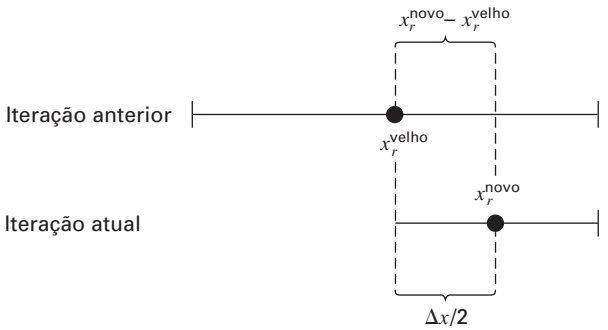


FIGURA 5.9
Descrição gráfica de por que o erro estimado para a bissecção ($\Delta x/2$) é equivalente à estimativa da raiz na iteração atual (x_r^{novo}) menos a raiz estimada na iteração anterior (x_r^{velho}).

Um outro benefício do método da bissecção é que o número de iterações necessárias para se chegar a um erro absoluto pode ser calculado *a priori* — isto é, antes de começar as iterações. Isso pode ser visto reconhecendo-se que, antes de começar a técnica, o erro absoluto é

$$E_a^0 = x_u^0 - x_l^0 = \Delta x^0$$

onde o sobrescrito indica a iteração. Então, antes de começar o método, estamos na "iteração zero". Depois da primeira iteração, o erro se torna

$$E_a^1 = \frac{\Delta x^0}{2}$$

Como cada iteração sucessiva divide o erro por dois, uma fórmula geral relacionando o erro e o número de iterações, n, é

$$E_a^n = \frac{\Delta x^0}{2^n} \tag{5.4}$$

Se $E_{a,d}$ é o erro desejado, essa equação pode ser resolvida por

$$n = \frac{\log\left(\Delta x^0 / E_{a,d}\right)}{\log 2} = \log_2\left(\frac{\Delta x^0}{E_{a,d}}\right) \tag{5.5}$$

Vamos testar a fórmula. Para o Exemplo 5.4, o intervalo inicial era $\Delta x_0 = 16 - 12 = 4$. Depois de seis iterações, o erro absoluto era

$$E_a = \frac{|14{,}875 - 14{,}75|}{2} = 0{,}0625$$

Pode-se substituir esses valores na Equação (5.5) para obter

$$n = \frac{\log(4/0{,}0625)}{\log 2} = 6$$

Logo, se soubermos antecipadamente que um erro de menos do que 0,0625 é aceitável, a fórmula diz que seis iterações forneceriam o resultado procurado.

Embora se tenha enfatizado o uso do erro relativo por razões óbvias, há casos em que (em geral, pelo conhecimento do contexto do problema) se poderá especificar um erro absoluto. Nesses casos, a bissecção junto com a Equação (5.5) podem fornecer um algoritmo útil na localização das raízes. Tais aplicações serão exploradas nos problemas do final do capítulo.

5.2.2 Algoritmo de Bissecção

O algoritmo na Figura 5.5 pode agora ser expandido para incluir uma verificação do erro (Figura 5.10). Esse algoritmo usa funções definidas pelo usuário para tornar a localização das raízes e os cálculos da função mais eficientes. Além disso, é colocado um limite superior no número de iterações. Finalmente, é incluída uma verificação do erro para evitar divisão por zero durante o cálculo do erro. Em tal situação, a Equação (5.2) se torna infinita. Se isso ocorrer, o programa pula o cálculo do erro nessa iteração.

O algoritmo na Figura 5.10 não é amigável ao usuário; ele é planejado estritamente para produzir a resposta correta. No Problema 5.14 no final deste capítulo, você terá a tarefa de torná-lo mais fácil de usar e entender.

5.2.3 Minimizando o Cálculo de Funções

O algoritmo de bissecção na Figura 5.10 é muito bom se você estiver calculando uma única raiz para uma função que seja fácil de calcular. Entretanto, há muitos exemplos em engenharia nos quais esse não é o caso. Por exemplo, suponha que você desenvolveu um programa de computador que precisa localizar uma raiz diversas vezes. Nesses casos, você poderia chamar o algoritmo da Figura 5.10 milhares e mesmo milhões de vezes durante uma única execução.

```
FUNCTION Bissecção(xl, xu, es, imax, xr, iter, ea)
  iter = 0
  DO
    xrold = xr
    xr = (xl + xu) / 2
    iter = iter + 1
    IF xr ≠ 0 THEN
       ea = ABS((xr − xrold) / xr) * 100
    END IF
    test = f(xl) * f(xr)
    IF test < 0 THEN
       xu = xr
    ELSE IF test > 0 THEN
       xl = xr
    ELSE
       ea = 0
    END IF
    IF ea < es OR iter ≥ imax EXIT
  END DO
  Bissecção = xr
END Bissecção
```

FIGURA 5.10
Pseudocódigo para a função que implementa a bissecção.

Além disso, no seu sentido mais geral, uma função de uma única variável é simplesmente uma entidade que devolve um único valor para um único valor que você tenha inserido. Entendidas nesse sentido, as funções não são sempre fórmulas simples como as equações de uma linha resolvidas nos exemplos anteriores deste capítulo. Por exemplo, uma função poderia consistir em muitas linhas de código que poderiam envolver uma quantidade significativa de tempo de execução para calcular. E, em alguns casos, a função poderia até mesmo representar um programa de computador independente.

FIGURA 5.11
Pseudocódigo para o subprograma de bissecção que minimiza o cálculo de funções.

```
FUNCTION Bissecção(xl, xu, es, imax, xr, iter, ea)
  iter = 0
  fl = f(xl)
  DO
    xrold = xr
    xr = (xl + xu) / 2
    fr = f(xr)
    iter = iter + 1
    IF xr ≠ 0 THEN
      ea = ABS((xr − xrold) / xr) * 100
    END IF
    test = fl * fr
    IF test < 0 THEN
      xu = xr
    ELSE IF test > 0 THEN
      xl = xr
      fl = fr
    ELSE
      ea = 0
    END IF
    IF ea < es OR iter ≥ imax EXIT
  END DO
  Bissecção = xr
END Bissecção
```

Por causa de ambos os fatores, é imperativo que os algoritmos numéricos minimizem os cálculos de funções. Desse ponto de vista, o algoritmo da Figura 5.10 é deficiente. Observe, em particular, que ao fazer dois cálculos de função por iteração, ele recalcula uma das funções que havia sido determinada na iteração anterior.

A Figura 5.11 fornece um algoritmo modificado que não tem essa deficiência. Destacamos as linhas que são diferentes das da Figura 5.10. Nesse caso, é calculado apenas um novo valor da função na estimativa da raiz. Os valores calculados anteriormente são salvos e meramente transferidos conforme o intervalo diminui. Então, são feitos $n + 1$ cálculos de funções, em vez de $2n$.

5.3 O MÉTODO DA FALSA POSIÇÃO

Embora a bissecção seja uma técnica perfeitamente válida para determinar raízes, sua abordagem do tipo "força bruta" é relativamente ineficiente. A falsa posição é uma alternativa baseada na percepção gráfica.

Uma deficiência do método da bissecção é que, na divisão do intervalo de x_l a x_u em metades iguais, não são levados em conta os módulos de $f(x_l)$ e $f(x_u)$. Por exemplo, se $f(x_l)$ estiver muito mais próximo de zero do que $f(x_u)$, será provável que a raiz esteja mais próxima de x_l do que de x_u (Figura 5.12). Um método alternativo que explora essa percepção gráfica é ligar $f(x_l)$ e $f(x_u)$ por uma reta. A intersecção dessa reta com o eixo x representa uma estimativa melhorada da raiz. O fato de a substituição da curva por uma reta dar uma "falsa posição" da raiz é a origem do nome, *método da falsa posição*, ou, em latim, *regula falsi*. Ele também é chamado de *método da interpolação linear*.

Usando triângulos semelhantes (Figura 5.12), a intersecção da reta com o eixo x pode ser calculada por

$$\frac{f(x_l)}{x_r - x_l} = \frac{f(x_u)}{x_r - x_u} \tag{5.6}$$

que pode ser reescrita como (ver Quadro 5.1 para detalhes):

$$x_r = x_u - \frac{f(x_u)(x_l - x_u)}{f(x_l) - f(x_u)} \tag{5.7}$$

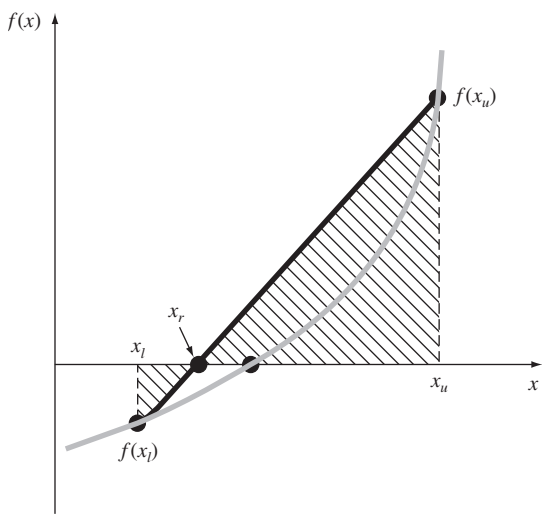


FIGURA 5.12
Uma descrição gráfica do método da falsa posição. Os triângulos semelhantes que foram usados para deduzir a fórmula do método estão hachurados.

Quadro 5.1 Dedução do Método da Falsa Posição

Multiplique a Equação (5.6) em cruz para obter

$$f(x_l)(x_r - x_u) = f(x_u)(x_r - x_l)$$

Agrupe os termos e reorganize:

$$x_r\,[f(x_l) - f(x_u)] = x_u\,f(x_l) - x_l\,f(x_u)$$

Divida por $f(x_l) - f(x_u)$:

$$x_r = \frac{x_u\,f(x_l) - x_l\,f(x_u)}{f(x_l) - f(x_u)} \qquad \text{(Q5.1.1)}$$

Esta é uma forma do método da falsa posição. Observe que ela permite o cálculo da raiz x_r como uma função das aproximações inferior x_l e superior x_u. Ele pode ser colocado em uma forma alternativa expandindo-o:

$$x_r = \frac{x_u\,f(x_l)}{f(x_l) - f(x_u)} - \frac{x_l\,f(x_u)}{f(x_l) - f(x_u)}$$

e então somando e subtraindo x_u no lado direito:

$$x_r = x_u + \frac{x_u\,f(x_l)}{f(x_l) - f(x_u)} - x_u - \frac{x_l\,f(x_u)}{f(x_l) - f(x_u)}$$

Agrupando os termos obtemos

$$x_r = x_u + \frac{x_u\,f(x_u)}{f(x_l) - f(x_u)} - \frac{x_l\,f(x_u)}{f(x_l) - f(x_u)}$$

ou

$$x_r = x_u - \frac{f(x_u)(x_l - x_u)}{f(x_l) - f(x_u)}$$

o que é o mesmo que a Equação (5.7). Essa forma foi usada porque ela envolve um cálculo de função a menos e uma multiplicação a menos do que a Equação (Q5.1.1). Além disso, ela é diretamente comparável ao método da secante, que será discutido no Capítulo 6.

Essa é a *fórmula da falsa posição*. O valor de x_r calculado com a Equação (5.7) então substitui qualquer das duas aproximações iniciais x_l ou x_u que forneça um valor da função com o mesmo sinal que $f(x_r)$. Dessa forma, os valores x_l e x_u sempre delimitam a raiz verdadeira. O processo é repetido até que a raiz seja estimada adequadamente. O algoritmo é idêntico ao da bissecção (Figura 5.5), com a diferença de que a Equação (5.7) é usada no passo 2. Além disso, é usado o mesmo critério de parada [Equação(5.2)] para encerrar os cálculos.

EXEMPLO 5.5 Falsa Posição

Enunciado do Problema. Use o método da falsa posição para determinar a raiz da mesma equação investigada no Exemplo 5.1 [Equação (E5.1.1)].

Solução. Como no Exemplo 5.3, comece os cálculos com aproximações $x_l = 12$ e $x_u = 16$.

Primeira iteração:

$$x_l = 12 \qquad f(x_l) = 6{,}0699$$
$$x_u = 16 \qquad f(x_u) = -2{,}2688$$
$$x_r = 16 - \frac{-2{,}2688(12 - 16)}{6{,}0669 - (-2{,}2688)} = 14{,}9113$$

que tem um erro relativo verdadeiro de 0,89%.

Segunda iteração:

$$f(x_l)\,f(x_r) = -1{,}5426$$

Portanto, a raiz está no primeiro subintervalo e x_r se torna a aproximação superior da próxima iteração, $x_u = 14{,}9113$:

$$x_l = 12 \qquad f(x_l) = 6{,}0699$$
$$x_u = 14{,}9113 \qquad f(x_u) = -0{,}2543$$
$$x_r = 14{,}9113 - \frac{-0{,}2543(12 - 14{,}9113)}{6{,}0669 - (-0{,}2543)} = 14{,}7942$$

que tem erros relativos verdadeiro e aproximado de 0,09 e 0,79%. Iterações adicionais podem ser feitas para refinar a estimativa da raiz.

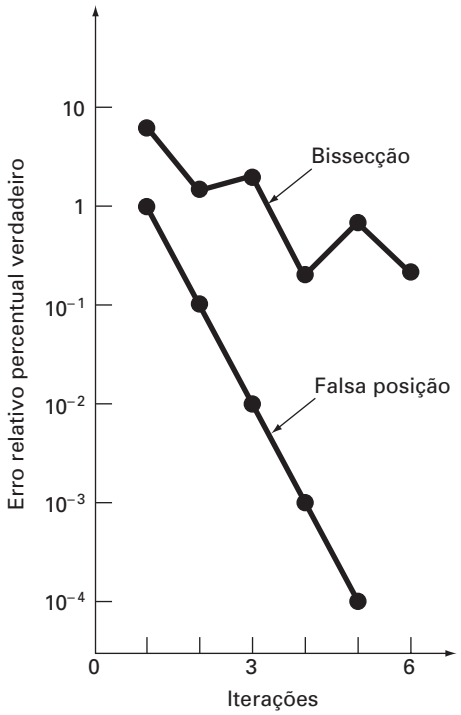


FIGURA 5.13
Comparação dos erros relativos dos métodos da bissecção e da falsa posição.

Uma impressão sobre a eficiência relativa dos métodos da bissecção e falsa posição pode ser obtida consultando-se a Figura 5.13, em que estão marcados os erros relativos percentuais verdadeiros para os Exemplos 5.4 e 5.5. Observe como o erro na falsa posição diminui muito mais rapidamente do que na bissecção por causa do esquema mais eficiente para a localização da raiz no método da falsa posição.

Lembre que, no método da bissecção, o intervalo entra x_l e x_u se tornava menor durante o cálculo. O intervalo, definido por $\Delta x/2 = |x_u - x_l|/2$ na primeira iteração, fornece, portanto, uma medida do erro nessa abordagem. Esse não é o caso no método da falsa posição porque uma das aproximações iniciais pode permanecer fixa durante o cálculo enquanto a outra converge para a raiz. Como ilustração, no Exemplo 5.6 a aproximação inferior x_l permanece em 12 enquanto x_u converge para a raiz. Para tais casos, o comprimento do intervalo não se contrai para zero, mas, em vez disso, aproxima-se de um valor constante.

O Exemplo 5.6 sugere que a Equação (5.2) representa um critério de erro muito conservador. Na realidade, a Equação (5.2) constitui de fato uma aproximação da discrepância da iteração anterior. Isso ocorre porque, para um caso tal como o Exemplo 5.6, no qual o método está convergindo rapidamente (por exemplo, o erro está sendo reduzido quase uma ordem de grandeza por iteração), a raiz da iteração atual x_r^{novo} é uma estimativa muito melhor do valor verdadeiro do que o resultado da iteração prévia x_r^{velho}. Assim, a quantidade no numerador da Equação (5.2) na realidade representa a discrepância da iteração anterior. Conseqüentemente, podemos estar certos de que o fato de a Equação (5.2) ser satisfeita garante que a raiz será conhecida com precisão maior do que a tolerância prescrita. Entretanto, como será descrito na próxima seção, há casos em que o método da falsa posição converge lentamente. Nesses casos, a Equação (5.2) se torna não-confiável e um critério de parada alternativo deve ser deduzido.

5.3.1 Armadilhas do Método da Falsa Posição

Embora o método da falsa posição possa parecer sempre ser o melhor dos métodos intervalares, há casos em que seu desempenho é deficiente. Na realidade, como ocorre no próximo exemplo, há certos casos nos quais a bissecção fornece resultados superiores.

EXEMPLO 5.6 Um Caso em que a Bissecção É Preferível à Falsa Posição

Enunciado do Problema. Use a bissecção e a falsa posição para localizar a raiz de

$$f(x) = x^{10} - 1$$

entre $x = 0$ e 1,3.

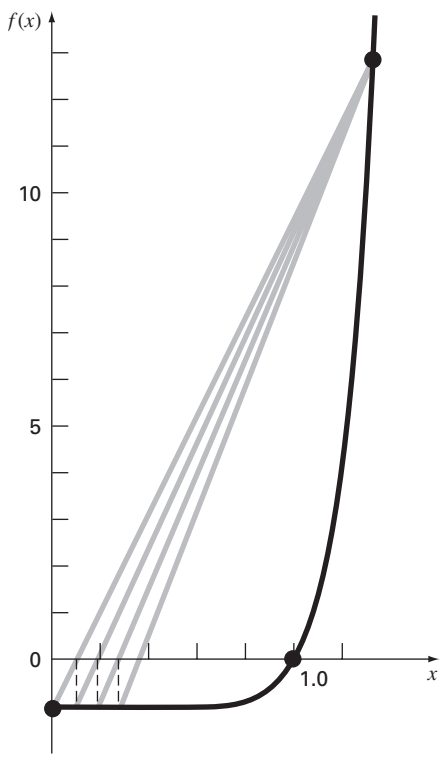
Solução. Usando a bissecção, os resultados podem ser resumidos por

Iteração	x_l	x_u	x_r	ε_a (%)	ε_t (%)
1	0	1,3	0,65	100,0	35,0
2	0,65	1,3	0,975	33,3	2,5
3	0,975	1,3	1,1375	14,3	13,8
4	0,975	1,1375	1,05625	7,7	5,6
5	0,975	1,05625	1,015625	4,0	1,6

Logo, depois de cinco iterações, o erro verdadeiro foi reduzido para menos de 2%. Na falsa posição, é obtida uma saída muito diferente:

Iteração	x_l	x_u	x_r	ε_a (%)	ε_t (%)
1	0	1,3	0,09430		90,6
2	0,09430	1,3	0,18176	48,1	81,8
3	0,18176	1,3	0,26287	30,9	73,7
4	0,26287	1,3	0,33811	22,3	66,2
5	0,33811	1,3	0,40788	17,1	59,2

FIGURA 5.14
Gráfico de $f(x) = x^{10} - 1$, ilustrando a convergência lenta do método da falsa posição.



Depois de cinco iterações. o erro verdadeiro diminuiu apenas para cerca de 59%. Além disso, observe que $\varepsilon_a < \varepsilon_t$. Logo, o erro aproximado é enganoso. Pode-se ganhar alguma intuição sobre esses resultados examinando o gráfico da função. Como na Figura 5.14, a curva viola a hipótese básica na qual a falsa posição foi baseada — isto é, se $f(x_l)$ está muito mais próximo de zero do que $f(x_u)$, então a raiz está mais próxima de x_l do que de x_u (lembre-se da Figura 5.12). Por causa da forma da função presente, o contrário é verdadeiro.

O exemplo anterior ilustra que generalizações abrangentes com relação aos métodos de localização de raízes, em geral, não são possíveis. Embora um método como a falsa posição seja quase sempre superior à bissecção, invariavelmente há casos que violam essas conclusões gerais. Portanto, além de se usar a Equação (5.2), os resultados deveriam sempre ser verificados substituindo a estimativa da raiz na equação original e verificando se o resultado está próximo a zero. Tal verificação deveria ser incorporada em todos os programas de computador para a localização de raízes.

O exemplo também ilustra uma grande fraqueza do método da falsa posição: ele é unilateral. Ou seja, conforme as iterações continuam, uma das extremidades do intervalo terá a tendência de permanecer fixa. Isso pode levar à convergência insatisfatória, particularmente para funções com curvatura significativa.

5.3.2 Falsa Posição Modificado

Uma forma de abrandar a natureza "unilateral" da falsa posição é fazer o algoritmo detectar quando uma das extremidades está presa. Se isso ocorrer, o valor da função na

FIGURA 5.15
Pseudocódigo para o método da falsa posição modificado.

```
FUNCTION FalsaPosMod(xl, xu, es, imax, xr, iter, ea)
  iter = 0
  fl = f(xl)
  fu = f(xu)
  DO
    xrold = xr
    xr = xu − fu * (xl − xu) / (fl − fu)
    fr = f(xr)
    iter = iter + 1
    IF xr <> 0 THEN
      ea = Abs((xr − xrold) / xr) * 100
    END IF
    test = fl * fr
    IF test < 0 THEN
      xu = xr
      fu = f(xu)
      iu = 0
      il = il + 1
      If il ≥ 2 THEN fl = fl / 2
    ELSE IF test > 0 THEN
      xl = xr
      fl = f(xl)
      il = 0
      iu = iu + 1
      IF iu ≥ 2 THEN fu = fu / 2
    ELSE
      ea = 0
    END IF
    IF ea < es OR iter ≥ imax THEN EXIT
  END DO
  FalsaPosMod = xr
END FalsaPosMod
```

extremidade presa pode ser dividido por dois, o que é chamado de *método da falsa posição modificado*.

O algoritmo na Figura 5.15 implementa essa estratégia. Observe como contadores são usados para determinar quando uma das extremidades permanece fixa por duas iterações. Se isso ocorrer, o valor da função na extremidade presa é dividido por dois.

A eficiência desse algoritmo pode ser ilustrada aplicando-o ao Exemplo 5.6. Se um critério de parada de 0,01% for usado, os métodos da bissecção e da falsa posição convergiriam em 14 e 39 iterações, respectivamente. Em contraste, o método da falsa posição modificado convergiria em 12 iterações. Portanto, para esse exemplo, ele é um pouco mais eficiente do que a bissecção e é muito superior ao método da falsa posição não-modificado.

5.4 BUSCAS INCREMENTAIS E DETERMINAÇÃO DE APROXIMAÇÕES INICIAIS

Além de verificar uma resposta individual, você deve determinar se todas as raízes possíveis foram localizadas. Como mencionado anteriormente, um gráfico da função, em geral, é muito útil como guia nessa tarefa. Uma outra opção é incorporar uma busca incremental no início do programa de computador. Isso consiste em começar em uma extremidade da região de interesse e então calcular a função em pequenos incrementos através da região. Quando a função muda de sinal, assume-se que uma raiz está dentro do incremento. Os valores de x no início e no fim do incremento podem, então, servir como aproximações iniciais para uma das técnicas descritas neste capítulo.

Um problema potencial para um busca incremental é a escolha do comprimento do incremento. Se o comprimento for pequeno demais, a busca pode gastar muito tempo. Por outro lado, se o comprimento for grande demais, existe a possibilidade de que raízes próximas possam ser perdidas (Figura 5.16). O problema é composto com a possível existência de raízes múltiplas. Um remédio parcial para tais casos é calcular a primeira derivada da função $f'(x)$ no início e no fim de cada intervalo. Se a derivada mudar de sinal, isso sugere que ocorreu um mínimo ou um máximo e que o intervalo deve ser examinado com mais atenção para se verificar a existência de uma possível raiz.

Embora tais modificações ou o uso de incrementos muito pequenos possam diminuir o problema, deve ficar claro que métodos do tipo força bruta, como a busca incremental, não são a prova de erros. Seria sensato suplementar tais técnicas automáticas com outras informações que forneçam uma intuição na localização das raízes. Essas informações podem ser encontradas fazendo-se gráficos e entendendo o problema físico do qual a equação se originou.

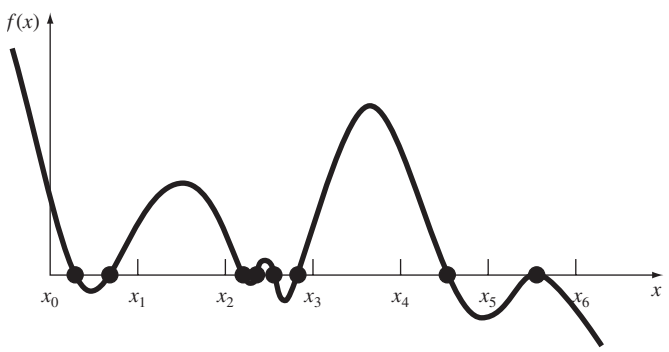


FIGURA 5.16
Casos nos quais se podem perder raízes em razão de o comprimento do incremento do procedimento de busca ser grande demais. Observe que a última raiz à direita é múltipla e seria perdida independentemente do comprimento do incremento.

PROBLEMAS

5.1 Determine as raízes reais de $f(x) = -0{,}5x^2 + 2{,}5x + 4{,}5$:
(a) Graficamente.
(b) Usando a fórmula quadrática.
(c) Usando três iterações do método da bissecção para determinar a maior raiz. Use as aproximações iniciais $x_l = 5$ e $x_u = 10$. Calcule o erro estimado ε_a e o erro verdadeiro ε_t depois de cada iteração.

5.2 Determine a raiz real de $f(x) = 5x^3 - 5x^2 + 6x - 2$:
(a) Graficamente.
(b) Usando o método da bissecção para localizar a raiz. Use as aproximações iniciais $x_l = 0$ e $x_u = 1$ e itere até que o erro estimado ε_a fique abaixo de um nível $\varepsilon_s = 10\%$.

5.3 Determine a raiz real de $f(x) = -25 + 82x - 90x^2 + 44x^3 - 8x^4 + 0{,}7x^5$:
(a) Graficamente.
(b) Usando o método da bissecção para determinar a raiz até $\varepsilon_s = 10\%$. Use as aproximações iniciais $x_l = 0{,}5$ e $x_u = 1{,}0$.
(c) Faça os mesmos cálculos que em **(b)** mas use o método da falsa posição e $\varepsilon_s = 0{,}2\ \%$.

5.4 **(a)** Determine as raízes de $f(x) = -12 - 21x + 18x^2 - 2{,}75x^3$ graficamente. Além disso, determine a primeira raiz da função pela **(b)** bissecção e **(c)** falsa posição. Para **(b)** e **(c)**, use as aproximações iniciais $x_l = -1$ e $x_u = 0$ e um critério de parada de 1%.

5.5 Localize a primeira raiz não-trivial de sen $x = x^3$, onde x está em radianos. Use uma técnica gráfica e a bissecção com o intervalo inicial de 0,5 a 1. Faça os cálculos até que ε_a seja menor do que $\varepsilon_s = 2\%$. Faça também uma verificação do erro substituindo sua resposta final na equação original.

5.6 Determine a raiz real positiva de $\ln(x^4) = 0{,}7$ **(a)** graficamente, **(b)** usando três iterações do método da bissecção, com aproximações iniciais $x_l = 0{,}5$ e $x_u = 2$, e **(c)** usando três iterações do método da falsa posição, com as mesmas aproximações iniciais que em **(b)**.

5.7 Determine a raiz real de $f(x) = (0{,}8 - 0{,}3x)/x$:
(a) Analiticamente.
(b) Graficamente.
(c) Usando três iterações do método da falsa posição e aproximações iniciais 1 e 3. Calcule o erro aproximado ε_a e o erro verdadeiro ε_t depois de cada iteração. Há algum problema com o resultado?

5.8 Encontre a raiz quadrada positiva de 18 usando o método da falsa posição até $\varepsilon_s = 0{,}5\%$. Use aproximações iniciais $x_l = 4$ e $x_u = 5$.

5.9 Encontre a menor raiz positiva da função $x^2\left|\cos\sqrt{x}\right| = 5$ (x está em radianos) usando o método da falsa posição. Para localizar a região na qual estão as raízes, inicialmente trace a função para valores de x entre 0 e 5. Faça os cálculos até que ε_a fique abaixo de $\varepsilon_s = 1\%$. Verifique sua resposta final substituindo-a na função original.

5.10 Encontre a raiz real positiva de $f(x) = x^4 - 8x^3 - 35x^2 + 450x - 1001$ usando o método da falsa posição. Use aproximações iniciais $x_l = 4{,}5$ e $x_u = 6$ e faça cinco iterações. Calcule os erros verdadeiro e aproximado baseado no fato de que a raiz é 5,60979. Use um gráfico para explicar seus resultados e faça os cálculos até $\varepsilon_s = 1{,}0\%$.

5.11 Determine a raiz real de $x^{3{,}5} = 80$: **(a)** analiticamente e **(b)** pelo método da falsa posição até $\varepsilon_s = 2{,}5\%$. Use aproximações iniciais de 2,0 e 5,0.

5.12 Dada

$$f(x) = -2x^6 - 1{,}5x^4 + 10x + 2$$

Use a bissecção para determinar o *máximo* dessa função. Use aproximações iniciais $x_l = 0$ e $x_u = 1$ e faça iterações até que o erro relativo aproximado fique abaixo de 5%.

5.13 A velocidade v do pára-quedista em queda livre é dada por

$$v = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right)$$

onde $g = 9{,}8\ \text{m/s}^2$. Para um pára-quedista com um coeficiente de arrasto $c = 15$ Kg/s calcule a massa m para que a velocidade seja $v = 35$ m/s em $t = 9$ s. Use o método da falsa posição para determinar m até o nível de $\varepsilon_s = 0{,}1\%$.

5.14 Uma viga está sob carga como mostrado na Figura P5.14. Use o método da bissecção para determinar a posição dentro da viga onde não existe nenhum torque.

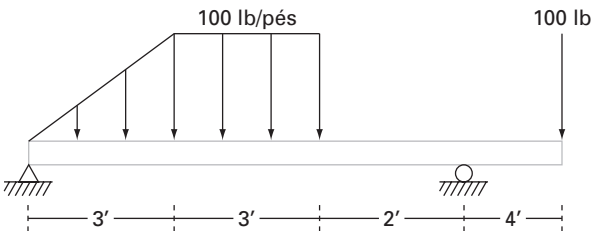


Figura P5.14

5.15 Água está escoando em um canal trapezoidal a uma vazão de $Q = 20\ \text{m}^3/\text{s}$. A profundidade crítica y para tal canal deve satisfazer a equação

$$0 = 1 - \frac{Q^2}{gA_c^3}B$$

onde $g = 9{,}81\ \text{m/s}^2$, A_c é a área da seção transversal (m^2), e B é a largura do canal na superfície (m). Para esse caso, a largura e a área transversal podem ser relacionadas à profundidade y por

$$B = 3 + y \qquad \text{e} \qquad A_c = 3y + \frac{y^2}{2}$$

Encontre a profundidade crítica usando **(a)** o método gráfico, **(b)** a bissecção e **(c)** a falsa posição. Para **(b)** e **(c),** use aproximações iniciais de $x_l = 0{,}5$ e $x_u = 2{,}5$, e itere até que o erro aproximado fique abaixo de 1% ou que o número de iterações ultrapasse 10. Discuta seus resultados.

5.16 Você está projetando um tanque esférico (Figura P5.16) para armazenar a água para uma pequena vila em uma região em desenvolvimento. O volume de líquido que ele armazena pode ser calculado por

$$V = \pi h^2\frac{[3R - h]}{3}2$$

onde V é o volume [m^3], h é a profundidade da água no tanque [m], R é o raio do tanque [m].

Se $R = 3$ m, até que profundidade o tanque deve estar cheio para que ele armazene 30 m^3? Use três iterações do método da falsa posição para determinar sua resposta. Determine o erro relativo aproximado depois de cada iteração.

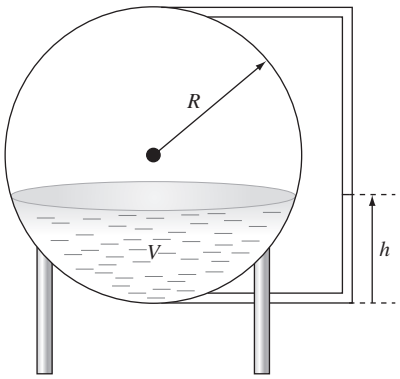


Figura P5.16

5.17 O coeficiente de saturação do oxigênio dissolvido em água fresca pode ser calculado pela equação (APHA,1992)

$$\ln o_{sf} = -139{,}34411 + \frac{1{,}575701 \times 10^5}{T_a} - \frac{6{,}642308 \times 10^7}{T_a^2} + \frac{1{,}243800 \times 10^{10}}{T_a^3} - \frac{8{,}621949 \times 10^{11}}{T_a^4}$$

onde o_{sf} é a concentração de saturação do oxigênio dissolvido em água fresca a 1 atm (mg/L) e T_a é a temperatura absoluta (K). Lembre-se de que $T_a = T + 273{,}15$, onde T é a temperatura (°C). De acordo com essa equação, a saturação diminui com o aumento da temperatura. Para águas naturais típicas em climas temperados, a equação pode ser usada para determinar a variação da concentração de oxigênio de 14,621 mg/L a 0° C até 6,413 mg/L a 40° C. Dado um valor da concentração de oxigênio, essa fórmula e o método da bissecção podem ser usados para determinar a temperatura em ° C.

(a) Se as aproximações iniciais forem tomadas como 0 e 40° C, quantas iterações da bissecção seriam necessárias para determinar a temperatura até um erro absoluto de 0,05° C?

(b) Desenvolva e teste um programa de bissecção para determinar T como uma função de uma dada concentração de oxigênio até um erro absoluto pré-especificado como em **(a)**. Dadas as aproximações iniciais 0 e 40° C, teste seu programa para um erro absoluto = 0,05° C e os seguintes casos: $o_{sf} = 8$, 10 e 12 mg/L. Verifique seus resultados.

5.18 Integre o algoritmo delineado na Figura 5.10 em um subprograma de bissecção completo, amigável ao usuário. Entre outras coisas:

(a) Coloque comentários de documentação em todo o subprograma para identificar o que foi planejado para cada seção realizar.

(b) Identifique a entrada e a saída.

(c) Acrescente uma verificação da resposta que substitua a estimativa da raiz na função original para verificar se o resultado final está próximo de zero.

(d) Teste o subprograma repetindo os cálculos dos Exemplos 5.3 e 5.4.

5.19 Desenvolva um subprograma para o método da bissecção que minimize cálculos de funções, baseado no pseudocódigo da Figura 5.11. Determine o número de cálculos de funções (n) para cada total de iterações. Teste o programa repetindo o Exemplo 5.6.

5.20 Desenvolva um programa amigável ao usuário para o método da falsa posição. A estrutura de seu programa deve ser parecida com o algoritmo de bissecção esboçado na Figura 5.10. Teste seu programa repetindo o Exemplo 5.5.

5.21 Desenvolva um subprograma para o método da falsa posição que minimize os cálculos de funções de uma forma parecida com a da Figura 5.11. Determine o número de cálculos de funções (n) para cada total de iterações. Teste o programa repetindo o Exemplo 5.6.

5.22 Desenvolva um programa amigável ao usuário para o método da falsa posição modificado, baseado na Figura 5.15. Teste seu programa determinando a raiz da função descrita no Exemplo 5.6. Faça diversas execuções até que o erro relativo porcentual verdadeiro fique abaixo de 0,01%. Marque em um gráfico os erros relativos percentuais verdadeiros e aproximados em função do número de iterações em um papel semilog. Interprete os seus resultados.

CAPÍTULO 6

Métodos Abertos

Nos métodos intervalares do capítulo anterior, a raiz era localizada dentro de um intervalo prescrito por um limitante inferior e um superior. As aplicações repetidas desses métodos sempre resultam em estimativas mais próximas do valor verdadeiro da raiz. Tais métodos são ditos *convergentes* porque se aproximam da verdade à medida que os cálculos prosseguem (Figura 6.1*a*).

Em contraste, os *métodos abertos* descritos neste capítulo são baseados em fórmulas que exigem apenas um único valor inicial de x ou dois valores iniciais que não delimitam necessariamente a raiz. Como tal, eles algumas vezes *divergem* ou se afastam da raiz verdadeira à medida que os cálculos prosseguem (Figura 6.1*b*). Entretanto, quando os métodos abertos convergem (Figura 6.1*c*), eles em geral o fazem muito mais rapidamente do que os métodos intervalares. Iniciaremos nossa discussão das técnicas abertas com uma versão simples, que será útil para ilustrar sua forma geral e também o conceito de convergência.

6.1 ITERAÇÃO DE PONTO FIXO SIMPLES

Como mencionado na abertura deste capítulo, os métodos abertos usam uma fórmula para prever a raiz. Tal fórmula pode ser deduzida para a *iteração de ponto fixo* simples (ou, como também é chamada, iteração de um ponto, substituições sucessivas ou aproximações

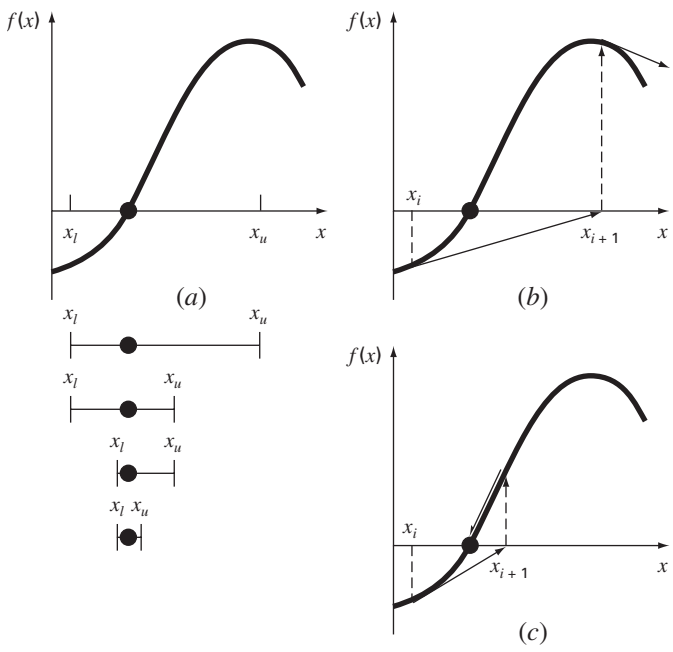


FIGURA 6.1
Descrição gráfica da diferença fundamental entre os métodos (*a*) intervalares e (*b*) e (*c*) abertos para a localização de raízes. Em (*a*), que é o método da bissecção, a raiz está restrita ao interior do intervalo definido por x_l e x_u. Em contraste, para o método aberto descrito em (*b*) e (*c*), é usada uma fórmula para avançar de x_i a x_{i+1} de forma iterativa. Assim, o método pode ou (*b*) divergir ou (*c*) convergir rapidamente, dependendo do valor da aproximação inicial.

sucessivas) reescrevendo a equação $f(x) = 0$ de modo que x esteja isolado no lado esquerdo da equação:

$$x = g(x) \tag{6.1}$$

Pode-se conseguir essa transformação ou por manipulação algébrica ou simplesmente somando x em ambos os lados da equação original. Por exemplo,

$$x^2 - 2x + 3 = 0$$

pode ser manipulada de forma simples para obtermos

$$x = \frac{x^2 + 3}{2}$$

enquanto sen $x = 0$ pode ser colocada na forma da Equação (6.1) somando x a ambos os lados para obter

$$x = \operatorname{sen} x + x$$

A utilidade da Equação (6.1) é que ela fornece uma fórmula para prever um novo valor de x em função de um velho valor de x. Portanto, dada uma aproximação inicial para a raiz x_i, a Equação (6.1) pode ser usada para calcular uma nova estimativa x_{i+1} expressa pela fórmula iterativa

$$x_{i+1} = g(x_i) \tag{6.2}$$

Como com as outras fórmulas iterativas deste livro, o erro aproximado para essa equação pode ser determinado usando-se o estimador de erro [Equação (3.5)]:

$$\varepsilon_a = \left| \frac{x_{i+1} - x_i}{x_{i+1}} \right| 100\%$$

EXEMPLO 6.1 Iteração de Ponto Fixo Simples

Enunciado do Problema. Use a iteração de ponto fixo simples para localizar a raiz de $f(x) = e^{-x} - x$.

Solução. A função pode ser separada diretamente e expressa na forma da Equação (6.2) como

$$x_{i+1} = e^{-x_i}$$

Começando com uma aproximação inicial $x_0 = 0$, essa equação iterativa pode ser usada para calcular

i	x_i	ε_a (%)	ε_t (%)
0	0		100,0
1	1,000000	100,0	76,3
2	0,367879	171,8	35,1
3	0,692201	46,9	22,1
4	0,500473	38,3	11,8
5	0,606244	17,4	6,89
6	0,545396	11,2	3,83
7	0,579612	5,90	2,20
8	0,560115	3,48	1,24
9	0,571143	1,93	0,705
10	0,564879	1,11	0,399

Assim, cada iteração traz o valor estimado para mais perto do valor verdadeiro: 0,56714329.

6.1.1 Convergência

Observe que o erro relativo porcentual verdadeiro para cada iteração do Exemplo 6.1 é aproximadamente proporcional (por um fator de cerca de 0,5 a 0,6) ao erro da iteração anterior. Essa propriedade, chamada de *convergência linear*, é característica da iteração de ponto fixo.

Além de comentarmos sobre a "taxa" de convergência, nesse ponto devemos comentar sobre a "possibilidade" de convergência. Os conceitos de convergência e divergência podem ser descritos graficamente. Lembre-se de que na Seção 5.1 traçamos o gráfico de uma função para visualizar sua estrutura e seu comportamento (Exemplo 5.1). Tal abordagem é usada na Figura 6.2*a* para a função $f(x) = e^{-x} - x$. Uma abordagem gráfica alternativa é separar a equação em duas componentes, como em

$$f_1(x) = f_2(x)$$

Então, as duas equações

$$y_1 = f_1(x) \tag{6.3}$$

e

$$y_2 = f_2(x) \tag{6.4}$$

podem ser traçadas separadamente (Figura 6.2*b*). Os valores de x correspondentes às intersecções dessas funções representam as raízes de $f(x) = 0$.

EXEMPLO 6.2 O Método Gráfico das Duas Curvas

Enunciado do Problema. Separe a equação $e^{-x} - x = 0$ em duas partes e determine suas raízes graficamente.

Solução. Reformule a equação como $y_1 = x$ e $y_2 = e^{-x}$. Os seguintes valores podem ser calculados:

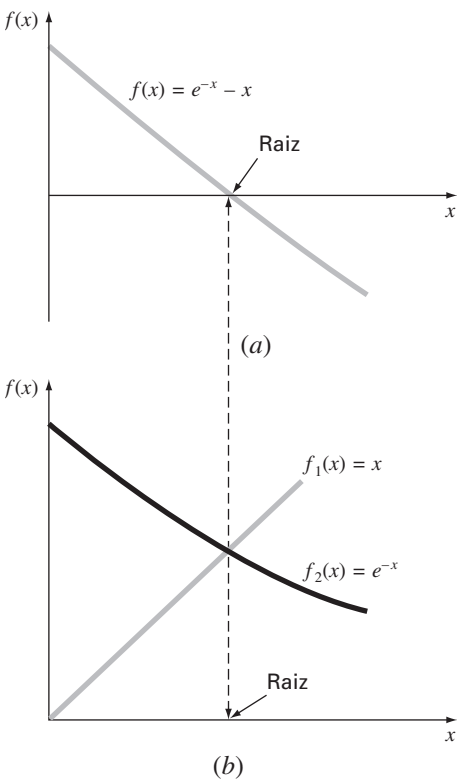


FIGURA 6.2
Dois métodos gráficos alternativos para determinar as raízes de $f(x) = e^{-x} - x$. (*a*) Raiz em um ponto onde o gráfico cruza o eixo x; (*b*) raiz na intersecção das funções componentes.

x	y_1	y_2
0,0	0,0	1,000
0,2	0,2	0,819
0,4	0,4	0,670
0,6	0,6	0,549
0,8	0,8	0,449
1,0	1,0	0,368

Esses pontos estão marcados na Figura 6.2*b*. A intersecção das duas curvas indica uma raiz em aproximadamente $x = 0{,}57$, que corresponde ao ponto onde a curva única na Figura 6.2*a* cruza o eixo x.

O método das duas curvas pode agora ser usado para ilustrar a convergência e a divergência da iteração de ponto fixo. Primeiro, a Equação (6.1) pode ser reescrita como um par de equações $y_1 = x$ e $y_2 = g(x)$. Essas duas equações podem então ser traçadas separadamente. Como era o caso com as Equações (6.3) e (6.4), as raízes de $f(x) = 0$ correspondem aos valores da abscissa na intersecção das duas curvas. A função $y_1 = x$ e quatro formas diferentes para $y_2 = g(x)$ são mostradas na Figura 6.3.

No primeiro caso (Figura 6.3*a*), a aproximação inicial x_0 é usada para determinar o ponto correspondente na curva y_2 $[x_0, g(x_0)]$. O ponto (x_1, x_1) é localizado movimentando-se horizontalmente para a esquerda até a curva y_1. Esses movimentos são equivalentes à primeira iteração no método do ponto fixo:

$$x_1 = g(x_0)$$

Portanto, em ambas as equações e no gráfico, um valor inicial x_0 é usado para se obter uma estimativa de x_1. A próxima iteração consiste em se movimentar para $[x_1, g(x_1)]$ e então para (x_2, x_2). Essa iteração é equivalente à equação

$$x_2 = g(x_1)$$

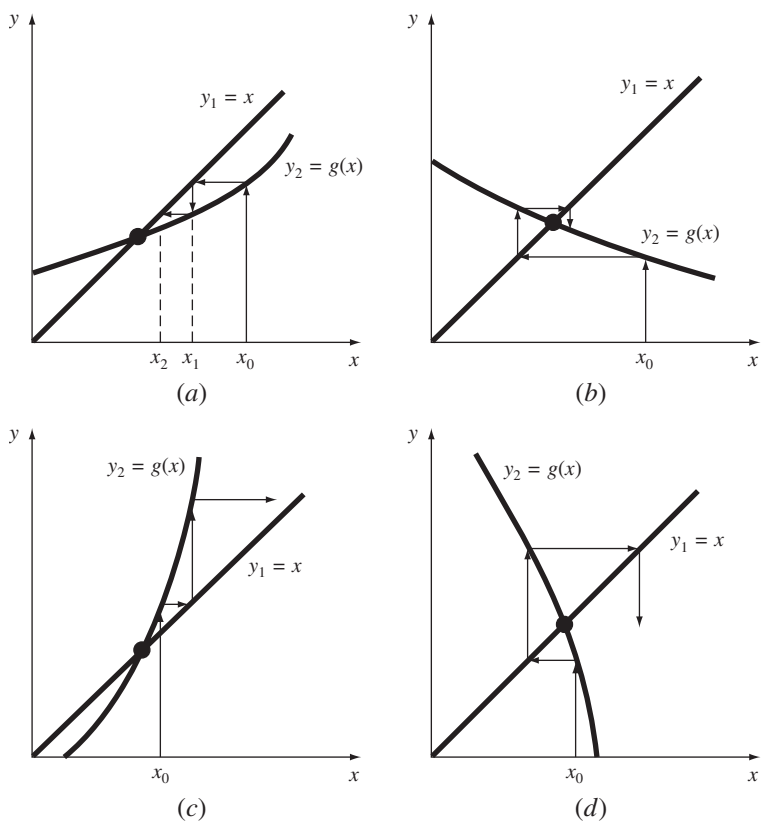


FIGURA 6.3
Descrição gráfica da convergência (*a*) e (*b*) e divergência (*c*) e (*d*) da iteração de ponto fixo simples. Os gráficos (*a*) e (*c*) são chamados padrões monótonos, enquanto (*b*) e (*d*) são chamados padrões oscilantes ou espirais. Observe que a convergência ocorre quando $|g'(x)| < 1$.

Quadro 6.1 Convergência da Iteração de Ponto Fixo

Do estudo da Figura 6.3, deveria estar claro que a iteração de ponto fixo converge se, na região de interesse, $|g'(x)| < 1$. Em outras palavras, a convergência ocorre se o módulo da inclinação de $g(x)$ for menor do que a inclinação da reta $f(x) = x$. Essa observação pode ser demonstrada teoricamente. Lembre-se de que a equação iterativa é

$$x_{i+1} = g(x_i)$$

Suponha que a solução verdadeira seja

$$x_r = g(x_r)$$

Subtraindo essas equações obtém-se:

$$x_r - x_{i+1} = g(x_r) - g(x_i) \tag{Q6.1.1}$$

O *teorema do valor médio para derivadas* (lembre-se da Seção 4.1.1) afirma que, se a função $g(x)$ e sua primeira derivada forem contínuas sobre um intervalo $a \leq x \leq b$, então existe pelo menos um valor de $x = \xi$ dentro do intervalo tal que

$$g'(\xi) = \frac{g(b) - g(a)}{b - a} \tag{Q6.1.2}$$

O lado direito dessa equação é a inclinação da reta ligando $g(a)$ e $g(b)$. Logo, o teorema do valor médio afirma que existe pelo menos um ponto entre a e b que tem uma inclinação, denotada por $g'(\xi)$, que é paralela à reta ligando $g(a)$ e $g(b)$ (lembre-se da Figura 4.3).

Agora, se fizermos $a = x_i$ e $b = x_r$, o lado direito da Equação (Q6.1.1) pode ser expresso como

$$g(x_r) - g(x_i) = (x_r - x_i)g'(\xi)$$

onde ξ está em algum lugar entre x_i e x_r. Esse resultado então pode ser substituído na Equação (Q6.1.1) para fornecer

$$x_r - x_{i+1} = (x_r - x_i)g'(\xi) \tag{Q6.1.3}$$

Se o erro verdadeiro para a i-ésima iteração for definido por

$$E_{t,i} = x_r - x_i$$

então a Equação (B6.1.3) torna-se

$$E_{t,i+1} = g'(\xi)E_{t,i}$$

Conseqüentemente, se $|g'(x)| < 1$, os erros diminuem a cada iteração. Para $|g'(x)| > 1$, o erro cresce. Observe também que, se a derivada for positiva, os erros serão positivos e, portanto, a solução iterativa será monótona (Figura 6.3*a* e *c*). Se a derivada for negativa, os erros oscilarão (Figura 6.3*b* e *d*).

Um desdobramento da análise é que ela também demonstra que, quando o método converge, o erro é aproximadamente proporcional e inferior ao erro no passo anterior. Por essa razão, a iteração de ponto fixo simples é dita *linearmente convergente*.

A solução na Figura 6.3*a* é *convergente* porque as estimativas de x se aproximam da raiz a cada iteração. O mesmo é verdade para a Figura 6.3*b*. Entretanto, esse não é o caso para as Figuras 6.3*c* e *d*, nas quais as iterações divergem da raiz. Observe que a convergência parece ocorrer apenas quando o valor absoluto da inclinação de $y_2 = g(x)$ é menor do que a inclinação de $y_1 = x$, isto é, quando $|g'(x)| < 1$. O Quadro 6.1 fornece uma dedução teórica desse resultado.

6.1.2 Algoritmo para a Iteração de Ponto Fixo

O algoritmo computacional para a iteração de ponto fixo é extremamente simples, e consiste em um laço para calcular iterativamente novas estimativas até que o critério de parada seja satisfeito. A Figura 6.4 apresenta um pseudocódigo para o algoritmo. Outros

FIGURA 6.4
Pseudocódigo para a iteração de ponto fixo. Observe que outros métodos abertos podem ser organizados nessa forma geral.

```
FUNCTION Fixpt(x0, es, imax, iter, ea)
  xr = x0
  iter = 0
  DO
    xrold = xr
    xr = g(xrold)
    iter = iter + 1
    IF xr ≠ 0 THEN
       ea = | (xr − xrold) / xr | · 100
    END IF
    IF ea < es OR iter ≥ imax EXIT
  END DO
  Fixpt = xr
END Fixpt
```

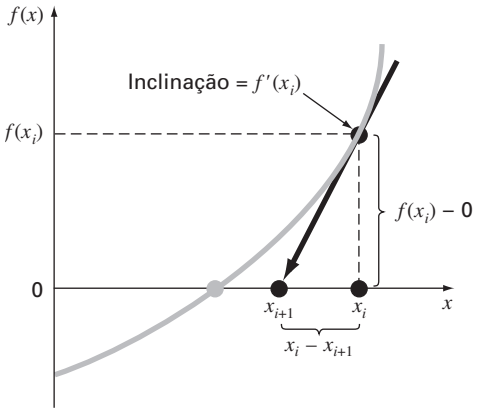


FIGURA 6.5
Descrição gráfica do método de Newton-Raphson. A tangente à função em x_i [isto é, $f'(x_i)$] é prolongada até o eixo x para fornecer uma estimativa da raiz em x_{i+1}.

métodos podem ser programados de forma parecida, sendo a principal mudança a troca da fórmula iterativa que é usada para calcular a nova estimativa da raiz.

6.2 O MÉTODO DE NEWTON-RAPHSON

Talvez a fórmula mais amplamente usada para localizar uma raiz seja a equação de Newton-Raphson (Figura 6.5). Se a aproximação inicial da raiz for x_i, pode-se estender uma reta tangente a partir do ponto $[x_i, f(x_i)]$. O ponto onde essa tangente cruza o eixo x usualmente representa uma estimativa melhorada da raiz.

O método de Newton-Raphson pode ser deduzido com base em sua interpretação geométrica (um método alternativo baseado na série de Taylor é descrito no Quadro 6.2). Como na Figura 6.5, a primeira derivada em x é equivalente à inclinação:

$$f'(x_i) = \frac{f(x_i) - 0}{x_i - x_{i+1}} \tag{6.5}$$

que pode ser reorganizada para fornecer

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \tag{6.6}$$

que é chamada de *fórmula de Newton-Raphson*.

EXEMPLO 6.3 Método de Newton-Raphson

Enunciado do Problema. Use o método de Newton-Raphson para fazer uma estimativa da raiz de $f(x) = e^{-x} - x$, utilizando uma aproximação inicial $x_0 = 0$.

Solução. A primeira derivada da função pode ser calculada como

$$f'(x) = -e^{-x} - 1$$

que pode ser substituída junto com a função original na Equação (6.6) para dar

$$x_{i+1} = x_i - \frac{e^{-x_i} - x_i}{-e^{-x_i} - 1}$$

Começando com uma aproximação inicial de $x_0 = 0$, essa equação iterativa pode ser aplicada para calcular

i	x_i	ε_t (%)
0	0	100
1	0,500000000	11,8
2	0,566311003	0,147
3	0,567143165	0,0000220
4	0,567143290	$< 10^{-8}$

Assim, a aproximação converge rapidamente para a raiz verdadeira. Observe que o erro relativo porcentual verdadeiro em cada iteração diminui muito mais rapidamente do que no método da iteração de ponto fixo simples (compare com o Exemplo 6.1).

6.2.1 Critério de Parada e Estimativas de Erro

De forma igual aos outros métodos de localização de raízes, a Equação (3.5) pode ser usada como critério de parada. Além disso, entretanto, a dedução pela série de Taylor do método (Quadro 6.2) fornece uma visão teórica da taxa de convergência, expressa por $E_{i+1} = O(E_i^2)$. Assim, o erro deveria ser aproximadamente proporcional ao quadrado do erro anterior. Em outras palavras, o número de algarismos significativos de acurácia aproximadamente dobra a cada iteração. Esse comportamento será examinado no próximo exemplo.

Quadro 6.2 Dedução e Análise de Erro do Método de Newton-Raphson

Além da dedução geométrica [Equações (6.5) e (6.6)], o método de Newton-Raphson pode ser deduzido a partir da expansão em série de Taylor. Essa dedução alternativa é útil, pois também dá informação sobre a taxa de convergência do método.

Lembre-se do Capítulo 4 que a expansão em série de Taylor pode ser representada por

$$f(x_{i+1}) = f(x_i) + f'(x_i)(x_{i+1} - x_i) + \frac{f''(\xi)}{2!}(x_{i+1} - x_i)^2 \qquad \text{(Q6.2.1)}$$

onde ξ está em algum ponto do intervalo de x_i a x_{i+1}. Uma versão aproximada é obtida truncando-se a série depois do termo da primeira derivada:

$$f(x_{i+1}) \cong f(x_i) + f'(x_i)(x_{i+1} - x_i)$$

Na intersecção com o eixo x, $f(x_{i+1})$ deveria ser igual a zero, ou

$$0 = f(x_i) + f'(x_i)(x_{i+1} - x_i) \qquad \text{(Q6.2.2)}$$

que pode ser reescrita como

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

que é idêntica à Equação (6.6). Portanto, deduzimos a fórmula de Newton-Raphson usando a série de Taylor.

Além da dedução, a série de Taylor também pode ser usada para fazer uma estimativa do erro da fórmula, o que se consegue percebendo que, se a série de Taylor completa fosse usada, seria obtido um resultado exato. Nessa situação, $x_{i+1} = x_r$, na qual x é o valor verdadeiro da raiz. Substituindo esse valor junto com $f(x_r) = 0$ na Equação (Q6.2.1) obtém-se

$$0 = f(x_i) + f'(x_i)(x_r - x_i) + \frac{f''(\xi)}{2!}(x_r - x_i)^2 \qquad \text{(Q6.2.3)}$$

A Equação (Q6.2.2) pode ser subtraída da Equação (Q6.2.3) fornecendo

$$0 = f'(x_i)(x_r - x_{i+1}) + \frac{f''(\xi)}{2!}(x_r - x_i)^2 \qquad \text{(Q6.2.4)}$$

Agora, perceba que o erro é igual à discrepância entre x_{i+1} e o valor verdadeiro x_r, como em

$$E_{t,i+1} = x_r - x_{i+1}$$

e a Equação (Q6.2.4) pode ser expressa por

$$0 = f'(x_i)E_{t,i+1} + \frac{f''(\xi)}{2!}E_{t,i}^2 \qquad \text{(Q6.2.5)}$$

Se supusermos a convergência, ambos x_i e ξ deveriam eventualmente ser aproximados pela raiz x_r e a Equação (Q6.2.5) pode ser reorganizada para fornecer

$$E_{t,i+1} = \frac{-f''(x_r)}{2f'(x_r)}E_{t,i}^2 \qquad \text{(Q6.2.6)}$$

De acordo com a Equação (Q6.2.6), o erro é aproximadamente proporcional ao quadrado do erro anterior. Isso significa que o número de casas decimais corretas aproximadamente dobra a cada iteração. Tal comportamento é chamado de *convergência quadrática*. O Exemplo 6.4 evidencia essa propriedade.

EXEMPLO 6.4 Análise de Erro do Método de Newton-Raphson

Enunciado do Problema. Como deduzido no Quadro 6.2, o método de Newton-Raphson é quadraticamente convergente. Ou seja, o erro é aproximadamente proporcional ao quadrado do erro anterior, como em

$$E_{t,i+1} \cong \frac{-f''(x_r)}{2f'(x_r)} E_{t,i}^2 \tag{E6.4.1}$$

Examine essa fórmula e veja se ela se aplica aos resultados do Exemplo 6.3.

Solução. A primeira derivada de $f(x) = e^{-x} - x$ é

$$f'(x) = -e^{-x} - 1$$

o que pode ser calculado em $x_r = 0{,}56714329$ como $f'(0{,}56714329) = -1{,}56714329$. A segunda derivada é

$$f''(x) = e^{-x}$$

o que pode ser calculado como $f''(0{,}56714329) = 0{,}56714329$. Esses resultados podem ser substituídos na equação (E6.4.1) para fornecer

$$E_{t,i+1} \cong -\frac{0{,}56714329}{2(-1{,}56714329)} E_{t,i}^2 = 0{,}18095 E_{t,i}^2$$

Do Exemplo 6.3, o erro inicial era $E_{t,\,0} = 0{,}56714329$, o que pode ser substituído na equação do erro para prever

$$E_{t,1} \cong 0{,}18095(0{,}56714329)^2 = 0{,}0582$$

o que está próximo do erro verdadeiro de 0,6714329. Para a próxima iteração,

$$E_{t,2} \cong 0{,}18095(0{,}06714329)^2 = 0{,}0008158$$

o que também se compara favoravelmente com o erro verdadeiro de 0,0008323. Para a terceira iteração,

$$E_{t,3} \cong 0{,}18095(0{,}0008323)^2 = 0{,}000000125$$

que é o erro obtido no Exemplo 6.3. A estimativa de erro melhora dessa forma porque, conforme nos aproximamos da raiz, x e ξ são mais bem aproximados por x_r [lembre-se de nossa hipótese ao passarmos da Equação (Q6.2.5) para a Equação (Q6.2.6) no Quadro 6.2]. Finalmente,

$$E_{t,4} \cong 0{,}18095(0{,}000000125)^2 = 2{,}83 \times 10^{-15}$$

Logo, esse exemplo ilustra que o erro no método de Newton-Raphson para esse caso é, de fato, aproximadamente proporcional (por um fator de 0,18095) ao quadrado do erro da iteração anterior.

6.2.2 Armadilhas do Método de Newton-Raphson

Embora o método de Newton-Raphson seja em geral muito eficiente, há situações nas quais ele tem um desempenho insatisfatório. Um caso especial — raízes múltiplas — será examinado adiante neste capítulo. Entretanto, mesmo quando tratando com raízes simples, podem aparecer dificuldades, como no próximo exemplo.

EXEMPLO 6.5 Exemplo de uma Função que Converge Lentamente com Newton-Raphson

Enunciado do Problema. Determine a raiz positiva de $f(x) = x^{10} - 1$ usando o método de Newton-Raphson e uma aproximação inicial $x = 0{,}5$.

Solução. A fórmula de Newton-Raphson para esse caso é

$$x_{i+1} = x_i - \frac{x_i^{10} - 1}{10x_i^9}$$

a qual pode ser usada para calcular

Iteração	x
0	0,5
1	51,65
2	46,485
3	41,8365
4	37,65285
5	33,887565
.	
.	
.	
∞	1,0000000

Logo, depois da primeira previsão fraca, a técnica está convergindo para a raiz verdadeira de 1, mas a uma taxa muito lenta.

Além da convergência lenta em razão da natureza da função, outras dificuldades podem aparecer, como ilustrado na Figura 6.6. Por exemplo, a Figura 6.6*a* descreve o caso no qual ocorre um ponto de inflexão [isto é, $f''(x) = 0$] na vizinhança de uma raiz. Observe que as iterações começando em x_0 se afastam progressivamente da raiz. A Figura 6.6*b* ilustra a tendência da técnica de Newton-Raphson em oscilar em torno de uma posição de máximo ou de mínimo. Tais oscilações podem persistir, ou, como na Figura 6.6*b*, é atingida uma inclinação próxima a zero, em que a solução é mandada para longe da área de interesse. A Figura 6.6*c* mostra como uma aproximação inicial que esteja próxima de uma raiz pode pular para uma posição longe por diversas raízes. Essa tendência de se afastar da área de interesse se dá porque foram encontradas inclinações próximas de zero. Obviamente, uma inclinação nula [$f'(x) = 0$] é, na verdade, um desastre porque provoca divisão por zero na fórmula de Newton-Raphson [Equação (6.6)]. Graficamente (ver Figura 6.6*d*), isso significa que a solução dispara horizontalmente e nunca atinge o eixo x.

Logo, não existe nenhum critério de convergência geral para o Newton-Raphson. Sua convergência depende da natureza da função e da precisão da aproximação inicial. O único remédio é ter uma aproximação inicial que esteja "suficientemente" próxima da raiz. E para algumas funções, nenhuma aproximação funcionará! Boas aproximações dependem em geral do conhecimento das condições do problema físico ou dos artifícios tais como os gráficos que fornecem informação sobre o comportamento da solução. A falta de um critério de convergência geral também sugere que um bom software de computador deveria ser planejado para reconhecer convergência lenta ou divergência. A próxima seção trata de algumas dessas questões.

6.2.3 Algoritmo para o Método de Newton-Raphson

Um algoritmo para o método de Newton-Raphson é prontamente obtido substituindo a Equação (6.6) na fórmula de predição [Equação (6.2)] na Figura 6.4. Observe, contudo, que o programa também deve ser modificado para calcular a primeira derivada, o que se pode conseguir simplesmente pela inclusão de uma função definida pelo usuário.

Além disso, com base na discussão anterior dos problemas em potencial do método de Newton-Raphson, o programa seria melhorado pela incorporação de diversas características adicionais:

1. Uma rotina de gráficos deveria ser incluída no programa.
2. No final dos cálculos, a estimativa final da raiz deveria sempre ser substituída na função original para calcular se o resultado está próximo a zero. Essa verificação nos resguarda parcialmente daqueles casos nos quais a convergência oscilante ou lenta pode levar a valores pequenos de ε_a enquanto a solução ainda está longe da raiz.

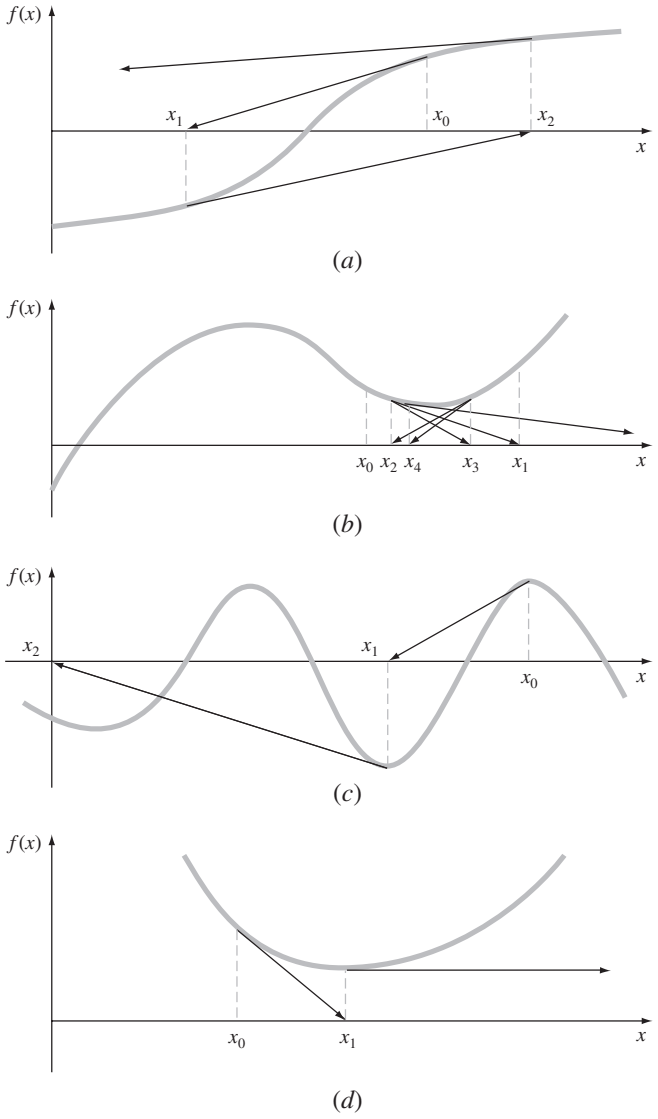


FIGURA 6.6
Quatro casos nos quais o método de Newton-Raphson exibe convergência insatisfatória.

3. O programa deveria sempre incluir um limite superior no número de iterações para se prevenir de soluções oscilantes, lentamente convergentes ou divergentes que poderiam persistir interminavelmente.
4. O programa deveria alertar o usuário e levar em conta a possibilidade de que $f'(x)$ possa ser igual a zero a qualquer instante durante os cálculos.

6.3 O MÉTODO DA SECANTE

Um problema em potencial na implementação do método de Newton-Raphson é o cálculo da derivada. Embora isso não seja inconveniente para polinômios e muitas outras funções, há certas funções cujas derivadas podem ser extremamente difíceis ou inconvenientes para calcular. Nesses casos, a derivada pode ser aproximada por uma diferença dividida regressiva, como em (Figura 6.7)

$$f'(x_i) \cong \frac{f(x_{i-1}) - f(x_i)}{x_{i-1} - x_i}$$

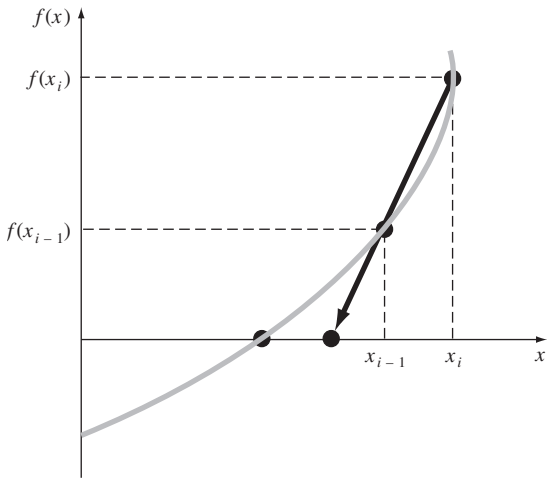


FIGURA 6.7
Descrição gráfica do método da secante. Essa técnica é parecida com a técnica de Newton-Raphson (Figura 6.5) no sentido que uma estimativa da raiz é prevista extrapolando-se uma tangente da função até o eixo *x*. Entretanto, o método da secante usa uma diferença em vez de uma derivada para fazer uma estimativa da inclinação.

Essa aproximação pode ser substituída na Equação (6.6) para fornecer a seguinte equação iterativa:

$$x_{i+1} = x_i - \frac{f(x_i)(x_{i-1} - x_i)}{f(x_{i-1}) - f(x_i)} \tag{6.7}$$

A Equação (6.7) é a fórmula do *método da secante*. Observe que a abordagem exige duas estimativas iniciais de *x*. Mas, como não é exigido que $f(x)$ mude de sinal entre as estimativas, ele não é classificado como um método intervalar.

EXEMPLO 6.6 O Método da Secante

Enunciado do Problema. Use o método da secante para fazer uma estimativa da raiz de $f(x) = e^{-x} - x$. Comece com as estimativas iniciais de $x_{-1} = 0$ e $x_0 = 1{,}0$.

Solução. Lembre-se de que a raiz verdadeira é 0,56714329....

Primeira iteração:

$$x_{-1} = 0 \qquad f(x_{-1}) = 1{,}00000$$

$$x_0 = 1 \qquad f(x_0) = -0{,}63212$$

$$x_1 = 1 - \frac{-0{,}63212(0 - 1)}{1 - (-0{,}63212)} = 0{,}61270 \qquad \varepsilon_t = 8{,}0\%$$

Segunda iteração:

$$x_0 = 1 \qquad f(x_0) = -0{,}63212$$

$$x_1 = 0{,}61270 \qquad f(x_1) = -0{,}07081$$

(Observe que ambas as estimativas estão agora do mesmo lado da raiz.)

$$x_2 = 0{,}61270 - \frac{-0{,}07081(1 - 0{,}61270)}{-0{,}63212 - (-0{,}07081)} = 0{,}56384 \qquad \varepsilon_t = 0{,}58\%$$

Terceira iteração:

$$x_1 = 0{,}61270 \qquad f(x_1) = -0{,}07081$$
$$x_2 = 0{,}56384 \qquad f(x_2) = 0{,}00518$$
$$x_3 = 0{,}56384 - \frac{0{,}00518(0{,}61270 - 0{,}56384)}{-0{,}07081 - (-0{,}00518)} = 0{,}56717 \qquad \varepsilon_t = 0{,}0048\%$$

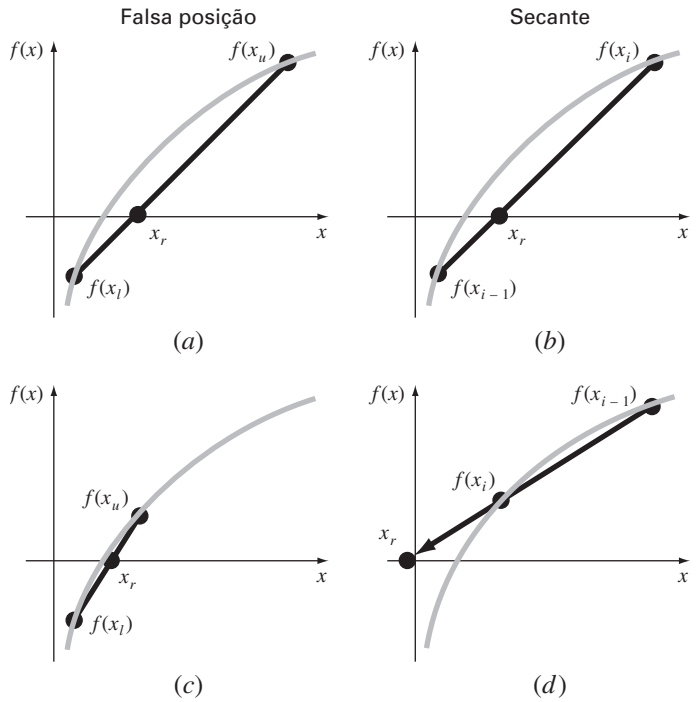
6.3.1 A Diferença entre os Métodos da Secante e da Falsa Posição

Observe a semelhança entre o método da secante e o método da falsa posição. Por exemplo, as Equações (6.7) e (5.7) são idênticas em uma comparação termo a termo. Ambos usam duas estimativas iniciais para calcular uma aproximação da inclinação da função que é utilizada para projetar para o eixo x para uma nova estimativa da raiz. Entretanto, uma diferença crítica entre os métodos é como um dos valores iniciais é substituído pela nova estimativa. Lembre-se de que no método da falsa posição a última estimativa da raiz substitui qualquer um dos valores iniciais que forneça o valor da função com o mesmo sinal que $f(x_r)$. Conseqüentemente, as duas estimativas sempre delimitam a raiz. Portanto, para todos os propósitos práticos, o método sempre converge porque a raiz é mantida dentro do intervalo. Em contraste, o método da secante substitui os valores em seqüência estrita, com o novo valor x_{i+1} substituindo x_i e x_i substituindo x_{i-1}. Para certos casos, isso pode levar à divergência.

EXEMPLO 6.7 Comparação da Convergência das Técnicas da Secante e da Falsa Posição

Enunciado do Problema. Use os métodos da falsa posição e da secante para fazer uma estimativa da raiz de $f(x) = \ln x$. Comece os cálculos com os valores $x_l = x_{i-1} = 0{,}5$ e $x_u = x_i = 5{,}0$.

FIGURA 6.8
Comparação dos métodos da falsa posição e da secante. As primeiras iterações (*a*) e (*b*) para ambas as técnicas são iguais. Entretanto, na segunda iteração (*c*) e (*d*), os pontos usados são diferentes. Como conseqüência, o método da secante pode divergir, como indicado em (*d*).



Solução. No método da falsa posição, o uso da Equação (5.7) e o critério de delimitação para substituir as estimativas resultam nas seguintes iterações:

Iteração	x_l	x_u	x_r
1	0,5	5,0	1,8546
2	0,5	1,8546	1,2163
3	0,5	1,2163	1,0585

Como pode ser visto (Figuras 6.8*a* e *c*), as estimativas estão convergindo para a raiz verdadeira, que é igual a 1.

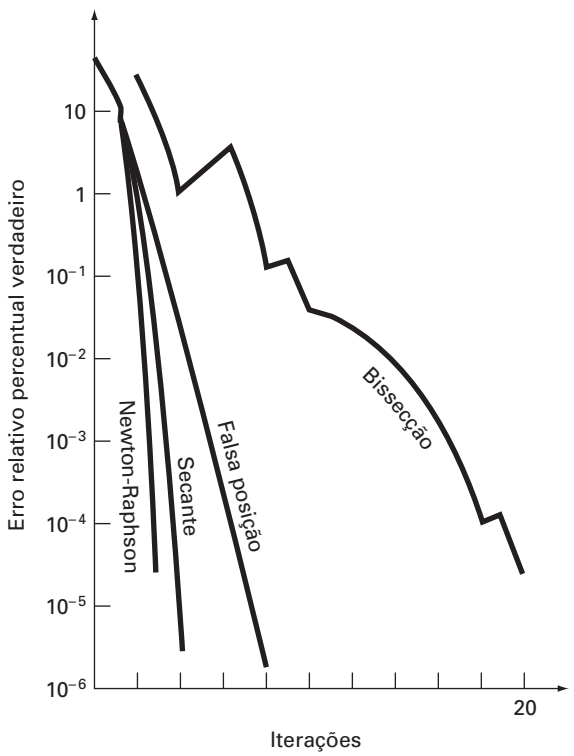
Para o método da secante, usar a Equação (6.7) e o critério seqüencial para substituir as estimativas resulta em

Iteração	x_{i-1}	x_i	x_{i+1}
1	0,5	5,0	1,8546
2	5,0	1,8546	−0,10438

Como na Figura 6.8*d*, a abordagem é divergente.

Embora o método da secante possa ser divergente, quando ele converge, usualmente o faz a uma taxa mais rápida do que o método da falsa posição. Por exemplo, a Figura 6.9 ilustra a superioridade do método da secante com relação a isso. A inferioridade do

FIGURA 6.9
Comparação dos erros relativos percentuais verdadeiros ε_t para os métodos para determinar raízes de $f(x) = e^{-x} - x$.



método da falsa posição resulta de uma extremidade permanecer fixa para continuar a delimitar a raiz. Essa propriedade, que é vantajosa do ponto de vista de impedir a divergência, é uma desvantagem com relação à taxa de convergência; torna a estimativa por diferença finita uma aproximação menos acurada da derivada.

6.3.2 Algoritmo para o Método da Secante

Como com os outros métodos abertos, um algoritmo para o método da secante é obtido modificando-se a Figura 6.4 de modo que as duas aproximações iniciais sejam entradas e usando-se a Equação (6.7) para calcular a raiz. Além disso, as opções sugeridas na Seção 6.2.3 para o método de Newton-Raphson também podem ser aplicadas com bastante vantagem ao programa da secante.

6.3.3 Método da Secante Modificado

Em vez de usar dois valores arbitrários para fazer uma estimativa da derivada, uma abordagem alternativa envolve uma pequena perturbação da variável independente para fazer uma estimativa de $f'(x)$,

$$f'(x_i) \cong \frac{f(x_i + \delta x_i) - f(x_i)}{\delta x_i}$$

onde δ é uma pequena fração de perturbação. Essa aproximação pode ser substituída na Equação (6.6) para fornecer a seguinte equação iterativa:

$$x_{i+1} = x_i - \frac{\delta x_i f(x_i)}{f(x_i + \delta x_i) - f(x_i)} \tag{6.8}$$

EXEMPLO 6.8 Método da Secante Modificado

Enunciado do Problema. Use o método da secante modificado para fazer uma estimativa da raiz de $f(x) = e^{-x} - x$. Use um valor de 0,01 para δ e comece com $x_0 = 1{,}0$. Lembre-se de que a raiz verdadeira é 0,56714329. . . .

Solução.
Primeira iteração:

$$x_0 = 1 \qquad f(x_0) = -0{,}63212$$
$$x_0 + \delta x_0 = 1{,}01 \qquad f(x_0 + \delta x_0) = -0{,}64578$$
$$x_1 = 1 - \frac{0{,}01(-0{,}63212)}{-0{,}64578 - (-0{,}63212)} = 0{,}537263 \qquad |\varepsilon_t| = 5{,}3\%$$

Segunda iteração:

$$x_0 = 0{,}537263 \qquad f(x_0) = 0{,}047083$$
$$x_0 + \delta x_0 = 0{,}542635 \qquad f(x_0 + \delta x_0) = 0{,}038579$$
$$x_1 = 0{,}537263 - \frac{0{,}005373(0{,}047083)}{0{,}038579 - 0{,}047083} = 0{,}56701 \qquad |\varepsilon_t| = 0{,}0236\%$$

Terceira iteração:

$$x_0 = 0{,}56701 \qquad f(x_0) = 0{,}000209$$
$$x_0 + \delta x_0 = 0{,}572680 \qquad f(x_0 + \delta x_0) = -0{,}00867$$
$$x_1 = 0{,}56701 - \frac{0{,}00567(0{,}000209)}{-0{,}00867 - 0{,}000209} = 0{,}567143 \qquad |\varepsilon_t| = 2{,}365 \times 10^{-5}\%$$

A escolha de um valor adequado para δ não é automática. Se δ for muito pequeno, o método pode ser sobrecarregado de erros de arredondamento causados pelo cancelamento

da subtração no denominador da Equação (6.8). Se for grande demais, a técnica pode tornar-se ineficiente e mesmo divergente. Entretanto, se escolhido corretamente, ele fornece uma boa alternativa para os casos nos quais o cálculo da derivada é difícil e encontrar duas aproximações iniciais é inconveniente.

6.4 RAÍZES MÚLTIPLAS

Uma *raiz múltipla* corresponde a um ponto onde a função é tangente ao eixo x. Por exemplo, uma raiz dupla aparece em

$$f(x) = (x - 3)(x - 1)(x - 1) \tag{6.9}$$

ou, multiplicando os termos, $f(x) = x^3 - 5x^2 + 7x - 3$. A equação tem uma *raiz dupla* porque um valor de x torna dois termos na Equação (6.9) iguais a zero. Graficamente, isso corresponde à curva tocar o eixo x tangencialmente na raiz dupla. Examine a Figura 6.10*a* em $x = 1$. Observe que, na raiz, a função toca o eixo, mas não o cruza.

Uma *raiz tripla* corresponde ao caso no qual um valor de x anula três termos em uma equação, como em

$$f(x) = (x - 3)(x - 1)(x - 1)(x - 1)$$

ou, multiplicando os termos, $f(x) = x^4 - 6x^3 + 12x^2 - 10x + 3$. Observe que a descrição gráfica (Figura 6.10*b*) indica novamente que a função é tangente ao eixo na raiz, mas que, nesse caso, o eixo é cruzado. Em geral, raízes com multiplicidade ímpar cruzam o eixo, enquanto as pares não. Por exemplo, a raiz quádrupla na Figura 6.10*c* não cruza o eixo.

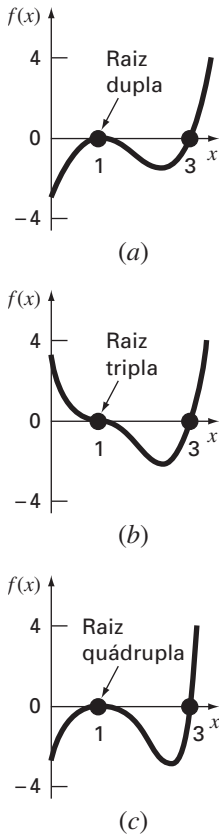


FIGURA 6.10
Exemplos de raízes múltiplas que são tangentes ao eixo x. Observe que a função não cruza o eixo em cada lado das raízes de multiplicidade par (*a*) e (*c*), enquanto cruza o eixo nos casos ímpares (*b*).

As raízes múltiplas causam algumas dificuldades para muitos dos métodos numéricos descritos na Parte Dois:

1. O fato de a função não mudar de sinal em raízes de multiplicidade par impede o uso dos métodos intervalares confiáveis, que foram discutidos no Capítulo 5. Portanto, dos métodos descritos neste livro, você está limitado aos métodos abertos, que podem divergir.
2. Um outro problema possível está relacionado ao fato de que não só $f(x)$ mas também $f'(x)$ vai a zero na raiz. Isso introduz problemas tanto no método de Newton-Raphson quanto no da secante, já que ambos contêm a derivada (ou sua estimativa) no denominador de suas respectivas fórmulas, o que pode resultar na divisão por zero quando a solução converge para muito próximo da raiz. Uma forma simples de contornar esse problema é baseada no fato que pode ser demonstrado teoricamente (Ralston e Rabinowitz, 1978) que $f(x)$ sempre atingirá zero antes de $f'(x)$. Portanto, se uma verificação do zero for incluída no programa de computador, os cálculos podem ser parados antes de $f'(x)$ atingir zero.
3. É possível demonstrar que os métodos de Newton-Raphson e da secante são linearmente — em vez de quadraticamente — convergentes para raízes múltiplas (Ralston e Rabinowitz, 1978). Foram propostas modificações para diminuir esse problema. Ralston e Rabinowitz (1978) indicaram que uma pequena mudança na formulação restaura sua convergência quadrática, como em

$$x_{i+1} = x_i - m\frac{f(x_i)}{f'(x_i)} \tag{6.9a}$$

onde m é a multiplicidade da raiz (isto é, $m = 2$ para uma raiz dupla, $m = 3$ para uma raiz tripla etc.). É claro que isso pode ser uma alternativa insatisfatória, porque depende do conhecimento prévio da multiplicidade da raiz.

Uma outra alternativa, também sugerida por Ralston e Rabinowitz (1978), é definir uma nova função $u(x)$, isto é, o quociente da função por sua derivada, como em

$$u(x) = \frac{f(x)}{f'(x)} \tag{6.10}$$

É possível mostrar que essa função tem raízes em todas as mesmas posições que a função original. Portanto, a Equação (6.10) pode ser substituída na Equação (6.6) para deduzir uma forma alternativa para o método de Newton-Raphson:

$$x_{i+1} = x_i - \frac{u(x_i)}{u'(x_i)} \tag{6.11}$$

A equação (6.10) pode ser derivada para fornecer

$$u'(x) = \frac{f'(x)f'(x) - f(x)f''(x)}{[f'(x)]^2} \tag{6.12}$$

As Equações (6.10) e (6.12) podem ser substituídas na Equação (6.11) e o resultado pode ser simplificado para fornecer

$$x_{i+1} = x_i - \frac{f(x_i)f'(x_i)}{[f'(x_i)]^2 - f(x_i)f''(x_i)} \tag{6.13}$$

EXEMPLO 6.9 Método de Newton-Raphson Modificado para Raízes Múltiplas

Enunciado do Problema. Use os métodos padrão e modificado de Newton-Raphson para calcular a raiz múltipla da Equação (6.9), com aproximação inicial $x_0 = 0$.

Solução. A primeira derivada da Equação (6.9) é $f'(x) = 3x^2 - 10x + 7$, e, portanto, o método padrão de Newton-Raphson para esse problema é [Equação (6.6)]

$$x_{i+1} = x_i - \frac{x_i^3 - 5x_i^2 + 7x_i - 3}{3x_i^2 - 10x_i + 7}$$

que pode ser resolvido iterativamente por

i	x_i	ε_t (%)
0	0	100
1	0,4285714	57
2	0,6857143	31
3	0,8328654	17
4	0,9133290	8,7
5	0,9557833	4,4
6	0,9776551	2,2

Como previsto, o método converge linearmente para o valor verdadeiro de 1,0.

Para o método modificado, a segunda derivada é $f''(x) = 6x - 10$ e a relação iterativa é [Equação (6.13)]

$$x_{i+1} = x_i - \frac{\left(x_i^3 - 5x_i^2 + 7x_i - 3\right)\left(3x_i^2 - 10x_i + 7\right)}{\left(3x_i^2 - 10x_i + 7\right)^2 - \left(x_i^3 - 5x_i^2 + 7x_i - 3\right)(6x_i - 10)}$$

que pode ser resolvida por

i	x_i	ε_t (%)
0	0	100
1	1,105263	11
2	1,003082	0,31
3	1,000002	0,00024

Logo, a fórmula modificada é quadraticamente convergente. Pode-se também usar ambos os métodos para procurar a raiz simples em $x = 3$. Usando uma aproximação inicial de $x_0 = 4$, obtêm-se os seguintes resultados:

i	Padrão	ε_t (%)	Modificado	ε_t (%)
0	4	33	4	33
1	3,4	13	2,636364	12
2	3,1	3,3	2,820225	6,0
3	3,008696	0,29	2,961728	1,3
4	3,000075	0,0025	2,998479	0,051
5	3,000000	2×10^{-7}	2,999998	$7{,}7 \times 10^{-5}$

Portanto, ambos os métodos convergem rapidamente, com o método padrão sendo um pouco mais eficiente.

Esse exemplo ilustra os prós e contras envolvidos na escolha do método modificado de Newton-Raphson. Embora ele seja preferível para raízes múltiplas, é um pouco menos eficiente e exige mais esforço computacional do que o método padrão para raízes simples.

Deve ser observado que uma versão modificada do método da secante adequado para raízes múltiplas também pode ser deduzida substituindo-se a Equação (6.10) na Equação (6.7). A fórmula resultante é (Ralston e Rabinowitz, 1978)

$$x_{i+1} = x_i - \frac{u(x_i)(x_{i-1} - x_i)}{u(x_{i-1}) - u(x_i)}$$

6.5 SISTEMAS DE EQUAÇÕES NÃO-LINEARES

Até este ponto, concentramo-nos na determinação das raízes de uma única equação. Um problema relacionado é a localização das raízes de um conjunto de equações simultâneas,

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0 \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \\ \cdot \qquad & \cdot \\ \cdot \qquad & \cdot \\ \cdot \qquad & \cdot \\ f_n(x_1, x_2, \ldots, x_n) &= 0 \end{aligned} \tag{6.14}$$

A solução desse sistema consiste em um conjunto de valores de x que resultam simultaneamente em igualar todas as equações a zero.

Na Parte Três, serão apresentados métodos para o caso em que todas as equações simultâneas são lineares — isto é, elas podem ser expressas na forma geral

$$f(x) = a_1x_1 + a_2x_2 + \cdots + a_nx_n - b = 0 \tag{6.15}$$

na qual o b's e os a's são constantes. Equações algébricas e transcendentais que não se ajustam a essa forma são chamadas *equações não-lineares*. Por exemplo,

$$x^2 + xy = 10$$

e

$$y + 3xy^2 = 57$$

são duas equações não-lineares simultâneas com duas incógnitas, x e y. Elas podem ser expressas na forma da Equação (6.14) como

$$u(x, y) = x^2 + xy - 10 = 0 \tag{6.16a}$$

$$v(x, y) = y + 3xy^2 - 57 = 0 \tag{6.16b}$$

Portanto, a solução seriam os valores de x e y que fizessem as funções $u(x, y)$ e $v(x, y)$ iguais a zero. Em sua maioria, as abordagens para determinar tais soluções são extensões dos métodos abertos para a resolução de uma única equação. Nesta seção, vamos investigar dois deles: iteração de ponto fixo e Newton-Raphson.

6.5.1 Iteração de Ponto Fixo

A abordagem por iteração de ponto fixo (Seção 6.1) pode ser modificada para resolver duas equações não-lineares simultâneas. Essa abordagem será ilustrada no seguinte exemplo.

EXEMPLO 6.10 Iteração de Ponto Fixo para um Sistema Não-Linear

Enunciado do Problema. Use a iteração de ponto fixo para determinar as raízes da Equação (6.16). Observe que um par correto de raízes é $x = 2$ e $y = 3$. Inicie os cálculos com as aproximações $x = 1{,}5$ e $y = 3{,}5$.

Solução. A Equação (6.16*a*) pode ser reescrita como

$$x_{i+1} = \frac{10 - x_i^2}{y_i} \tag{E6.10.1}$$

e a Equação (6.16*b*) pode ser resolvida por

$$y_{i+1} = 57 - 3x_i y_i^2 \tag{E6.10.2}$$

Observe que os subscritos serão omitidos no restante do exemplo.

Com base nas aproximações iniciais, a Equação (E6.10.1) pode ser usada para determinar um novo valor de x:

$$x = \frac{10 - (1{,}5)^2}{3{,}5} = 2{,}21429$$

Esse resultado e o valor inicial $y = 3{,}5$ podem ser substituídos na Equação (E6.10.2) para determinar um novo valor de y:

$$y = 57 - 3(2{,}21429)(3{,}5)^2 = -24{,}37516$$

Logo, a abordagem parece estar divergindo, comportamento ainda mais pronunciado na segunda iteração:

$$x = \frac{10 - (2{,}21429)^2}{-24{,}37516} = -0{,}20910$$

$$y = 57 - 3(-0{,}20910)(-24{,}37516)^2 = 429{,}709$$

Obviamente, o método está deteriorando.

Agora, vamos repetir os cálculos, mas com as equações originais escritas em uma forma diferente. Por exemplo, uma formulação alternativa para a Equação (6.16*a*) é

$$x = \sqrt{10 - xy}$$

e para a Equação (6.16*b*) é

$$y = \sqrt{\frac{57 - y}{3x}}$$

Agora os resultados são mais satisfatórios:

$$x = \sqrt{10 - 1{,}5(3{,}5)} = 2{,}17945$$

$$y = \sqrt{\frac{57 - 3{,}5}{3(2{,}17945)}} = 2{,}86051$$

$$x = \sqrt{10 - 2{,}17945(2{,}86051)} = 1{,}94053$$

$$y = \sqrt{\frac{57 - 2{,}86051}{3(1{,}94053)}} = 3{,}04955$$

Logo, a abordagem está convergindo para os valores verdadeiros de $x = 2$ e $y = 3$.

O exemplo anterior ilustra a desvantagem mais grave da iteração de ponto fixo — isto é, a convergência freqüentemente depende da maneira como as equações são formuladas. Adicionalmente, mesmo nos casos em que a convergência é possível, a divergência pode ocorrer se as aproximações iniciais não estiverem suficientemente próximas da solução verdadeira. Usando um raciocínio análogo ao do Quadro 6.1, pode ser demonstrado que condições suficientes para a convergência no caso de duas equações são

$$\left|\frac{\partial u}{\partial x}\right| + \left|\frac{\partial u}{\partial y}\right| < 1$$

e

$$\left|\frac{\partial v}{\partial x}\right| + \left|\frac{\partial v}{\partial y}\right| < 1$$

Esses critérios são tão restritivos que a iteração de ponto fixo tem utilidade limitada na solução de sistemas não-lineares. Entretanto, como descreveremos mais adiante neste livro, ela pode ser muito útil para resolver sistemas lineares.

6.5.2 Newton-Raphson

Lembre-se de que o método de Newton-Raphson era baseado no uso da derivada (isto é, da inclinação) de uma função para fazer uma estimativa de sua intersecção com o eixo das variáveis independentes — ou seja, a raiz (Figura 6.5). Essa estimativa era baseada na expansão em série de Taylor (lembre-se do Quadro 6.2),

$$f(x_{i+1}) = f(x_i) + (x_{i+1} - x_i) f'(x_i) \tag{6.17}$$

onde x_i é a aproximação inicial da raiz e x_{i+1} é o ponto no qual a tangente intercepta o eixo x. Nessa intersecção, $f(x_{i+1})$ por definição é igual a zero e a Equação (6.17) pode ser reorganizada para fornecer

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \tag{6.18}$$

que é a forma para uma única equação do método de Newton-Raphson.

A forma para várias equações é deduzida de maneira idêntica. Entretanto, uma série de Taylor para diversas variáveis deve ser usada para representar o fato de que mais de uma variável independente contribui para a determinação da raiz. Para o caso de duas variáveis, a série de Taylor de primeira ordem pode ser escrita [lembre-se da Equação (4.26)] para cada equação não-linear como

$$u_{i+1} = u_i + (x_{i+1} - x_i)\frac{\partial u_i}{\partial x} + (y_{i+1} - y_i)\frac{\partial u_i}{\partial y} \tag{6.19a}$$

e

$$v_{i+1} = v_i + (x_{i+1} - x_i)\frac{\partial v_i}{\partial x} + (y_{i+1} - y_i)\frac{\partial v_i}{\partial y} \tag{6.19b}$$

Da mesma forma como para a versão para uma única equação, a estimativa da raiz corresponde aos valores de x e y nos quais u_{i+1} e v_{i+1} são iguais a zero. Para tal situação, a Equação (6.19) pode ser reorganizada para fornecer

$$\frac{\partial u_i}{\partial x}x_{i+1} + \frac{\partial u_i}{\partial y}y_{i+1} = -u_i + x_i\frac{\partial u_i}{\partial x} + y_i\frac{\partial u_i}{\partial y} \tag{6.20a}$$

$$\frac{\partial v_i}{\partial x}x_{i+1} + \frac{\partial v_i}{\partial y}y_{i+1} = -v_i + x_i\frac{\partial v_i}{\partial x} + y_i\frac{\partial v_i}{\partial y} \tag{6.20b}$$

Já que todos os valores subscritos com i são conhecidos (eles correspondem à última aproximação), as únicas incógnitas são x_{i+1} e y_{i+1}. Portanto, a Equação (6.20) é um conjunto de duas equações lineares com duas incógnitas [compare com a Equação (6.15)].

Conseqüentemente, manipulações algébricas (por exemplo, a regra de Cramer) podem ser usadas para determinar

$$x_{i+1} = x_i - \frac{u_i \dfrac{\partial v_i}{\partial y} - v_i \dfrac{\partial u_i}{\partial y}}{\dfrac{\partial u_i}{\partial x}\dfrac{\partial v_i}{\partial y} - \dfrac{\partial u_i}{\partial y}\dfrac{\partial v_i}{\partial x}} \tag{6.21a}$$

$$y_{i+1} = y_i - \frac{v_i \dfrac{\partial u_i}{\partial x} - u_i \dfrac{\partial v_i}{\partial x}}{\dfrac{\partial u_i}{\partial x}\dfrac{\partial v_i}{\partial y} - \dfrac{\partial u_i}{\partial y}\dfrac{\partial v_i}{\partial x}} \tag{6.21b}$$

O denominador de cada uma dessas equações é chamado formalmente de determinante da *matriz Jacobiana* do sistema.

A Equação (6.21) é a versão para duas equações do método de Newton-Raphson. Como no exemplo a seguir, ele pode ser usado iterativamente para aproximar as raízes de duas equações simultâneas.

EXEMPLO 6.11 Newton-Raphson para um Sistema Não-linear

Enunciado do Problema. Use o método de Newton-Raphson para equações múltiplas para determinar as raízes da Equação (6.16). Observe que o par correto de raízes é $x = 2$ e $y = 3$. Inicie os cálculos com a aproximação $x = 1{,}5$ e $y = 3{,}5$.

Solução. Determine primeiro as derivadas parciais e calcule-as nas aproximações iniciais de x e y:

$$\frac{\partial u_0}{\partial x} = 2x + y = 2(1{,}5) + 3{,}5 = 6{,}5 \qquad \frac{\partial u_0}{\partial y} = x = 1{,}5$$

$$\frac{\partial v_0}{\partial x} = 3y^2 = 3(3{,}5)^2 = 36{,}75 \qquad \frac{\partial v_0}{\partial y} = 1 + 6xy = 1 + 6(1{,}5)(3{,}5) = 32{,}5$$

Portanto, o determinante da matriz Jacobiana para a primeira iteração é

$$6{,}5(32{,}5) - 1{,}5(36{,}75) = 156{,}125$$

Os valores das funções podem ser calculados nas aproximações iniciais como sendo

$$u_0 = (1{,}5)^2 + 1{,}5(3{,}5) - 10 = -2{,}5$$

$$v_0 = 3{,}5 + 3(1{,}5)(3{,}5)^2 - 57 = 1{,}625$$

Esses valores podem ser substituídos na Equação (6.21) para fornecer

$$x = 1{,}5 - \frac{-2{,}5(32{,}5) - 1{,}625(1{,}5)}{156{,}125} = 2{,}03603$$

$$y = 3{,}5 - \frac{1{,}625(6{,}5) - (-2{,}5)(36{,}75)}{156{,}125} = 2{,}84388$$

Desse modo, os resultados estão convergindo para os valores verdadeiros $x = 2$ e $y = 3$. O cálculo pode ser repetido até que uma acurácia aceitável seja obtida.

Da mesma forma como na iteração de ponto fixo, a abordagem de Newton-Raphson irá com freqüência divergir se as aproximações iniciais não estiverem suficientemente próximas das raízes verdadeiras. Enquanto os métodos gráficos podiam ser usados para se encontrar boas aproximações para o caso de uma única equação, nenhum procedimento simples está disponível para a versão com diversas equações. Embora existam algumas abordagens avançadas para se obter primeiras estimativas aceitáveis, em geral as aproximações iniciais precisam ser obtidas com base na tentativa e erro e no conhecimento do sistema físico que está sendo modelado.

A abordagem de Newton-Raphson para duas equações pode ser generalizada para resolver n equações simultâneas. Já que a forma mais eficiente para fazer isso envolve a álgebra de matrizes e a resolução de equações lineares simultâneas, vamos adiar nossa discussão da abordagem geral até a Parte Três.

PROBLEMAS

6.1 Use a iteração de ponto fixo simples para localizar a raiz de $f(x) = 2\,\text{sen}(\sqrt{x}) - x$
Use a aproximação inicial $x_0 = 0{,}5$ e itere até $\varepsilon_a \leq 0{,}001\%$. Verifique que o processo é linearmente convergente, como descrito no Quadro 6.1.

6.2 Determine a maior raiz real de
$f(x) = 2x^3 - 11{,}7x^2 + 17{,}7x - 5$
(a) Graficamente.
(b) Pelo método da iteração de ponto fixo (três iterações, $x_0 = 3$). Observação: certifique-se de desenvolver uma solução que convirja para a raiz.
(c) Pelo método de Newton-Raphson (três iterações, $x_0 = 3$).
(d) Pelo método da secante (três iterações, $x_{-1} = 3$, $x_0 = 4$).
(e) Pelo método da secante modificado (três iterações, $x_0 = 3$, $\delta = 0{,}01$).
Calcule os erros relativos percentuais aproximados para suas soluções.

6.3 Use **(a)** a iteração de ponto fixo e **(b)** o método de Newton-Raphson para determinar a raiz de $f(x) = -x^2 + 1{,}8x + 2{,}5$ usando $x_0 = 5$. Faça os cálculos até que ε_a seja menor do que $\varepsilon_s = 0{,}05\%$. Faça também uma verificação do erro na sua resposta final.

6.4 Determine as raízes reais de $f(x) = -1 + 5{,}5x - 4x^2 + 0{,}5x^3$: **(a)** graficamente e **(b)** usando o método de Newton-Raphson até $\varepsilon_s = 0{,}01\%$.

6.5 Use o método de Newton-Raphson para determinar uma raiz real de $f(x) = -1 + 5{,}5x - 4x^2 + 0{,}5x^3$ usando aproximações iniciais **(a)** 4,52 e **(b)** 4,54. Discuta e use métodos gráficos e analíticos para explicar quaisquer peculiaridades nos seus resultados.

6.6 Determine a menor raiz real de $f(x) = -12 - 21x + 18x^2 - 2{,}4x^3$: **(a)** graficamente e **(b)** usando o método da secante para um valor de ε_s correspondente a três algarismos significativos.

6.7 Localize a primeira raiz positiva de
$f(x) = \text{sen}\, x + \cos(1 + x^2) - 1$
onde x está em radianos. Use quatro iterações do método da secante com aproximações iniciais **(a)** $x_{i-1} = 1{,}0$ e $x_i = 3{,}0$; **(b)** $x_{i-1} = 1{,}5$ e $x_i = 2{,}5$, e **(c)** $x_{i-1} = 1{,}5$ e $x_i = 2{,}25$ para localizar a raiz. **(d)** Use o método gráfico para explicar seus resultados.

6.8 Determine a raiz real de $x^{3{,}5} = 80$, com o método da secante modificado até $\varepsilon_s = 0{,}1\%$ usando uma aproximação inicial $x_0 = 3{,}5$ e $\delta = 0{,}01$.

6.9 Determine a maior raiz real de $f(x) = 0{,}95x^3 - 5{,}9x^2 + 10{,}9x - 6$:
(a) Graficamente.
(b) Usando o método de Newton-Raphson (três iterações, $x_i = 3{,}5$).
(c) Usando o método da secante (três iterações, $x_{i-1} = 2{,}5$ e $x_i = 3{,}5$).
(d) Usando o método da secante modificado (três iterações, $x_i = 3{,}5$ e $\delta = 0{,}01$).

6.10 Determine a menor raiz real de $f(x) = 8\,\text{sen}(x)e^{-x} - 1$:
(a) Graficamente.
(b) Usando o método de Newton-Raphson (três iterações, $x_i = 0{,}3$).
(c) Usando o método da secante (três iterações, $x_{i-1} = 0{,}5$ e $x_i = 0{,}4$).
(d) Usando o método da secante modificado (cinco iterações, $x_i = 0{,}3$ e $\delta = 0{,}01$).

6.11 A função $x^3 - 2x^2 - 4x + 8$ tem uma raiz dupla em $x = 2$. Use **(a)** Newton-Raphson padrão [Equação (6.6)], **(b)** Newton-Raphson modificado [Equação (6.9a)] e **(c)** Newton-Raphson modificado [Equação (6.13)] para determinar a raiz em $x = 2$. Compare e discuta a taxa de convergência usando uma aproximação inicial de $x_0 = 1{,}2$.

6.12 Determine as raízes das seguintes equações não-lineares simultâneas usando **(a)** iteração de ponto fixo e **(b)** o método de Newton-Raphson:
$y = -x^2 + x + 0{,}75$
$y + 5xy = x^2$
Use aproximações iniciais $x = y = 1{,}2$ e discuta os resultados.

6.13 Determine as raízes das equações não-lineares simultâneas
$(x - 4)^2 + (y - 4)^2 = 5$
$x^2 + y^2 = 16$
Use uma abordagem gráfica para obter suas aproximações iniciais. Determine estimativas refinadas com o método de Newton-Raphson para duas equações descrito na Seção 6.5.2.

6.14 Repita o Problema 6.13, agora para
$y = x^2 + 1$
$y = 2\cos x$

6.15 Um balanço de massa para um poluente em um lago bem misturado pode ser escrito como

$$V\frac{dc}{dt} = W - Qc - kV\sqrt{c}$$

Dados os valores dos parâmetros $V = 1 \times 10^6 \text{m}^3$, $Q = 1 \times 10^5\ \text{m}^3/\text{ano}$, $W = 1 \times 10^6$ g/ano e $k = 0{,}25\ \text{m}^{0,5}/\text{g}^{0,5}/\text{ano}$, use o método da secante modificado para determinar a concentração estacionária. Use uma aproximação inicial de $c = 4\ \text{g/m}^3$ e $\delta = 0{,}5$. Faça três iterações e determine o erro porcentual relativo depois da terceira iteração.

6.16 Para o Problema 6.15, a raiz pode ser localizada com a iteração de ponto fixo por

$$c = \left(\frac{W - Qc}{kV}\right)^2$$

ou por

$$c = \frac{W - kV\sqrt{c}}{Q}$$

Apenas um caso convergirá para aproximações iniciais $2 < c < 6$. Escolha o correto e mostre por que ele sempre funcionará.

6.17 Desenvolva um programa amigável ao usuário para o método de Newton-Raphson baseado na Figura 6.4 e na Seção 6.2.3. Teste-o repetindo os cálculos do Exemplo 6.3.

6.18 Desenvolva um programa amigável ao usuário para o método da secante baseado na Figura 6.4 e na Seção 6.2.3. Teste-o repetindo os cálculos do Exemplo 6.6.

6.19 Desenvolva um programa amigável ao usuário para o método da secante modificado baseado na Figura 6.4 e na Seção 6.2.3. Teste-o repetindo os cálculos do Exemplo 6.8.

6.20 Desenvolva um programa amigável ao usuário para o método de Newton-Raphson para duas equações baseado na Seção 6.5. Teste-o resolvendo o Exemplo 6.10.

6.21 Use o programa que você desenvolveu no Problema 6.20 para resolver os Problemas 6.12 e 6.13 até uma tolerância de $\varepsilon_s = 0{,}01\%$.

6.22 O método "divisão e média", um método antigo para aproximar a raiz quadrada de qualquer número positivo a, pode ser formulado como

$$x = \frac{x + a/x}{2}$$

Demonstre que isso é equivalente ao algoritmo de Newton-Raphson.

6.23 **(a)** Aplique o método de Newton-Raphson à função $f(x) = \text{tgh}\,(x^2 - 9)$ para calcular sua raiz real conhecida em $x = 3$. Use uma aproximação inicial $x_0 = 3{,}1$ e faça um mínimo de quatro iterações. **(b)** O método convergiu para sua raiz real? Esboce o gráfico, com os resultados para cada iteração identificados.

6.24 O polinômio $f(x) = 0{,}0074x^4 - 0{,}284x^3 + 3{,}355x^2 - 12{,}183x + 5$ tem uma raiz real entre 15 e 20. Aplique o método de Newton-Raphson a essa função usando uma aproximação inicial $x_0 = 16{,}15$. Explique seus resultados.

6.25 Use o método da secante na função do círculo $(x + 1)^2 + (y - 2)^2 = 16$ para encontrar uma raiz positiva. Tome sua aproximação inicial como $x_i = 3$ e $x_{i-1} = 0{,}5$. Aproxime-se da solução a partir do primeiro e quarto quadrantes. Quando calcular $f(x)$ no quarto quadrante, certifique-se de tomar o valor negativo da raiz quadrada. Por que a sua solução diverge?

6.26 Você está projetando um tanque esférico (Figura P6.26) para armazenar água para uma pequena vila em um país em desenvolvimento. O volume de líquido que ele pode armazenar pode ser calculado por

$$V = \pi h^2 \frac{[3R - h]}{3}$$

onde V é o volume [m^3], h é a profundidade da água no tanque [m] e R é o raio do tanque [m].

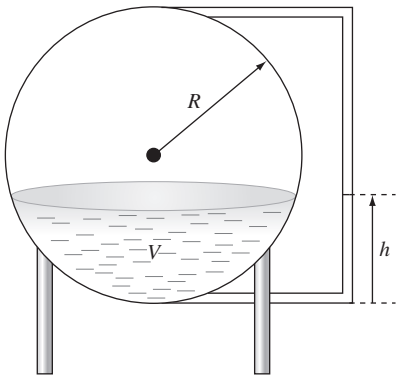


Figura P6.26

Se $R = 3$ m, até qual profundidade o tanque deve ser enchido para que armazene 30 m^3? Use três iterações do método de Newton-Raphson para determinar sua resposta. Determine o erro relativo aproximado depois de cada iteração. Observe que uma aproximação inicial R será sempre convergente.

CAPÍTULO 7

Raízes de Polinômios

Neste capítulo, serão discutidos métodos para encontrar as raízes de equações polinomiais da forma geral

$$f_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{7.1}$$

onde n é o grau do polinômio e os a's são coeficientes constantes. Embora os coeficientes possam ser números complexos, a discussão se limitará aos casos em que eles são reais. Para tais casos, as raízes podem ser reais e/ou complexas.

As raízes de tais polinômios seguem estas regras:

1. Para uma equação de grau n, existem n raízes reais ou complexas. Deve-se observar que essas raízes não serão necessariamente distintas.
2. Se n for ímpar, existe pelo menos uma raiz real.
3. Se existirem raízes complexas, elas existem em pares conjugados (isto é, $\lambda + \mu i$ e $\lambda - \mu i$) onde $i = \sqrt{-1}$.

Antes de descrever as técnicas para localizar as raízes dos polinômios, serão fornecidos alguns fundamentos. A primeira seção oferece alguma motivação para o estudo das técnicas; a segunda trata de algumas manipulações computacionais fundamentais envolvendo polinômios.

7.1 POLINÔMIOS NA ENGENHARIA E NA CIÊNCIA

Os polinômios têm muitas aplicações na engenharia e na ciência. Por exemplo, eles são usados amplamente no ajuste de curvas. Entretanto, uma de suas aplicações mais interessantes e poderosas é na caracterização de sistemas dinâmicos e, em particular, dos sistemas lineares. Os exemplos incluem aparelhos mecânicos, estruturas e circuitos elétricos. Serão explorados exemplos específicos por todo o resto deste texto.

Por agora, será mantida a discussão simples e geral focalizando em um sistema de segunda ordem simples definido pela seguinte *equação diferencial ordinária* linear (ou EDO):

$$a_2\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0y = F(t) \tag{7.2}$$

onde y e t são as variáveis dependente e independente, respectivamente, os a's são coeficientes constantes e $F(t)$ é um termo forçante. Se entender como essa equação é deduzida a partir de um sistema físico ajudar a motivar seu estudo de matemática, você poderia ler atentamente o início da Seção 8.4 antes de continuar.

Além disso, deve-se observar que a Equação (7.2) pode ser alternativamente expressa como um par de EDOs de primeira ordem definindo uma nova variável z,

$$z = \frac{dy}{dt} \tag{7.3}$$

A Equação (7.3) pode ser substituída junto com sua derivada na Equação (7.2) para se remover o termo da segunda derivada. Isso reduz o problema a resolver

$$\frac{dz}{dt} = \frac{F(t) - a_1z - a_0y}{a_2} \tag{7.4}$$

$$\frac{dy}{dt} = z \tag{7.5}$$

De forma análoga, uma EDO linear de ordem n pode sempre ser expressa como um sistema de n EDOs de primeira ordem.

Agora, a solução será considerada. O termo forçante representa o efeito do mundo externo no sistema. A solução homogênea ou *solução geral* da equação trata do caso no qual o termo forçante é igualada a zero,

$$a_2\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0y = 0 \tag{7.6}$$

Logo, como o nome sugere, a *solução geral* de todos os sistemas lineares com termo forçante nulo diz algo muito fundamental sobre o sistema que está sendo simulado — isto é, como o sistema responde na ausência de estímulos externos.

Agora, a solução geral de todos os sistemas lineares homogêneos é da forma $y = e^{rt}$. Se essa função for derivada e substituída na Equação (7.6), o resultado é

$$a_2r^2e^{rt} + a_1re^{rt} + a_0e^{rt} = 0$$

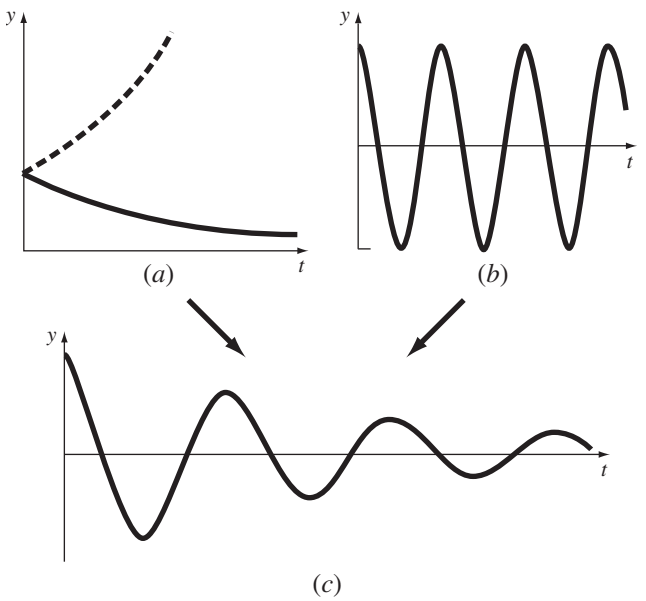
ou, cancelando os termos exponenciais,

$$a_2r^2 + a_1r + a_0 = 0 \tag{7.7}$$

Observe que o resultado é um polinômio chamado de *equação característica*. As raízes desse polinômio são os valores de r que satisfazem a Equação (7.7). Esses r's são chamados de valores característicos do sistema, ou *autovalores*.

Logo, aqui está a conexão entre raízes de polinômios e engenharia e ciência. O autovalor diz algo fundamental sobre o sistema que se está modelando e encontrar os autovalores envolve encontrar as raízes de polinômios. E, enquanto encontrar as raízes de uma equação de segundo grau é fácil usando a fórmula quadrática, encontrar raízes de sistemas de ordem superior (e, portanto, de polinômios de grau mais alto) é difícil (ou impossível) analiticamente. Logo, a melhor abordagem geral exige métodos numéricos do tipo descrito neste capítulo.

FIGURA 7.1
A solução geral das EDOs lineares pode ser composta de componentes (*a*) exponenciais e (*b*) senoidais. A combinação das duas formas resulta na função senoidal amortecida mostrada em (*c*).



Antes de prosseguir para esses métodos, é preciso avançar um pouco mais na análise investigando o que valores específicos de autovalores podem implicar no comportamento de sistemas físicos. Inicialmente, calculam-se as raízes da Equação (7.7) com a fórmula quadrática,

$$\begin{matrix} r_1 \\ r_2 \end{matrix} = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2a_0}}{a_0}$$

Portanto, obtêm-se duas raízes. Se o *discriminante* ($a_1^2 - 4a_2a_0$) for positivo, as raízes são reais e a solução geral pode ser representada por

$$y = c_1e^{r_1t} + c_2e^{r_2t} \tag{7.8}$$

onde os c's são onstantes que podem ser determinadas a partir das condições iniciais. Esse é o chamado *caso superamortecido*.

Se o discriminante for nulo, resulta uma única raiz real e a solução geral pode ser formulada como

$$y = (c_1 + c_2t)e^{\lambda t} \tag{7.9}$$

Esse é o chamado *caso criticamente amortecido*.

Se o discriminante for negativo, as raízes serão números complexos conjugados,

$$\begin{matrix} r_1 \\ r_2 \end{matrix} = \lambda \pm \mu i$$

e a solução geral pode ser formulada como

$$y = c_1e^{(\lambda+\mu i)t} + c_2e^{(\lambda-\mu i)t}$$

O comportamento físico dessa solução pode ser elucidada usando-se a fórmula de Euler

$$e^{\mu it} = \cos \mu t + i \text{ sen } \mu\text{t}$$

para reformular a solução geral como (ver Boyce e DiPrima, 1992, para detalhes da dedução)

$$y = c_1e^{\lambda t} \cos \mu t + c_2e^{\lambda t} \text{ sen } \mu\text{t} \tag{7.10}$$

Esse é chamado o *caso subamortecido*.

As Equações (7.8), (7.9) e (7.10) expressam as possíveis maneiras de sistemas lineares responderem dinamicamente. Os termos exponenciais significam que as soluções podem decair (parte real negativa) ou crescer (parte real positiva) exponencialmente com o tempo (Figura 7.1*a*). Os termos senoidais (parte imaginária) significam que a solução pode oscilar (Figura 7.1*b*). Se o autovalor tiver tanto parte real quanto imaginária, as formas exponencial e senoidal são combinadas (Figura 7.1*c*). Como tal conhecimento é um elemento-chave no entendimento, no projeto e no controle do comportamento de um sistema físico, os polinômios característicos são muito importantes na engenharia e em muitos ramos da ciência. Será explorada a dinâmica de diversos sistemas em engenharia nas aplicações cobertas no Capítulo 8.

7.2 CÁLCULOS COMPUTACIONAIS COM POLINÔMIOS

Antes de se descrever os métodos de localização de raízes, serão abordadas algumas operações computacionais fundamentais envolvendo polinômios. Elas têm utilidade por si só, além de fornecerem suporte na procura das raízes.

7.2.1 Cálculo e Derivação de Polinômios

Embora seja a forma mais comum, a Equação (7.1) fornece métodos insatisfatórios para determinar o valor de um polinômio em um valor de x particular. Por exemplo, calcular um polinômio de terceiro grau como

$$f_3(x) = a_3x^3 + a_2x^2 + a_1x + a_0 \tag{7.11}$$

envolve seis multiplicações e três adições. Em geral, para um polinômio de grau n, essa abordagem exige $n(n + 1)/2$ multiplicações e n adições.

Em contraste, uma forma estruturada,

$$f_3(x) = ((a_3x + a_2)x + a_1)x + a_0 \tag{7.12}$$

envolve três multiplicações e três adições. Para um polinômio de grau n, essa abordagem exige n multiplicações e n adições. Como a forma estruturada minimiza o número de operações, tende a minimizar os erros de arredondamento. Observe que, dependendo de sua preferência, a ordem da estrutura pode ser invertida:

$$f_3(x) = a_0 + x(a_1 + x(a_2 + xa_3)) \tag{7.13}$$

Um pseudocódigo sucinto para implementar a forma estruturada pode ser escrito simplesmente como

```
DOFOR j = n, 0, -1
  p = p * x+a(j)
END DO
```

onde p guarda o valor do polinômio (definido por seus coeficientes, os a's) calculado em x.

Há casos (como no método de Newton-Raphson) em que se pode querer calcular tanto a função quanto sua derivada. Esse cálculo também pode ser naturalmente incluído adicionando-se uma única linha ao pseudocódigo anterior,

```
DOFOR j = n, 0, -1

  df = df * x+p

  p = p * x+a(j)
END DO
```

onde df guarda o valor da primeira derivada do polinômio.

7.2.2 Deflação Polinomial

Suponha que você tenha determinado uma única raiz de um polinômio de grau n. Se repetir seu procedimento para a localização de raízes, você poderá encontrar a mesma raiz. Portanto, seria bom remover a raiz encontrada antes de continuar. Esse processo de remoção é chamado de *deflação polinomial*.

Antes de mostrar como isso é feito, alguma orientação poderia ser útil. Os polinômios são tipicamente representados na forma da Equação (7.1). Por exemplo, um polinômio de grau cinco poderia ser escrito como

$$f_5(x) = -120 - 46x + 79x^2 - 3x^3 - 7x^4 + x^5 \tag{7.14}$$

Embora essa seja uma forma familiar, ela não é necessariamente a melhor expressão para entender o comportamento matemático do polinômio. Por exemplo, o polinômio de grau cinco poderia ser expresso alternativamente por

$$f_5(x) = (x + 1)(x - 4)(x - 5)(x + 3)(x - 2) \tag{7.15}$$

Essa é a chamada forma *fatorada* do polinômio. Se fossem completadas as multiplicações e os termos semelhantes fossem agrupados, seria obtida a Equação (7.14). Entretanto, a forma da Equação (7.15) tem a vantagem de indicar claramente as raízes da função. Portanto, é claro que $x = -1$, 4, 5, -3 e 2 são todas raízes porque cada uma faz com que um termo individual na Equação (7.15) se anule.

Agora, suponha que esse polinômio de grau cinco seja dividido por qualquer um de seus fatores, por exemplo, $x + 3$. Nesse caso, o resultado será um polinômio de grau quatro

$$f_4(x) = (x + 1)(x - 4)(x - 5)(x - 2) = -40 - 2x + 27x^2 - 10x^3 + x^4 \qquad (7.16)$$

com um resto nulo.

No passado distante, você provavelmente aprendeu a dividir polinômios usando uma abordagem chamada *divisão sintética*[1]. Diversos algoritmos de computador (baseados tanto na divisão sintética quanto em outros métodos) estão disponíveis para fazer a operação. Um esquema simples é fornecido pelo seguinte pseudocódigo, que divide um polinômio de grau n por um fator do tipo monômio $x - t$:

```
r = a(n)
a(n) = 0
DOFOR i = n-1, 0, -1
  s = a(i)
  a(i) = r
  r = s + r * t
END DO
```

Se o monômio corresponder a uma raiz do polinômio, o resto r será nulo e os coeficientes do quociente estarão armazenados em a no final do laço.

EXEMPLO 7.1 Deflação Polinomial

Enunciado do Problema. Divida o polinômio de segunda ordem

$$f(x) = (x - 4)(x + 6) = x^2 + 2x - 24$$

pelo fator $x - 4$.

Solução. Usando a abordagem delineada no pseudocódigo anterior, os parâmetros são $n = 2$, $a_0 = -24$, $a_1 = 2$, $a_2 = 1$ e $t = 4$, que podem ser usados para calcular

$$r = a_2 = 1$$
$$a_2 = 0$$

O laço é então iterado de $i = 2 - 1 = 1$ a 0. Para $i = 1$,

$$s = a_1 = 2$$
$$a_1 = r = 1$$
$$r = s + rt = 2 + 1(4) = 6$$

Para $i = 0$,

$$s = a_0 = -24$$
$$a_0 = r = 6$$
$$r = -24 + 6(4) = 0$$

Portanto, o resultado é como esperado — o quociente é $a_0 + a_1x = 6 + x$, com resto nulo.

Também é possível dividir por polinômios de grau mais alto. Como veremos mais adiante neste capítulo, a tarefa mais comum envolve a divisão por um polinômio de grau dois ou uma parábola. A sub-rotina na Figura 7.2 trata do problema mais geral de dividir um polinômio a de grau n por um polinômio d de grau m. O resultado é um polinômio q de grau $n - m$, com um polinômio de grau $m - 1$ como resto.

Como cada raiz calculada é conhecida apenas aproximadamente, deve ser observado que a deflação é sensível a erros de arredondamento. Em alguns casos, eles podem crescer a ponto de os resultados ficarem sem significado.

[1] N.R.T.: Algoritmo de Briot-Ruffine.

```
SUB poldiv(a, n, d, m, q, r)
DOFOR j = 0, n
  r(j) = a(j)
  q(j) = 0
END DO
DOFOR k = n-m, 0, -1
  q(k+1) = r(m+k) / d(m)
  DOFOR j = m+k-1, k, -1
    r(j) = r(j)-q(k+1) * b(j-k)
  END DO
END DO
DOFOR j = m, n
  r(j) = 0
END DO
n = n-m
DOFOR i = 0, n
  a(i) = q(i+1)
END DO
END SUB
```

FIGURA 7.2
Algoritmo para dividir um polinômio *a* (definido por seus coeficientes) por um polinômio *d* de grau menor.

Algumas estratégias gerais podem ser aplicadas para minimizar esse problema. Por exemplo, o erro de arredondamento é afetado pela ordem na qual os termos são calculados. A *deflação progressiva* se refere ao caso no qual os coeficientes dos novos polinômios estão em ordem de potências decrescente de x (isto é, do que tem maior grau para o termo de grau zero). Nesse caso, é preferível dividir pelas raízes de menor valor absoluto primeiro. Reciprocamente, para a *deflação regressiva* (isto é, do que tem grau zero para o termo de maior grau), é preferível dividir pelas raízes de maior valor absoluto primeiro.

Uma outra forma de reduzir os erros de arredondamento é considerar cada estimativa sucessiva da raiz obtida durante a deflação como uma boa primeira aproximação. Esses arredondamentos podem ser usados como aproximações iniciais, e as raízes determinadas novamente com o polinômio original não-deflacionado. Isso é chamado *polimento da raiz*.

Finalmente, aparece um problema quando duas raízes deflacionadas são suficientemente inacuradas para que ambas convirjam para uma mesma raiz não-deflacionada. Nesse caso, alguém pode ser levado erroneamente a acreditar que o polinômio tenha uma raiz múltipla (lembre-se da Seção 6.4). Uma forma de detectar tal problema é comparar cada raiz polida com aquelas localizadas anteriormente. Press et al. (1992) discutiram esse problema em mais detalhes.

7.3 MÉTODOS CONVENCIONAIS

Agora que cobrimos algum material fundamental sobre polinômios, podemos começar a descrever os métodos para localizar suas raízes. O primeiro passo óbvio seria investigar a viabilidade das abordagens intervalares e dos métodos abertos descritos nos Capítulos 5 e 6.

A eficácia dessas abordagens depende do fato de o problema que está sendo resolvido envolver ou não raízes complexas. Se houver apenas raízes reais, qualquer um dos métodos descritos anteriormente pode ser utilizado. Entretanto, encontrar boas aproximações iniciais complica tanto os métodos intervalares quanto os métodos abertos, enquanto que os métodos abertos são suscetíveis a divergência.

Quando são possíveis raízes complexas, os métodos que isolam as raízes não podem ser usados por causa do problema óbvio de que o critério para definir um intervalo que delimite a raiz (isto é, onde ocorra mudança de sinal) não se estende para as aproximações complexas.

Dos métodos abertos, o método de Newton-Raphson tradicional forneceria uma abordagem viável. Em particular, pode-se desenvolver um código conciso que inclua a deflação. Se for usada uma linguagem que permita variáveis complexas (como o Fortran), tal algoritmo localizará tanto as raízes reais quanto as complexas. Contudo, como se poderia esperar, ele seria suscetível a problemas de convergência. Por essa razão, foram desenvolvidos métodos especiais para encontrar as raízes reais e complexas de polinômios. Descreveremos dois deles — os métodos de Müller e de Bairstow — na próxima seção. Como você verá, ambos estão relacionados à abordagem com os métodos abertos mais convencionais descrita no Capítulo 6.

7.4 O MÉTODO DE MÜLLER

Lembre-se de que o método da secante obtém uma estimativa da raiz prolongando uma reta por dois valores da função até o eixo x (Figura 7.3*a*). O método de Müller adota uma abordagem parecida, mas prolonga uma parábola através de três pontos (Figura 7.3*b*).

O método consiste na determinação dos coeficientes da parábola que passa pelos três pontos. Esses coeficientes podem então ser substituídos na fórmula quadrática para se obter o ponto no qual a parábola intercepta o eixo x — isto é, a estimativa da raiz. O método se torna mais fácil escrevendo-se a equação da parábola em uma forma conveniente,

$$f_2(x) = a(x - x_2)^2 + b(x - x_2) + c \tag{7.17}$$

Queremos que essa parábola passe por três pontos $(x_0, f(x_0))$, $(x_1, f(x_1))$ e $(x_2, f(x_2))$. Os coeficientes da Equação (7.17) podem ser calculados substituindo cada um dos três pontos para obter

$$f(x_0) = a(x_0 - x_2)^2 + b(x_0 - x_2) + c \tag{7.18}$$

$$f(x_1) = a(x_1 - x_2)^2 + b(x_1 - x_2) + c \tag{7.19}$$

$$f(x_2) = a(x_2 - x_2)^2 + b(x_2 - x_2) + c \tag{7.20}$$

Observe que omitimos o subscrito "2" da função para sermos concisos. Como temos três equações, é possível resolver para determinar os três coeficientes desconhecidos a, b e c.

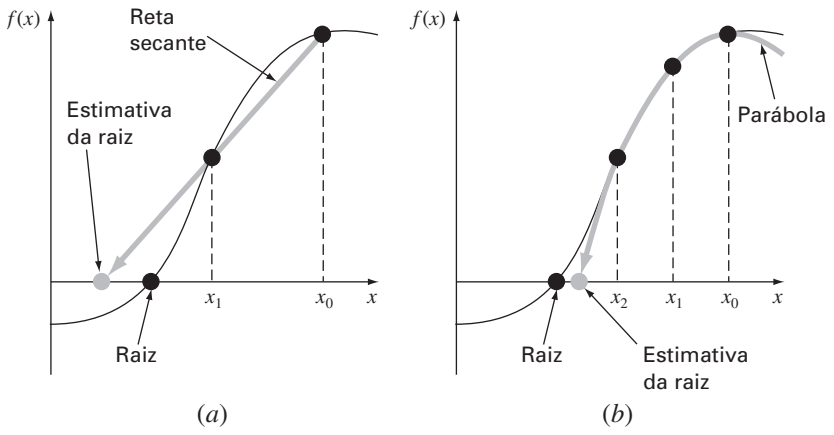


FIGURA 7.3
Uma comparação de duas abordagens relacionadas para localizar raízes: (*a*) o método da secante e (*b*) o método de Müller.

Já que dois dos termos na Equação (7.20) são nulos, ela pode ser imediatamente resolvida, fornecendo $c = f(x_2)$. Logo, o coeficiente c é simplesmente igual ao valor da função calculada na terceira aproximação, x_2. Esse resultado pode ser substituído nas Equações (7.18) e (7.19) para fornecer duas equações com duas incógnitas:

$$f(x_0) - f(x_2) = a(x_0 - x_2)^2 + b(x_0 - x_2) \tag{7.21}$$

$$f(x_1) - f(x_2) = a(x_1 - x_2)^2 + b(x_1 - x_2) \tag{7.22}$$

Pode-se então usar manipulação algébrica para determinar os coeficientes restantes a e b. Uma maneira de fazer isso envolve a definição de algumas diferenças,

$$h_0 = x_1 - x_0 \qquad h_1 = x_2 - x_1$$
$$\delta_0 = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \qquad \delta_1 = \frac{f(x_2) - f(x_1)}{x_2 - x_1} \tag{7.23}$$

Elas podem ser substituídas nas Equações (7.21) e (7.22) para fornecer

$$(h_0 + h_1)b - (h_0 + h_1)^2 a = h_0\delta_0 + h_1\delta_1$$
$$h_1 \quad b - \quad h_1^2 \quad a = \quad h_1\delta_1$$

a partir das quais se pode isolar a e b. O resultado pode ser resumido como

$$a = \frac{\delta_1 - \delta_0}{h_1 + h_0} \tag{7.24}$$

$$b = ah_1 + \delta_1 \tag{7.25}$$

$$c = f(x_2) \tag{7.26}$$

Para encontrar a raiz, aplicamos a fórmula quadrática à Equação (7.17). No entanto, por causa do possível erro de arredondamento, em vez de usar a forma convencional, será usada a formulação alternativa [Equação (3.13)] para obter

$$x_3 - x_2 = \frac{-2c}{b \pm \sqrt{b^2 - 4ac}} \tag{7.27a}$$

ou, isolando a incógnita x_3 no lado esquerdo da igualdade,

$$x_3 = x_2 + \frac{-2c}{b \pm \sqrt{b^2 - 4ac}} \tag{7.27b}$$

Observe que o uso da fórmula quadrática significa que podem ser localizadas tanto raízes reais quanto complexas. Essa é uma importante vantagem do método.

Além disso, a Equação (7.27*a*) fornece uma forma natural de determinar o erro aproximado. Como o lado esquerdo representa a diferença entre a estimativa atual (x_3) e anterior (x_2) da raiz, o erro pode ser calculado por

$$\varepsilon_a = \left|\frac{x_3 - x_2}{x_3}\right| 100\%$$

Agora, um problema com a Equação (7.27*a*) é que ela fornece duas raízes, correspondendo ao termo $\pm$ no denominador. No método de Müller, o sinal é escolhido para coincidir com o sinal de b. Essa escolha resultará no denominador maior e, portanto, dará a estimativa da raiz que está mais próxima de x_2.

Uma vez que x_3 esteja determinada, o processo é repetido. Isso traz à tona a questão de qual ponto descartar. Duas estratégias gerais são tipicamente usadas:

1. Se estiverem sendo localizadas apenas raízes reais, escolhemos os dois pontos originais que estejam mais próximos da nova estimativa da raiz, x_3.
2. Se estiverem sendo calculadas tanto raízes reais quanto complexas, é usada uma abordagem seqüencial. Isto é, da mesma forma que no método da secante, x_1, x_2 e x_3 tomam o lugar de x_0, x_1 e x_2.

EXEMPLO 7.2 O Método de Müller

Enunciado do Problema. Use o Método de Müller com aproximações x_0, x_1 e x_2 dadas por 4,5, 5,5 e 5, respectivamente, para determinar uma raiz da equação

$$f(x) = x^3 - 13x - 12$$

Observe que as raízes dessa equação são -3, -1 e 4.

Solução. Inicialmente, é calculada a função nas aproximações

$$f(4{,}5) = 20{,}625 \qquad f(5{,}5) = 82{,}875 \qquad f(5) = 48$$

que podem ser usadas para calcular

$$h_0 = 5{,}5 - 4{,}5 = 1 \qquad h_1 = 5 - 5{,}5 = -0{,}5$$
$$\delta_0 = \frac{82{,}875 - 20{,}625}{5{,}5 - 4{,}5} = 62{,}25 \qquad \delta_1 = \frac{48 - 82{,}875}{5 - 5{,}5} = 69{,}75$$

Esses valores, por sua vez, podem ser substituídos nas Equações (7.24) a (7.26) para calcular

$$a = \frac{69{,}75 - 62{,}25}{-0{,}5 + 1} = 15 \qquad b = 15(-0{,}5) + 69{,}75 = 62{,}25 \qquad c = 48$$

A raiz quadrada do discriminante pode ser calculada por

$$\sqrt{62{,}25^2 - 4(15)48} = 31{,}54461$$

Então, como $|62{,}25 + 31{,}54451| > |62{,}25 - 31{,}54451|$, o sinal positivo é usado no denominador da Equação (7.27*b*) e uma nova estimativa da raiz é determinada por

$$x_3 = 5 + \frac{-2(48)}{62{,}25 + 31{,}54451} = 3{,}976487$$

e a estimativa de erro por

$$\varepsilon_a = \left|\frac{-1{,}023513}{3{,}976487}\right| 100\% = 25{,}74\%$$

Como o erro é grande, novas aproximações são escolhidas; x_0 é substituído por x_1, x_1 é substituído por x_2, x_2 é substituído por x_3. Portanto, para a nova iteração,

$$x_0 = 5{,}5 \qquad x_1 = 5 \qquad x_2 = 3{,}976487$$

e os cálculos são repetidos. Os resultados, tabulados a seguir, mostram que o método converge rapidamente para a raiz $x_r = 4$:

i	x_r	ε_a (%)
0	5	
1	3,976487	25,74
2	4,00105	0,6139
3	4	0,0262
4	4	0,0000119

O pseudocódigo para implementar o método de Müller para raízes reais é apresentado na Figura 7.4. Observe que essa rotina está construída de forma a usar como entrada uma única aproximação inicial não-nula que então é perturbada para determinar as outras duas aproximações. É claro que o algoritmo poderia ser programado para acomodar três aproximações. Para linguagens como o Fortran, o código encontrará raízes complexas se as variáveis adequadas forem declaradas como complexas.

```
SUB Muller(xr, h, eps, maxit)
x2 = xr
x1 = xr + h*xr
x0 = xr − h*xr
DO
  iter = iter + 1
  h0 = x1 − x0
  h1 = x2 − x1
  d0 = (f(x1) − f(x0 )) / h0
  d1 = (f(x2) − f(x1)) / h1
  a = (d1 − d0) / (h1 + h0 )
  b = a*h1 + d1
  c = f(x2)
  rad = SQRT(b*b − 4*a*c)
  If |b+rad| > |b−rad| THEN
     den = b + rad
  ELSE
    den = b − rad
  END IF
  dxr = −2*c / den
  xr = x2 + dxr
  PRINT iter, xr
  IF (|dxr| < eps*xr OR iter >= maxit) EXIT
  x0 = x1
  x1 = x2
  x2 = xr
END DO
END Müller
```

FIGURA 7.4
Pseudocódigo para o método de Müller.

7.5 O MÉTODO DE BAIRSTOW

O método de Bairstow é uma abordagem iterativa vagamente relacionada a ambos os métodos, de Müller e de Newton-Raphson. Antes de se fazer uma descrição matemática da técnica, lembre-se da forma fatorada do polinômio,

$$f_5(x) = (x + 1)(x - 4)(x - 5)(x + 3)(x - 2) \tag{7.28}$$

Se dividirmos por um fator que não corresponda a uma raiz (por exemplo, $x + 6$), o quociente será um polinômio de quarto grau. Entretanto, nesse caso, resultará um resto não-nulo.

Com base no exposto, pode-se elaborar um algoritmo para determinar uma raiz de um polinômio: (1) escolha um valor para a raiz $x = t$, (2) divida o polinômio pelo fator $x - t$ e (3) determine se existe um resto. Se não existir, a escolha foi perfeita e a raiz é igual a t. Se existir um resto, a escolha pode ser sistematicamente ajustada e o procedimento ser repetido até que o resto desapareça e a raiz seja localizada. Depois de conseguir isso, todo o procedimento pode ser repetido para o quociente, para determinar uma outra raiz.

O método de Bairstow é, de modo geral, baseado nessa abordagem. Conseqüentemente, ele depende do processo matemático de dividir um polinômio por um fator. Da discussão sobre deflação de polinômios (Seção 7.2.2), lembre-se de que a divisão sintética envolve a divisão de um polinômio por um fator $x - t$. Por exemplo, o polinômio geral [Equação (7.1)]

$$f_n(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{7.29}$$

pode ser dividido pelo fator $x - t$ para fornecer um segundo polinômio que tem um grau a menos,

$$f_{n-1}(x) = b_1 + b_2 x + b_3 x^2 + \cdots + b_n x^{n-1} \tag{7.30}$$

com um resto $R = b_0$, onde os coeficientes podem ser calculados pela relação de recorrência

$$b_n = a_n$$
$$b_i = a_i + b_{i+1} t \qquad \text{para } i = n - 1 \text{ a } 0$$

Observe que, se t fosse uma raiz do polinômio original, o resto b_0 seria igual a zero.

Para permitir o cálculo de raízes complexas, o método de Bairstow divide o polinômio por um fator quadrático $x^2 - rx - s$. Isso é feito para a Equação (7.29), o resultado é um novo polinômio

$$f_{n-2}(x) = b_2 + b_3 x + \cdots + b_{n-1} x^{n-3} + b_n x^{n-2}$$

com um resto

$$R = b_1(x - r) + b_0 \tag{7.31}$$

Como no caso da divisão sintética normal, uma relação de recorrência simples pode ser usada para fazer a divisão pelo fator quadrático:

$$b_n = a_n \tag{7.32a}$$

$$b_{n-1} = a_{n-1} + r b_n \tag{7.32b}$$

$$b_i = a_i + r b_{i+1} + s b_{i+2} \qquad \text{para } i = n - 2 \text{ a } 0 \tag{7.32c}$$

O fator quadrático é introduzido para permitir a determinação de raízes complexas. Isso se relaciona ao fato de que, se os coeficientes do polinômio original forem reais, as raízes complexas ocorrem em pares conjugados. Se $x^2 - rx - s$ for um divisor exato do polinômio, as raízes podem ser determinadas pela fórmula quadrática. Portanto, o método se reduz a determinar valores de r e s que tornam o fator quadrático um divisor exato. Em outras palavras, procuram-se os valores que tornam o termo do resto igual a zero.

Uma inspeção da Equação (7.31) leva a concluir que, para o resto ser nulo, b_0 e b_1 precisam ser nulos. Como é pouco provável que nossas aproximações iniciais dos valores de r e s levem a esse resultado, devemos determinar uma maneira sistemática de modificar nossas aproximações de modo que b_0 e b_1 se aproximem de zero. Para fazê-lo, o método de Bairstow usa uma estratégia parecida com a abordagem de Newton-Raphson. Como b_0 e b_1 são funções tanto de r quanto de s, podem ser expandidos usando-se uma série de Taylor, como em [lembre-se da Equação (4.26)]

$$b_1(r + \Delta r, s + \Delta s) = b_1 + \frac{\partial b_1}{\partial r}\Delta r + \frac{\partial b_1}{\partial s}\Delta s$$
$$b_0(r + \Delta r, s + \Delta s) = b_0 + \frac{\partial b_0}{\partial r}\Delta r + \frac{\partial b_0}{\partial s}\Delta s \tag{7.33}$$

onde os valores no lado direito são todos calculados em r e s. Observe que o termo de segunda ordem e os de ordem mais alta foram desprezados, o que representa uma hipótese implícita de que Δr e Δs são suficientemente pequenos para que os termos de ordem superior sejam desprezíveis. Uma outra forma de expressar essa hipótese é dizer que as aproximações iniciais estão adequadamente próximas dos valores de r e s nas raízes.

As variações, Δr e Δs, necessárias para melhorar nossas aproximações podem ser estimadas igualando-se a Equação (7.33) a zero para obter

$$\frac{\partial b_1}{\partial r}\Delta r + \frac{\partial b_1}{\partial s}\Delta s = -b_1 \tag{7.34}$$

$$\frac{\partial b_0}{\partial r}\Delta r + \frac{\partial b_0}{\partial s}\Delta s = -b_0 \tag{7.35}$$

Se as derivadas parciais dos b's puderem ser determinadas, elas são um sistema de duas equações que podem ser resolvidas simultaneamente para as duas incógnitas, Δr e Δs. Bairstow mostrou que as derivadas parciais podem ser obtidas por uma divisão sintética dos b's de um modo análogo à maneira como os próprios b's foram determinados:

$$c_n = b_n \tag{7.36a}$$

$$c_{n-1} = b_{n-1} + rc_n \tag{7.36b}$$

$$c_i = b_i + rc_{i+1} + sc_{i+2} \qquad \text{para } i = n-2 \text{ a } 1 \tag{7.36c}$$

onde $\partial b_0/\partial r = c_1$, $\partial b_0/\partial s = \partial b_1/\partial r = c_2$, e $\partial b_1/\partial s = c_3$. Portanto, as derivadas parciais são obtidas por uma divisão sintética dos b's. A seguir, as derivadas parciais podem ser substituídas nas Equações (7.34) e (7.35) junto com os b's para fornecer

$$c_2\Delta r + c_3\Delta s = -b_1$$

$$c_1\Delta r + c_2\Delta s = -b_0$$

Essas equações podem ser resolvidas para determinar Δr e Δs, os quais, por sua vez, podem ser usados para melhorar as aproximações iniciais de r e s. Em cada passo, pode-se fazer uma estimativa de um erro em r e um em s, como em

$$|\varepsilon_{a,r}| = \left|\frac{\Delta r}{r}\right| 100\% \tag{7.37}$$

e

$$|\varepsilon_{a,s}| = \left|\frac{\Delta s}{s}\right| 100\% \tag{7.38}$$

Quando ambas as estimativas de erro caírem abaixo de um critério de parada pré-especificado ε_s, os valores das raízes podem ser determinados por

$$x = \frac{r \pm \sqrt{r^2 + 4s}}{2} \tag{7.39}$$

Nesse ponto, há três possibilidades:

1. *O quociente é um polinômio de grau maior ou igual a três*. Nesse caso, o método de Bairstow seria aplicado ao quociente para calcular novos valores de r e s. Os valores anteriores de r e s podem servir como aproximações iniciais nessa aplicação.
2. *O quociente é um polinômio quadrático*. Nesse caso, as duas raízes restantes poderiam ser calculadas diretamente com a Equação (7.39).
3. *O quociente é um polinômio de grau um*. Nesse caso, a única raiz restante pode ser calculada simplesmente por

$$x = -\frac{s}{r} \tag{7.40}$$

EXEMPLO 7.3 Método de Bairstow

Enunciado do Problema. Use o método de Bairstow para determinar as raízes do polinômio

$$f_5(x) = x^5 - 3{,}5x^4 + 2{,}75x^3 + 2{,}125x^2 - 3{,}875x + 1{,}25$$

Use aproximações iniciais $r = s = -1$ e itere até o nível de $\varepsilon_s = 1\%$.

Solução. As Equações (7.32) e (7.36) podem ser usadas para calcular

$$b_5 = 1 \qquad b_4 = -4{,}5 \qquad b_3 = 6{,}25 \qquad b_2 = 0{,}375 \qquad b_1 = -10{,}5$$
$$b_0 = 11{,}375$$
$$c_5 = 1 \qquad c_4 = -5{,}5 \qquad c_3 = 10{,}75 \qquad c_2 = -4{,}875 \qquad c_1 = -16{,}375$$

Logo, as equações simultâneas para Δr e Δs que devem ser resolvidas são

$$-4{,}875\Delta r + 10{,}75\Delta s = 10{,}5$$
$$-16{,}375\Delta r - 4{,}875\Delta s = -11{,}375$$

e cuja solução é $\Delta r = 0{,}3558$ e $\Delta s = 1{,}1381$. Portanto, nossas aproximações iniciais podem ser corrigidas para

$$r = -1 + 0{,}3558 = -0{,}6442$$
$$s = -1 + 1{,}1381 = 0{,}1381$$

e os erros aproximados podem ser calculados pelas Equações (7.37) e (7.38),

$$|\varepsilon_{a,r}| = \left|\frac{0{,}3558}{-0{,}6442}\right| 100\% = 55{,}23\% \qquad |\varepsilon_{a,s}| = \left|\frac{1{,}1381}{0{,}1381}\right| 100\% = 824{,}1\%$$

A seguir, os cálculos são repetidos usando-se os valores revistos de r e s. Aplicando as Equações (7.32) e (7.36) obtem-se:

$$b_5 = 1 \qquad b_4 = -4{,}1442 \qquad b_3 = 5{,}5578 \qquad b_2 = -2{,}0276 \qquad b_1 = -1{,}8013$$
$$b_0 = 2{,}1304$$
$$c_5 = 1 \qquad c_4 = -4{,}7884 \qquad c_3 = 8{,}7806 \qquad c_2 = -8{,}3454 \qquad c_1 = 4{,}7874$$

Portanto, é preciso resolver

$$-8{,}3454\Delta r + 8{,}7806\Delta s = 1{,}8013$$
$$4{,}7874\Delta r - 8{,}3454\Delta s = -2{,}1304$$

e obtêm-se $\Delta r = 0{,}1331$ e $\Delta s = 0{,}3316$, os quais podem ser usados para corrigir as estimativas da raiz para

$$r = -0{,}6442 + 0{,}1331 = -0{,}5111 \qquad |\varepsilon_{a,r}| = 26{,}0\%$$
$$s = 0{,}1381 + 0{,}3316 = 0{,}4697 \qquad |\varepsilon_{a,s}| = 70{,}6\%$$

Pode-se continuar os cálculos, com o resultado de que após quatro iterações o método converge para os valores $r = -0{,}5$ ($|\varepsilon_{a,r}| = 0{,}063\%$) e $s = 0{,}5$ ($|\varepsilon_{a,s}| = 0{,}040\%$). A Equação (7.39) pode então ser usada para calcular as raízes como sendo

$$x = \frac{-0{,}5 \pm \sqrt{(-0{,}5)^2 + 4(0{,}5)}}{2} = 0{,}5, -1{,}0$$

Nesse ponto, o quociente é a equação cúbica

$$f(x) = x^3 - 4x^2 + 5{,}25x - 2{,}5$$

O método de Bairstow pode ser aplicado a esse polinômio usando-se os resultados do passo anterior, $r = -0{,}5$ e $s = 0{,}5$, como aproximações iniciais. Cinco iterações fornecem as estimativas $r = 2$ e $s = -1{,}249$, as quais podem ser usadas para calcular

$$x = \frac{2 \pm \sqrt{2^2 + 4(-1{,}249)}}{2} = 1 \pm 0{,}499i$$

Nesse ponto, o quociente é um polinômio de grau um que pode ser resolvido diretamente pela Equação (7.40) para determinar a quinta raiz: 2.

Observe que o cerne do método de Bairstow é o cálculo dos b's e dos c's usando as Equações (7.32) e (7.36). Um dos principais pontos fortes do método é a maneira concisa como essas relações de recorrência podem ser programadas.

A Figura 7.5 lista um pseudocódigo para implementar o método de Bairstow. O cerne do algoritmo consiste no laço para calcular os b's e os c's. Observe também que o código para resolver as equações simultâneas faz uma verificação para prevenir a divisão por zero. Se esse for o caso, os valores de r e s são ligeiramente perturbados e o

procedimento é iniciado novamente. Além disso, o algoritmo coloca um limitante superior, definido pelo usuário, no número máximo de iterações (MAXIT) e deveria ser planejado para evitar a divisão por zero ao calcular as estimativas de erros. Finalmente, o algoritmo necessita das aproximações iniciais para *r* e *s* (*rr* e *ss* no código). Se não houver nenhum conhecimento prévio sobre as raízes, elas podem ser igualadas a zero no programa de chamada.

FIGURA 7.5
(*a*) Algoritmo para implementar o método de Bairstow, junto com (*b*) um algoritmo para determinar as raízes de um polinômio quadrático.

(*a*) Algoritmo de Bairstow

```
SUB Bairstow (a,nn,es,rr,ss,maxit,re,im,ier)
DIMENSION b(nn), c(nn)
r = rr; s = ss; n = nn
ier = 0; ea1 = 1; ea2 = 1
DO
  IF n < 3 OR iter ≥ maxit EXIT
  iter = 0
  DO
    iter = iter + 1
    b(n) = a(n)
    b(n − 1) = a(n − 1) + r * b(n)
    c(n) = b(n)
    c(n − 1) = b(n − 1) + r * c(n)
    DO i = n − 2, 0, −1
      b(i) = a(i) + r * b(i+1) + s * b(i + 2)
      c(i) = b(i) + r * c(i+1) + s * c(i + 2)
    END DO
    det = c(2) * c(2)− c(3) * c(1)
    IF det ≠ 0 THEN
       dr = (−b(1) * c(2) + b(0) * c(3))/det
       ds = (−b(0) * c(2) + b(1) * c(1))/det
       r = r + dr
       s = s + ds
       IF r≠0 THEN ea1 = ABS(dr/r) * 100
       IF s≠0 THEN ea2 = ABS(ds/s) * 100
    ELSE
      r = r + 1
      s = s + 1
      iter = 0
    END IF
    IF ea1 ≤ es AND ea2 ≤ es OR iter ≥ maxit EXIT
  END DO
  CALL Quadroot(r,s,r1,i1,r2,i2)
  re(n) = r1
  im(n) = i1
  re(n − 1) = r2
  im(n − 1) = i2
  n = n−2
  DO  i = 0, n
    a(i) = b(i + 2)
  END DO
END DO
IF iter < maxit  THEN
   IF n = 2  THEN
      r = −a(1)/a(2)
      s = −a(0)/a(2)
      CALL Quadroot(r,s,r1,i1,r2,i2)
      re(n) = r1
      im(n) = i1
      re(n − 1) = r2
      im(n − 1) = i2
   ELSE
      re(n) = −a(0)/a(1)
      im(n) = 0
   END IF
ELSE
   ier = 1
END IF
END Bairstow
```

(*b*) Algoritmo para as Raízes de um Polinômio Quadrático

```
SUB Quadroot(r,s,r1,i1,r2,i2)
disc = r ^ 2 + 4 * s
IF disc > 0  THEN
   r1 = (r + SQRT(disc))/2
   r2 = (r − SQRT(disc))/2
   i1 = 0
   i2 = 0
ELSE
   r1 = r/2
   r2 = r1
   i1 = SQRT(ABS(disc))/2
   i2 = −i1
END IF
END QuadRoot
```

7.6 OUTROS MÉTODOS

Há outros métodos disponíveis para localizar as raízes de polinômios. O *método de Jenkins-Traub* (Jenkins e Traub, 1970) é usado comumente em bibliotecas de software como o IMSL. Ele é bastante complicado, e um bom ponto para começar a entendê-lo é encontrado em Ralston e Rabinowitz (1978).

O *método de Laguerre*, que aproxima tanto raízes reais quanto complexas e tem convergência cúbica, está entre as melhores abordagens. Uma discussão completa pode ser encontrada em Householder (1970). Além disso, Press et al. (1992) apresentam um bom algoritmo para implementar esse método.

7.7 LOCALIZAÇÃO DE RAÍZES COM BIBLIOTECAS E PACOTES

As bibliotecas e os pacotes de software têm excelentes recursos para localizar raízes. Nesta seção, será dada uma pequena amostra de alguns dos mais úteis.

7.7.1 Excel

Uma planilha como o Excel pode ser usada para localizar a raiz por tentativa e erro. Por exemplo, se queremos encontrar uma raiz de

$$f(x) = x - \cos x$$

primeiro, entramos um valor de x em uma célula. Então construímos uma outra célula para $f(x)$ que obteria seu valor para o x da primeira célula. Você pode então variar x na célula até que a célula de $f(x)$ se aproxime de zero. Esse processo pode ser mais aprimorado usando-se os recursos gráficos do Excel para obter boas aproximações iniciais (Figura 7.6).

Embora o Excel de fato facilite a abordagem por tentativa e erro, ele também tem duas ferramentas padrão que podem ser usadas na localização de raízes: *Goal Seek* e *Solver*. Ambas as ferramentas podem ser usadas para ajustar sistematicamente as aproximações iniciais. O Goal Seek é usado explicitamente para conduzir uma equação para um valor (no nosso caso, o zero) variando um único parâmetro.

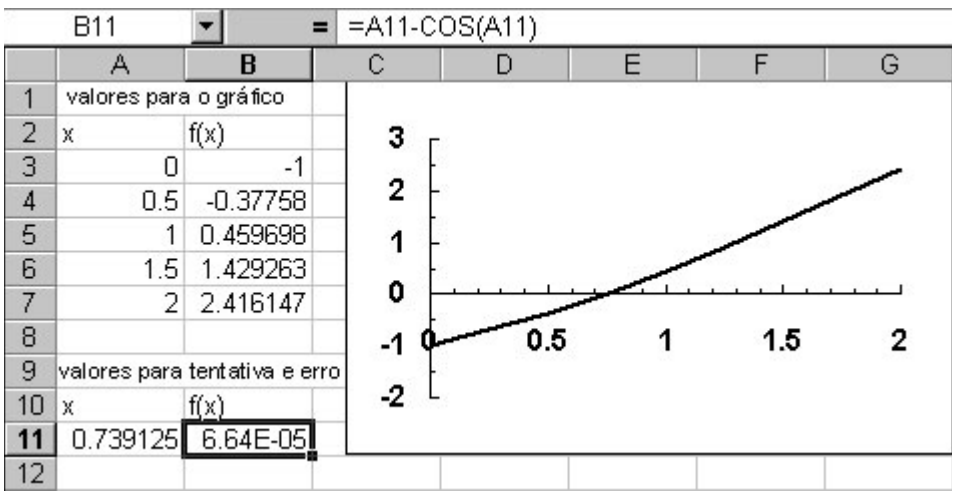


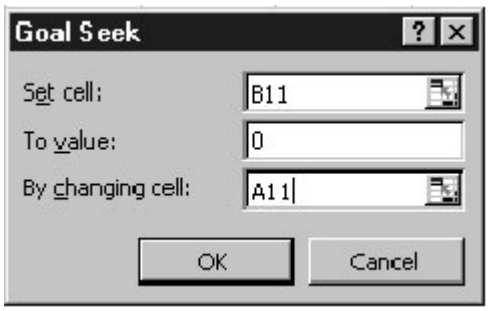
FIGURA 7.6
Uma planilha construída para determinar a raiz de $f(x) = x - \cos x$ por tentativa e erro. O gráfico é usado para obter uma boa aproximação inicial.

EXEMPLO 7.4 Usando a Ferramenta Goal Seek do Excel para Localizar uma Única Raiz

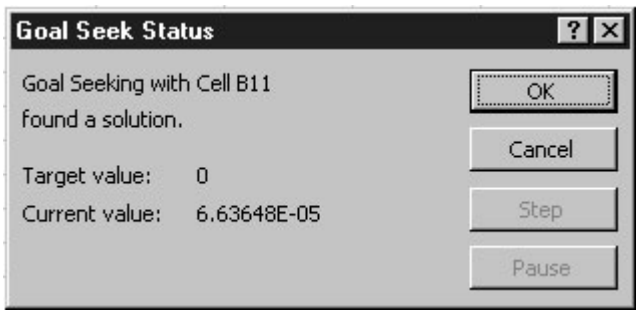
Enunciado do Problema. Use o Goal Seek para determinar a raiz da função transcendental

$$f(x) = x - \cos x$$

Solução. Como na Figura 7.6, a chave para resolver uma única equação com o Excel é criar uma célula para armazenar o valor em questão da função e então considerar o valor dependente em uma outra célula. Uma vez que isso tenha sido feito, a opção Goal Seek é escolhida no menu de ferramentas (Tools). Nesse ponto, um quadro de diálogo será mostrado, pedindo-lhe para impor um valor para uma célula, por meio da mudança de uma outra célula. Por exemplo, suponha que, como na Figura 7.6, sua aproximação foi inserida na célula A11 e o resultado de sua função está na célula B11. O quadro de diálogo do Goal Seek seria preenchido como



Quando a opção OK é escolhida, um quadro de mensagem mostra os resultados,



As células na planilha também seriam modificadas para os novos valores (como mostrado na Figura 7.6).

A ferramenta *Solver* é mais sofisticada do que o Goal Seek nos aspectos: (1) ela pode variar várias células simultaneamente e, (2) ao mesmo tempo em que conduz uma célula-alvo para um valor, ela pode minimizar e maximizar seu valor. O próximo exemplo ilustra como isso pode ser usado para resolver um sistema de equações não-lineares.

EXEMPLO 7.5 Usando o Solver do Excel para um Sistema Não-linear

Enunciado do Problema. Lembre-se de que na Seção 6.5 obtivemos a solução do seguinte conjunto de equações simultâneas

$$u(x, y) = x^2 + xy - 10 = 0$$
$$v(x, y) = y + 3xy^2 - 57 = 0$$

Observe que o par correto de raízes é $x = 2$ e $y = 3$. Use o Solver para determinar as raízes utilizando aproximações iniciais $x = 1$ e $y = 3{,}5$.

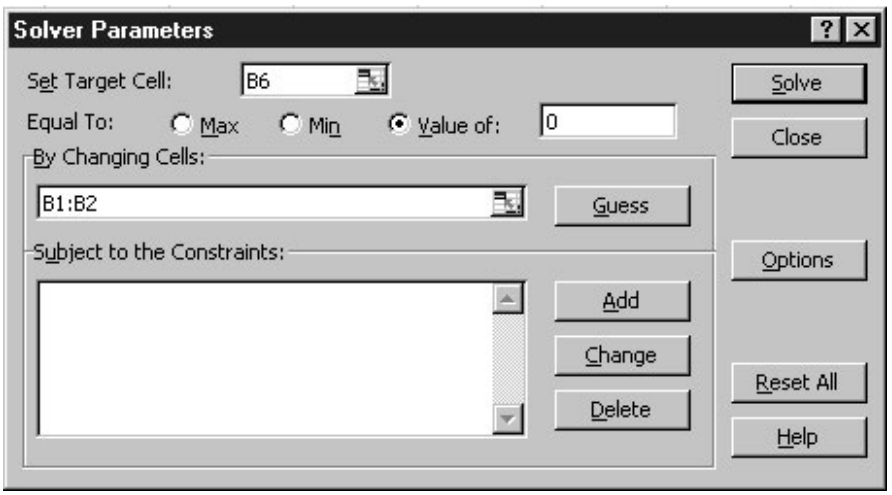
Solução. Como mostrado a seguir, duas células (B1 e B2) podem ser criadas para armazenar as aproximações para x e y. Os valores da própria função, $u(x, y)$ e $v(x, y)$, podem então ser inseridos em duas outras células (B3 e B4). Como pode ser visto, as aproximações iniciais resultam em valores da função que estão longe de zero.

B6	=	=B3^2+B4^2

	A	B	C
1	x	1	
2	y	3.5	
3	u(x, y)	-5.5	
4	v(x, y)	-16.75	
5			
6	soma dos quadrados	310.8125	
7			

A seguir, uma outra célula pode ser criada para armazenar uma única variável que reflita quão próximas de zero ambas as funções estão. Uma forma de fazer isso é somar os quadrados dos valores das funções; em seguida, o resultado é armazenado na célula B6. Se ambas as funções forem nulas, essa função também deveria ser nula. Além disso, o uso do quadrado das funções evita a possibilidade de que ambas as funções pudessem ter o mesmo valor não-nulo, mas com sinais opostos. Nesse caso, a célula-alvo (B6) seria zero, mas as raízes seriam incorretas.

Uma vez que a planilha seja criada, a opção **Solver** é escolhida do menu **Tools**. Nesse ponto, um quadro de diálogos será mostrado, pedindo-lhe as informações pertinentes. As células pertinentes do quadro de diálogo do Solver seriam preenchidas como



Quando a opção OK for escolhida, um quadro de diálogo será aberto com um relatório sobre o sucesso da operação. No caso presente, o Solver obtém a solução correta:

	A	B	C	D
1	x	2.00003		
2	y	2.999984		
3	u(x, y)	0.000176		
4	v(x, y)	0.000202		
5				
6	soma dos quadrados	7.19E-08		
7				

Deve-se observar que o Solver pode falhar. Seu sucesso depende (1) do condicionamento do sistema de equações e/ou (2) da qualidade das aproximações iniciais. Logo, a saída bem-sucedida do exemplo anterior não é garantida. Apesar disso, consideramos o Solver suficientemente útil para torná-lo uma opção possível para obter rapidamente raízes em uma ampla variedade de aplicações em engenharia.

7.7.2 MATLAB

Como resumido na Tabela 7.1, o software MATLAB é capaz de localizar raízes de uma única equação algébrica ou transcendental. Ele é excelente na manipulação e determinação de raízes de polinômios.

A função *fzero* foi desenvolvida para localizar uma raiz de uma única equação. Uma representação simplificada de sua sintaxe é

```
fzero(f,x0,options)
```

onde f é a função que você está analisando, x_0 é a aproximação inicial, e *options* são os parâmetros de otimização (estes são mudados usando-se a função *optimset*). Se *options* for omitido, serão usados valores padrão. Observe que podem ser utilizadas uma ou duas aproximações. Se forem usadas duas aproximações, é suposto que elas delimitem uma raiz. O exemplo a seguir ilustra como fzero pode ser usado.

TABELA 7.1 Funções comuns no MATLAB relacionadas com a localização de raízes e manipulação de polinômios.

Função	Descrição
fzero	Raiz de uma única função.
roots	Encontra raízes de polinômios.
poly	Constrói polinômios com raízes especificadas.
polyval	Calcula valores de polinômios.
polyvalm	Calcula valores de polinômios com argumentos matriciais.
residue	Expansão em frações parciais (resíduos).
polyder	Deriva polinômios.
conv	Multiplica polinômios.
deconv	Divide polinômios.

EXEMPLO 7.6 Usando o MATLAB para a Localização de Raízes

Enunciado do Problema. Use a função fzero do MATLAB para encontrar as raízes de

$$f(x) = x^{10} - 1$$

no intervalo entre $x_l = 0$ e $x_u = 4$. Obviamente, ocorrem duas raízes, em -1 e 1. Lembre-se de que, no Exemplo 5.6, usamos o método da falsa posição com aproximações iniciais de 0 e 1,3 para determinar a raiz positiva.

Solução. Usando as mesmas condições iniciais do Exemplo 5.6, podemos utilizar o MATLAB para determinar a raiz, como em

```
>> x0=[0 1.3];
>> x=fzero(inline('x^10-1'),x0)

x =
     1
```

De modo análogo, podemos usar as aproximações iniciais de –1,3 e 0 para determinar a raiz negativa,

```
>> x0=[-1.3 0];
>> x=fzero(inline('x^10-1'),x0)

x =
    -1
```

Pode-se utilizar também uma única aproximação. Um caso interessante seria usar uma aproximação inicial 0,

```
>> x0=0;
>> x=fzero(inline('x^10-1'),x0)

x =
    -1
```

Assim, para essa aproximação, acontece que o algoritmo subjacente converge para a raiz negativa.

O uso do optimset pode ser ilustrado utilizando-o para mostrar as próprias iterações à medida que a solução progride:

```
>> x0=0;
>> option=optimset('DISP','ITER');
>> x=fzero(inline('x^10-1'),x0,option)

Func-count        x              f(x)          Procedure
    1             0               -1          initial
    2     -0.0282843               -1          search
    3      0.0282843               -1          search
    4          -0.04               -1          search
    •
    •
    •
   21           0.64         -0.988471          search
   22       -0.905097         -0.631065          search
   23        0.905097         -0.631065          search
   24           -1.28           10.8059          search

  Looking for a zero in the interval [-1.28], 0.9051]

   25        0.784528         -0.911674       interpolation
   26       -0.247736         -0.999999       bisection
   27       -0.763868         -0.932363       bisection
   28        -1.02193          0.242305       bisection
   29       -0.968701          -0.27239       interpolation
   30       -0.996873        -0.0308299       interpolation
   31       -0.999702       -0.00297526       interpolation
   32              -1      5.53132e-006       interpolation
   33              -1     -7.41965e-009       interpolation
   34              -1     -1.88738e-014       interpolation
   35              -1                 0       interpolation
Zero found in the interval: [-1.28, 0.9051].

x =
    -1
```

Esses resultados ilustram a estratégia usada pelo fzero quando lhe é fornecida uma única aproximação. Inicialmente, ele procura na vizinhança da aproximação até detectar uma mudança de sinal. Então, usa uma combinação de bissecção e interpolação para procurar a raiz. A interpolação envolve tanto o método da secante quanto a interpolação quadrática inversa (lembre-se da Seção 7.4). Deve-se observar que o algoritmo fzero é muito mais complexo do que essa descrição básica poderia indicar. Você pode consultar Press et al. (1992) para detalhes adicionais.

EXEMPLO 7.7 Usando o MATLAB para Manipular e Determinar as Raízes de Polinômios

Enunciado do Problema. Investigue como o MATLAB pode ser usado para manipular e determinar as raízes de polinômios. Use a seguinte equação do Exemplo 7.3,

$$f_5(x) = x^5 - 3{,}5x^4 + 2{,}75x^3 + 2{,}125x^2 - 3{,}875x + 1{,}25 \qquad \text{(E7.7.1)}$$

que tem três raízes reais: 0,5, $-1{,}0$ e 2, e um par de raízes complexas: $-1 \pm 0.5i$.

Solução. Os polinômios são inseridos no MATLAB armazenando-se seus coeficientes como um vetor. Por exemplo, quando o MATLAB solicitar (≫), digitar e inserir a seguinte linha armazena os coeficientes em um vetor **a**,

```
>> a=[1 -3,5 2,75 2,125 -3,875 1,25];
```

Pode-se então prosseguir para manipular o polinômio. Por exemplo, é possível calculá-lo em $x = 1$ digitando

```
>> polyval(a,1)
```

com o resultado $1(1)^5 - 3{,}5(1)^4 + 2{,}75(1)^3 + 2{,}125(1)^2 - 3{,}875(1) + 1{,}25 = -0{,}25$,

```
ans =
   -0,2500
```

Pode-se calcular a derivada $f'(x) = 5x^4 - 14x^3 + 8{,}25x^2 + 4{,}25x - 3{,}875$ por

```
>> polyder(a)
ans =
    5,0000    -14,0000    8,2500    4,2500    -3,8750
```

A seguir, vamos criar um polinômio quadrático que tenha raízes correspondentes a duas das raízes originais da Equação (E7.7.1): 0,5 e -1. Esse polinômio quadrático é $(x - 0{,}5)(x + 1) = x^2 + 0{,}5x - 0{,}5$ e pode ser inserido no MATLAB como o vetor **b**,

```
>> b=[1 0,5 -0,5];
```

Divide-se então o polinômio original por esse polinômio por meio de

```
>> [d,e]=deconv(a,b)
```

com o resultado sendo um quociente (um polinômio de grau três **d**) e um resto (**e**),

```
d =
    1,0000    -4,0000    5,2500    -2,5000
e =
     0      0      0      0      0      0
```

Como o polinômio é um divisor perfeito, o polinômio resto tem coeficientes nulos. Agora, as raízes do polinômio quociente podem ser determinadas por

```
>> roots(d)
```

com o resultado esperado de que as raízes restantes do polinômio original (E7.7.1) são encontradas,

```
ans =
   2,0000
   1,0000 + 0,5000i
   1,0000 - 0,5000i
```

Pode-se agora multiplicar **d** por **b** para reencontrar o polinômio original,

```
>> conv(d,b)
ans =
    1,0000   -3,5000   2,7500   2,1250   -3,8750   1,2500
```

Finalmente, determinam-se todas as raízes do polinômio original por

```
>> r=roots(a)
r =
  -1,0000
   2,0000
   1,0000 + 0,5000i
   1,0000 — 0,5000i
   0,5000
```

7.7.3 IMSL

O IMSL tem diversas sub-rotinas para determinar raízes de equações (Tabela 7.2). Na discussão presente, vamo-nos concentrar na rotina ZREAL. Essa rotina localiza os zeros reais de uma função real usando o método de Müller.

O ZREAL é implementado pelo seguinte comando CALL,

```
CALL ZREAL(F, ERABS, ERREL, EPS, ETA, NROOT, IMAX, X0, X, INFO)
```

TABELA 7.2 Rotinas IMSL para localizar raízes.

Categoria	Rotina	Capacidade
Raízes de uma função		
	ZREAL	Encontra os zeros reais de uma função real usando o método de Müller.
	ZBREN	Encontra um zero de uma função real que muda de sinal em um intervalo dado.
	ZANLY	Encontra os zeros de uma função complexa univalorizada usando o método de Müller.
Raízes de um sistema de equações		
	NEQNF	Resolve um sistema de equações não-lineares usando um algoritmo híbrido modificado de Powell (uma variação do método de Newton) e uma aproximação do jacobiano por diferença finita.
	NEQNJ	Resolve um sistema de equações não-lineares usando um algoritmo híbrido modificado de Powell (uma variação do método de Newton) com um jacobiano fornecido pelo usuário.
	NEQBF	Resolve um sistema de equações não-lineares usando atualizações fatoradas da secante e uma aproximação do jacobiano por diferença finita.
	NEQBJ	Resolve um sistema de equações não-lineares usando atualizações fatoradas da secante com um jacobiano fornecido pelo usuário.
Raízes de polinômios		
	ZPORC	Encontra os zeros de polinômios com coeficientes reais com o algoritmo de Jenkins-Traub.
	ZPLRC	Encontra os zeros de polinômios com coeficientes reais com o método de Laguerre.
	ZPOCC	Encontra os zeros de polinômios com coeficientes complexos com o algoritmo de Jenkins-Traub.

onde

F = Função definida pelo usuário, para a qual os zeros devem ser encontrados.

ERABS = Primeiro critério de parada, termina se $|f(x_i)| <$ ERABS. (Entrada)

ERREL = Segundo critério de parada; termina se $|(x_i - x_{i-1})/x_i| <$ ERREL. (Entrada)

EPS = Ver ETA. (Entrada)

ETA = Critério de dispersão para raízes múltiplas. (Entrada)

Se o zero, x_i, foi calculado e $|x_i - x_j| <$ EPS, onde x_j é um zero calculado anteriormente, o cálculo é reiniciado com uma nova aproximação $x_i +$ ETA.

NROOT = Número de raízes a serem encontradas. (Entrada)

IMAX = Número máximo de iterações por zero. (Entrada)

X0 = Vetor de comprimento NROOT contendo as aproximações iniciais. (Entrada)

X = Vetor de comprimento NROOT contendo os zeros calculados. (Saída)

INFO = Vetor inteiro de comprimento NROOT. (Saída)

Contém o número de iterações para encontrar cada raiz.

Observe que a iteração termina quando qualquer um dos critérios de parada for satisfeito ou quando o número máximo de iteração for atingido. A função F tem a forma geral

```
FUNCTION F(X)
REAL F,X
F = ...
END
```

em que a linha "F = …" é onde a função da variável desconhecida X é escrita.

EXEMPLO 7.8 Usando o IMSL para Localizar uma Única Raiz

Enunciado do Problema. Use o ZREAL para determinar a raiz da função transcendental

$$f(x) = x - \cos x$$

Solução. Um exemplo de um programa principal em Fortran 90 e uma função usando ZREAL para resolver esse problema pode ser escrito como

```
PROGRAM Root
IMPLICIT NONE
INTEGER::nroot
PARAMETER(nroot=1)
INTEGER::itmax=50
REAL::errabs=0.,errrel=1.E-5,eps=0.,eta=0.
REAL::f,x0(nroot),x(nroot)
EXTERNAL f
INTEGER::info(nroot)
PRINT *, "Enter initial guess"
READ *, x0
CALL ZREAL(f,errabs,errrel,eps,eta,nroot,itmax,x0,x,info)
PRINT *, "root = ", x
PRINT *, "iterations = ", info
END PROGRAM
```

```
FUNCTION f(x)
IMPLICIT NONE
REAL::f,x
f = x - cos(x)
END FUNCTION
```

A saída é::

```
Enter initial guess
0.5
root =    7.390851E-01
iterations =           5
```

PROBLEMAS

7.1 Divida o polinômio $f(x) = x^4 - 7{,}5x^3 + 14{,}5x^2 + 3x - 20$ pelo monômio $x - 2$. É verdade que $x = 2$ é uma raiz?

7.2 Divida o polinômio $f(x) = x^5 - 5x^4 + x^3 - 6x^2 - 7x + 10$ pelo monômio $x - 2$.

7.3 Use o método de Müller para determinar a raiz positiva de

(a) $f(x) = x^3 + x^2 - 3x - 5$

(b) $f(x) = x^3 - 0{,}5x^2 + 4x - 3$

7.4 Use o método de Müller ou o MATLAB para determinar as raízes reais e complexas de

(a) $f(x) = x^3 - x^2 + 3x - 2$

(b) $f(x) = 2x^4 + 6x^2 + 10$

(c) $f(x) = x^4 - 2x^3 + 6x^2 - 8x + 8$

7.5 Use o método de Bairstow para determinar as raízes de

(a) $f(x) = -2 + 6{,}2x - 4x^2 + 0{,}7x^3$

(b) $f(x) = 9{,}34 - 21{,}97x + 16{,}3x^2 - 3{,}704x^3$

(c) $f(x) = x^4 - 3x^3 + 5x^2 - x - 10$

7.6 Desenvolva um programa para implementar o método de Müller. Teste-o repetindo o Exemplo 7.2.

7.7 Use o programa desenvolvido no Problema 7.6 para determinar as raízes reais do Problema 7.4**a**. Construa um gráfico (à mão ou com o Excel ou algum outro pacote gráfico) para encontrar aproximações iniciais adequadas.

7.8 Desenvolva um programa para implementar o método de Bairstow. Teste-o repetindo o Exemplo 7.3.

7.9 Use o programa desenvolvido no Problema 7.8 para determinar as raízes das equações no Problema 7.5.

7.10 Determine as raízes reais de $x^{3{,}5} = 80$ com o recurso Goal Seek do Excel ou uma biblioteca ou um pacote de sua escolha.

7.11 A velocidade de um pára-quedista em queda livre é dada por

$$v = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right)$$

onde $g = 9{,}8$ m/s^2. Para um pára-quedista com coeficiente de arrasto $c = 14$ kg/s, calcule a massa m para que a velocidade seja $v = 35$ m/s em $t = 8$ s. Use o recurso Goal Seek do Excel ou uma biblioteca ou um pacote de sua escolha.

7.12 Determine as raízes das equações não-lineares simultâneas

$$y = -x^2 + x + 0{,}75$$
$$y + 5xy = x^2$$

Use aproximações iniciais $x = y = 1{,}2$ e a ferramenta Solver do Excel ou uma biblioteca ou um pacote de sua escolha.

7.13 Determine as raízes das equações não-lineares simultâneas

$$(x - 4)^2 + (y - 4)^2 = 5$$
$$x^2 + y^2 = 16$$

Use uma abordagem gráfica para obter suas aproximações iniciais. Determine estimativas refinadas com a ferramenta Solver do Excel ou uma biblioteca ou um pacote de sua escolha.

7.14 Faça operações no MATLAB idênticas àquelas do Exemplo 7.7 ou use uma biblioteca ou um pacote de sua escolha para encontrar todas as raízes do polinômio

$$f(x) = (x - 4)(x + 2)(x - 1)(x + 5)(x - 7)$$

Observe que a função `poly` pode ser usada para converter as raízes para um polinômio.

7.15 Use o MATLAB ou uma biblioteca ou um pacote de sua escolha para determinar as raízes das equações no Problema 7.5.

7.16 Desenvolva um subprograma para encontrar as raízes de um polinômio usando a rotina IMSL, o ZREAL ou uma biblioteca ou um pacote de sua escolha. Teste-o determinando as raízes das equações dos Problemas 7.4 e 7.5.

7.17 Um cilindro circular bidimensional é colocado em um escoamento uniforme a alta velocidade. São emitidos vórtices do cilindro a uma freqüência constante, e sensores de pressão na superfície de trás do cilindro detectam essa freqüência calculando quão freqüentemente a pressão varia. Dados três pontos, use o método de Müller para encontrar o instante no qual a pressão era nula.

Tempo	0,60	0,62	0,64
Pressão	20	50	60

7.18 Quando tentamos determinar a acidez de uma solução de hidróxido de magnésio em ácido clorídrico, obtemos a seguinte equação

$$A(x) = x^3 + 3{,}5x^2 - 40$$

onde x é a concentração do íon hidrônio. Encontre a concentração do íon hidrônio para uma solução saturada (acidez nula) usando dois métodos diferentes no MATLAB (por exemplo, graficamente e usando a função `roots`).

7.19 Considere o seguinte sistema com três incógnitas a, u, e v:

$$u^2 - 2v^2 = a^2$$
$$u + v = 2$$
$$a^2 - 2a - u = 0$$

Determine os valores reais das incógnitas usando: **(a)** o Solver do Excel e **(b)** um pacote de software de manipulação simbólica.

7.20 Na análise de sistemas de controle, são desenvolvidas funções de transferência que relacionam matematicamente a dinâmica da entrada de um sistema com sua saída. Uma função de transferência para um sistema de posicionamento de um robô é dada por

$$G(s) = \frac{C(s)}{N(s)} = \frac{s^3 + 12{,}5s^2 + 50{,}5s + 66}{s^4 + 19s^3 + 122s^2 + 296s + 192}$$

onde $G(s)$ é o ganho do sistema, $C(s)$ é a saída do sistema, $N(s)$ é a entrada do sistema e s é a transformada de Laplace complexa de freqüência. Use uma técnica numérica para encontrar as raízes do numerador e do denominador e fatore-os na forma

$$G(s) = \frac{(s + a_1)(s + a_2)(s + a_3)}{(s + b_1)(s + b_2)(s + b_3)(s + b_4)}$$

onde a_i e b_i são as raízes do numerador e do denominador, respectivamente.

7.21 Desenvolva uma função M-file para a bissecção de forma análoga à Figura 5.10. Teste a função repetindo os cálculos dos Exemplos 5.3 e 5.4.

7.22 Desenvolva uma função M-file para o método da falsa posição. A estrutura de sua função deveria ser análoga ao algoritmo da bissecção esboçado na Figura 5.10. Teste o programa repetindo o Exemplo 5.5.

7.23 Desenvolva uma função M-file para o método de Newton-Raphson com base na Figura 6.4 e na Seção 6.2.3. Além da aproximação inicial, passe a função e sua derivada como argumentos. Teste a função repetindo os cálculos do Exemplo 6.3.

7.24 Desenvolva uma função M-file para o método da secante baseado na Figura 6.4 e na Seção 6.3.2. Além das duas aproximações iniciais, passe a função como um argumento. Teste a função repetindo os cálculos do Exemplo 6.6.

7.25 Desenvolva uma função M-file para o método da secante modificado baseado na Figura 6.4 e na Seção 6.3.2. Além da aproximação inicial e da perturbação, passe a função como um argumento. Teste a função repetindo os cálculos do Exemplo 6.8.

CAPÍTULO 8

Estudo de Casos: Raízes de Equações

O propósito deste Capítulo é usar os procedimentos numéricos discutidos nos Capítulos 5, 6 e 7 para resolver problemas reais de engenharia. As técnicas numéricas são importantes nas aplicações práticas porque a engenharia freqüentemente encontra problemas que não podem ser abordados usando-se técnicas analíticas. Por exemplo, modelos matemáticos simples que podem ser resolvidos analiticamente nem sempre são aplicáveis quando problemas reais estão envolvidos. Logo, modelos mais complicados precisam ser usados. Nesses casos, é apropriado implementar uma solução numérica com o computador. Em outras situações, problemas de projeto de engenharia exigem soluções para as variáveis implícitas em equações complicadas.

Os estudos de casos a seguir são encontrados rotineiramente nos cursos de graduação mais avançados e nos de pós-graduação. Além disso, são representativos dos problemas tratados profissionalmente. Tais problemas são tirados de quatro ramos importantes da engenharia: química, civil, elétrica e mecânica. Essas aplicações também servem para ilustrar os prós e contras entre as diversas técnicas numéricas.

A primeira aplicação, tirada da engenharia química, fornece um exemplo excelente de como os métodos para localizar raízes permitem usar fórmulas realísticas na prática da engenharia. Além disso, também ilustra como a eficiência da técnica de Newton-Raphson é usada com vantagem quando um grande número de localizações de raízes é necessário.

Os problemas de projeto em engenharia seguintes são tirados da engenharia civil, elétrica e mecânica. A Seção 8.2 usa tanto os métodos intervalares quanto os métodos abertos para determinar a profundidade e a velocidade da água escoando em um canal aberto. A Seção 8.3 mostra como as raízes das equações transcendentais podem ser utilizadas no projeto de um circuito elétrico. As Seções 8.2 e 8.3 também ilustram como os métodos gráficos fornecem percepção do processo de localização de raízes. Finalmente, a Seção 8.4 usa a localização de raízes de polinômios para analisar a vibração de um automóvel.

8.1 LEIS DOS GASES IDEAIS E NÃO-IDEAIS (ENGENHARIA QUÍMICA/BIOQUÍMICA)

Fundamentos. A *lei dos gases ideais* é dada por

$$pV = nRT \tag{8.1}$$

onde p é a pressão absoluta, V é o volume, n é o número de moles, R é a constante universal dos gases e T é a temperatura absoluta. Embora essa equação seja amplamente usada pelos engenheiros e cientistas, ela é acurada apenas sobre um intervalo limitado de pressão e temperatura. Além disso, a Equação (8.1) é mais apropriada para alguns gases do que para outros.

Uma equação de estado alternativa para os gases é dada por

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT \tag{8.2}$$

conhecida como *equação de van der Waals*, onde $v = V/n$ é o volume molar e a e b são constantes empíricas que dependem do gás particular.

Um projeto de engenharia química exige que se faça uma estimativa acurada do volume (v) tanto do dióxido de carbono como do oxigênio para diversas temperaturas diferentes e diversas combinações de pressão de modo que recipientes de armazenamento adequados possam ser escolhidos. Também é de interesse examinar quão bem cada gás obedece à lei dos gases ideais comparando os volumes molares calculados pelas Equações (8.1) e (8.2). São fornecidos os seguintes dados:

$$R = 0{,}082054 \text{ L atm/(mol K)}$$

$$\left.\begin{aligned} a &= 3{,}592 \\ b &= 0{,}04267 \end{aligned}\right\} \text{ dióxido de carbono}$$

$$\left.\begin{aligned} a &= 1{,}360 \\ b &= 0{,}03183 \end{aligned}\right\} \text{ oxigênio}$$

As pressões de interesse para o projeto são 1, 10 e 100 atm combinadas com temperaturas de 300, 500 e 700 K.

Solução. Os volumes molares para ambos os gases são calculados usando a lei dos gases ideais, com $n = 1$. Por exemplo, se $p = 1$ atm e $T = 300$ K,

$$v = \frac{V}{n} = \frac{RT}{p} = 0{,}082054 \frac{\text{L atm}}{\text{mol K}} \frac{300 \text{ K}}{1 \text{ atm}} = 24{,}6162 \text{ L/mol}$$

Esses cálculos são repetidos para todas as combinações de temperatura e pressão apresentadas na Tabela 8.1.

Os cálculos do volume molar a partir da equação de van der Waals podem ser feitos usando qualquer um dos métodos numéricos para encontrar raízes de equações discutidos nos Capítulos 5, 6 e 7, com

$$f(v) = \left(p + \frac{a}{v^2}\right)(v - b) - RT \tag{8.3}$$

Nesse caso, a derivada de $f(v)$ é fácil de determinar e o método de Newton-Raphson é implementado de forma conveniente e eficiente. A derivada de $f(v)$ com relação a v é dada por

$$f'(v) = p - \frac{a}{v^2} + \frac{2ab}{v^3} \tag{8.4}$$

O método de Newton-Raphson é descrito pela Equação (6.6):

$$v_{i+1} = v_i - \frac{f(v_i)}{f'(v_i)}$$

TABELA 8.1 Cálculos do volume molar.

Temperatura, K	Pressão, atm	Volume Molar (Lei dos Gases Ideais), L/mol	Volume Molar (van der Waals) Dióxido de Carbono, L/mol	Volume Molar (van der Waals) Oxigênio, L/mol
300	1	24,6162	24,5126	24,5928
	10	2,4616	2,3545	2,4384
	100	0,2462	0,0795	0,2264
500	1	41,0270	40,9821	41,0259
	10	4,1027	4,0578	4,1016
	100	0,4103	0,3663	0,4116
700	1	57,4378	57,4179	57,4460
	10	5,7438	5,7242	5,7521
	100	0,5744	0,5575	0,5842

que pode ser utilizada para fazer uma estimativa da raiz. Por exemplo, usando a aproximação inicial 24,6162, o volume molar do dióxido de carbono a 300 K e 1 atm é calculado como sendo 24,5126 L/mol. Esse resultado foi obtido depois de apenas duas iterações e tem um ε_a de menos que 0,001%.

Cálculos análogos para todas as combinações de pressão e temperatura para ambos os gases estão apresentados na Tabela 8.1. Vê-se que os resultados da lei dos gases ideais são diferentes daqueles provenientes da equação de van der Waals para ambos os gases, dependendo dos valores específicos de p e T. Além disso, como alguns dos resultados são significativamente diferentes, seus projetos de recipientes de armazenamento seriam bem diferentes, dependendo de qual equação de estado fosse usada.

Nesse caso, uma equação de estado complicada foi examinada usando o método de Newton-Raphson. Os resultados diferem significativamente da lei dos gases ideais em diversos casos. Do ponto de vista prático, o método de Newton-Raphson foi apropriado para essa aplicação, pois $f'(v)$ era fácil de calcular. Portanto, as propriedades de convergência rápida do método de Newton-Raphson puderam ser exploradas.

Além de demonstrar seu poder para um único cálculo, o problema de projeto do qual se está tratando também ilustra como o método de Newton-Raphson é especialmente atraente quando são necessários diversos cálculos. Por causa da velocidade dos computadores digitais, as eficiências dos diversos métodos numéricos para a maioria das raízes de equações são indistinguíveis para um único cálculo. Mesmo uma diferença de 1 s entre a abordagem grosseira por bissecção e a eficiência do método de Newton-Raphson não corresponde a uma perda significativa de tempo quando apenas um cálculo está sendo feito. Entretanto, suponha que milhões de cálculos de raízes são necessários para resolver um problema. Nesse caso, a eficiência do método pode ser um fator decisivo na escolha da técnica.

Por exemplo, suponha que lhe tenha sido pedido para projetar um sistema de controle computadorizado automático para um processo de produção química. O sistema precisa de estimativas acuradas do volume molar em uma base essencialmente contínua para fabricar adequadamente o produto final. São instalados medidores que fornecem leituras instantâneas da pressão e da temperatura. Os cálculos de v devem ser obtidos para uma variedade de gases usados no processo.

Para tais aplicações, os métodos intervalares, como a bissecção ou o método da falsa posição, seriam, provavelmente, muito lentos. Além disso, as duas aproximações iniciais necessárias nessas abordagens também poderiam interpor um atraso crítico no processo. Essas deficiências também são relevantes no método da secante, que, de modo similar, necessita de duas estimativas iniciais.

Em contraste, o método de Newton-Raphson necessita de apenas uma aproximação para a raiz. A lei dos gases ideais poderia ser usada para obter essa aproximação no início do processo. Então, supondo que a escala de tempo seja suficientemente pequena para que a pressão e a temperatura não variem muito entre os cálculos, a solução para a raiz anterior serviria como uma boa aproximação para a próxima aplicação. Portanto, a boa aproximação que, em geral, é um pré-requisito para a convergência do método de Newton-Raphson estaria automaticamente disponível. Todas essas considerações favoreceriam muito a técnica de Newton-Raphson para tais problemas.

8.2 ESCOAMENTO DE CANAL ABERTO (ENGENHARIA CIVIL/AMBIENTAL)

Fundamentos. A engenharia civil é um amplo campo que inclui áreas tão diversas quanto as engenharias estrutural, geotécnica, de transporte, ambiental e de recursos hídricos. As duas últimas especialidades tratam tanto da poluição quanto do fornecimento da água e, portanto, fazem um extenso uso da ciência da mecânica dos fluidos.

Um problema geral está relacionado com o escoamento de água em canais abertos como rios e canais. A vazão, que é rotineiramente medida na maioria dos principais rios e córregos, é definida como o volume de água passando por um determinado ponto em um canal por unidade de tempo, Q (m^3/s).

Embora a vazão seja uma quantidade útil, uma outra questão se relaciona com o que acontece quando se põe uma vazão específica em um canal inclinado (Figura 8.1). Na realidade, acontecem duas coisas: a água atinge uma profundidade específica H (m) e se move a uma velocidade específica U (m/s). Os engenheiros ambientalistas poderiam estar interessados em saber essas quantidades para prever o transporte e o destino de poluentes em um rio. Assim, a questão geral é: se for dada a vazão para um canal, como calcular a profundidade e a velocidade?

Solução. A relação mais fundamental entre o escoamento e a profundidade é a *equação da continuidade*:

$$Q = UA_c \tag{8.5}$$

onde A_c é a área da seção transversal do canal (m^2). Dependendo da forma do canal, a área pode ser relacionada com a profundidade por alguma relação funcional. Para o canal retangular descrito na Figura 8.1, $A_c = BH$. Substituindo essa relação na Equação (8.5), obtém-se

$$Q = UBH \tag{8.6}$$

onde B é a largura (m). Deve-se observar que a equação da continuidade é deduzida a partir da *conservação da massa* (lembre-se da Tabela 1.1).

Agora, embora a Equação (8.6) certamente relacione os parâmetros do canal, não é suficiente para responder à questão. Supondo que B tenha sido especificado, há uma equação para duas incógnitas (U e H). Portanto, é necessária uma equação adicional. Para escoamentos uniformes (significando que o escoamento não varia nem no espaço nem no tempo), o engenheiro irlandês Robert Manning propôs a seguinte fórmula semi-empírica (apropriadamente chamada de equação de Manning)

$$U = \frac{1}{n} R^{2/3} S^{1/2} \tag{8.7}$$

onde n é o coeficiente de rugosidade de Manning (um número adimensional usado para parametrizar o atrito no canal), S é a inclinação do canal (adimensional, queda em metros por metro de comprimento) e R é o raio hidráulico (m), o qual se relaciona com os parâmetros mais fundamentais por

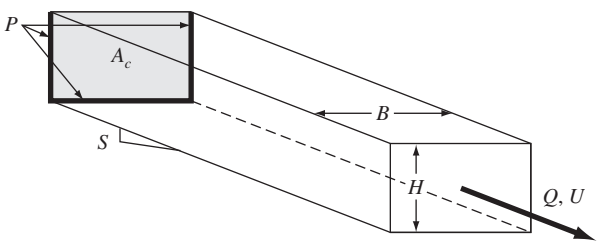
$$R = \frac{A_c}{P} \tag{8.8}$$

onde P é o perímetro molhado (m). Como o nome indica, o perímetro molhado é o comprimento dos lados e do fundo do canal que estão submersos. Por exemplo, para um canal retangular, ele é definido por

$$P = B + 2H \tag{8.9}$$

Deve-se observar que, do mesmo modo como a equação da continuidade é deduzida da conservação da massa, a equação de Manning é uma expressão da *conservação do momento*. Em particular, ela indica como a velocidade depende da rugosidade, uma manifestação do atrito.

FIGURA 8.1



Embora o sistema de equações não-lineares (8.6 e 8.7) possa ser resolvido simultaneamente (por exemplo, usando a abordagem multidimensional de Newton-Raphson descrita na Seção 6.5.2), uma abordagem mais simples seria combinar essas equações. A Equação (8.7) pode ser substituída na Equação (8.6) para fornecer

$$Q = \frac{BH}{n} R^{2/3} S^{1/2} \tag{8.10}$$

Então, o raio hidráulico, Equação (8.8), juntamente com as diversas relações para o canal retangular podem ser substituídos,

$$Q = \frac{S^{1/2}}{n} \frac{(BH)^{5/3}}{(B + 2H)^{2/3}} \tag{8.11}$$

Portanto, a equação agora contém uma única incógnita H juntamente com os valores dados para Q e para os parâmetros do canal (n, S e B).

Embora se tenha uma equação com uma incógnita, é impossível resolvê-la explicitamente para determinar H. Entretanto, a profundidade pode ser determinada numericamente reformulando-se a equação como um problema de raiz

$$f(H) = \frac{S^{1/2}}{n} \frac{(BH)^{5/3}}{(B + 2H)^{2/3}} - Q = 0 \tag{8.12}$$

A Equação (8.12) pode ser resolvida prontamente com qualquer um dos métodos para localização de raízes descritos nos Capítulos 5 e 6. Por exemplo, se $Q = 5$ m^3/s, $B = 20$ m, $n = 0{,}03$, e $S = 0{,}0002$, a equação é

$$f(H) = 0{,}471405 \frac{(20H)^{5/3}}{(20 + 2H)^{2/3}} - 5 = 0 \tag{8.13}$$

Ela pode ser resolvida, fornecendo $H = 0{,}7023$m. O resultado é verificado por substituição na Equação (8.13) para fornecer

$$f(H) = 0{,}471405 \frac{(20 \times 0{,}7023)^{5/3}}{(20 + 2 \times 0{,}7023)^{2/3}} - 5 = 7{,}8 \times 10^{-5} \tag{8.14}$$

que está bastante próximo de zero.

A outra incógnita, a velocidade, pode agora ser determinada substituindo-se novamente na Equação (8.6),

$$U = \frac{Q}{BH} = \frac{5}{20(0{,}7023)} = 0{,}356 \text{ m/s} \tag{8.15}$$

Portanto, tivemos sucesso em resolver as equações, determinando a profundidade e a velocidade.

A seguir, a abordagem se aprofundará um pouco mais nos aspectos numéricos desse problema. Uma questão pertinente seria: como encontrar boas aproximações iniciais para o método numérico? A resposta depende do tipo de método.

Para os métodos intervalares como a bissecção e a falsa posição, uma abordagem seria determinar se é possível encontrar estimativas inferiores e superiores que sempre delimitem uma única raiz. Uma abordagem conservadora seria escolher zero como aproximação inferior. Então, se for conhecida a máxima profundidade que poderia ocorrer, esse valor servirá como aproximação superior. Por exemplo, todos os rios do mundo, exceto os maiores, têm menos de 10 m de profundidade. Portanto, pode-se escolher 0 e 10 como limitantes para H.

Se $Q > 0$ e $H = 0$, a Equação (8.12) será sempre negativa para a aproximação inferior. Conforme H aumenta, a Equação (8.12) aumentará monotonamente e eventualmente se tornará positiva. Portanto, as aproximações devem delimitar uma única raiz para a maioria dos casos confrontados rotineiramente nos rios e córregos naturais.

Agora, uma técnica como a bissecção encontraria a raiz de forma muito confiável. Mas qual o preço pago? Usando-se um intervalo tão grande e uma técnica como a bissecção, o número de iterações para atingir a precisão desejada poderia ser computa-

cionalmente excessivo. Por exemplo, se fosse escolhida uma tolerância de 0,001 m, a Equação (5.5) poderia ser usada para calcular

$$n = \frac{\log(10/0{,}001)}{\log 2} = 13{,}3$$

Portanto, seriam necessárias 14 iterações. Embora isso certamente não seja caro para um único cálculo, poderia tornar-se exorbitante se fossem feitos muitos cálculos como esse. A alternativa seria estreitar o intervalo inicial (com base no conhecimento específico do sistema), mudar para uma técnica intervalar mais eficiente (como a falsa posição), ou aceitar uma precisão mais grosseira.

Uma outra forma de obter melhor eficiência seria usar um método aberto como os métodos de Newton-Raphson ou da secante. É claro que, nesses casos, a questão das aproximações iniciais é complicada pela questão de convergência.

Pode-se conseguir alguma intuição sobre essas questões examinando a menos eficiente das abordagens abertas — a iteração de ponto fixo. Examinando a Equação (8.11), há duas formas diretas de isolar H, isto é, pode-se isolar H no numerador,

$$H = \frac{(Qn)^{3/5}(B + 2H)^{2/5}}{BS^{3/10}} \tag{8.16}$$

ou no denominador

$$H = \frac{1}{2}\left[\frac{S^3(BH)^{5/2}}{(Qn)^{3/2}} - B\right] \tag{8.17}$$

Esse é o ponto em que uma argumentação física pode ser útil. Para a maioria dos rios e córregos, a largura é muito maior do que a profundidade. Portanto, a quantidade $B + 2H$ não deve variar muito. De fato, ela deveria ser aproximadamente igual a B. Em comparação, BH é diretamente proporcional a H. Conseqüentemente, a Equação (8.16) deveria convergir mais rapidamente para a raiz. Isso pode ser verificado substituindo as aproximações $H = 0$ e 10 em ambas as equações. Para a Equação (8.16), os resultados são 0,6834 e 0,9012, que estão ambos próximos do valor verdadeiro de 0,7023. Em contraste, os resultados para a Equação (8.17) são -10 e 8,178, que estão claramente distantes da raiz.

A superioridade da Equação (8.16) é ainda mais confirmada pelos gráficos das componentes (lembre-se da Figura 6.3). Como na Figura 8.2, a componente $g(H)$ da Equação (8.16) é quase horizontal. Portanto, ela não só convergirá, como o fará rapidamente. Em contraste, a componente $g(H)$ da Equação (8.17) é quase vertical, implicando uma divergência forte e rápida.

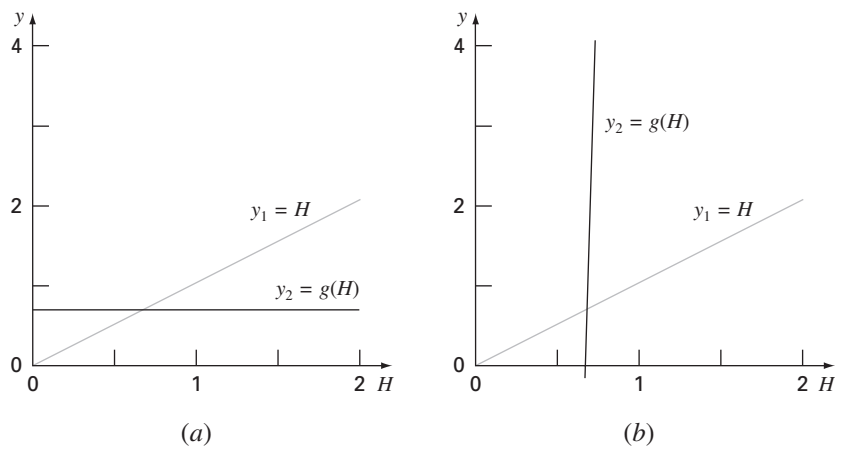


FIGURA 8.2
Gráficos das componentes para dois casos de iteração de ponto fixo, uma que convergirá [(*a*), Equação (8.16)] e uma que divergirá [(*b*) Equação(8.17)].

Há duas formas práticas de tirar vantagem de tal análise:

1. Na eventualidade de um método aberto mais refinado ser usado, a Equação (8.16) fornece uma maneira de obter uma excelente aproximação inicial. Por exemplo, se H for escolhido como zero, a Equação (8.12)
se tornará

$$H_0 = \frac{(Qn/B)^{3/5}}{S^{3/10}}$$

onde H_0 seria o valor inicial usado no método de Newton-Raphson ou da secante.

2. Foi mostrado que a iteração de ponto fixo fornece uma opção viável para esse problema particular. Por exemplo, usando uma aproximação inicial de $H = 0$, a Equação (8.16) atingiria seis dígitos de precisão em quatro iterações no caso que está sendo examinado. Uma situação na qual a fórmula de ponto fixo poderia vir a calhar seria em uma aplicação de planilha. Isto é, as planilhas são ideais para uma fórmula iterativa convergente que dependa de uma única célula.

8.3 PROJETO DE UM CIRCUITO ELÉTRICO (ENGENHARIA ELÉTRICA)

Fundamentos. Os engenheiros elétricos geralmente usam as leis de Kirchhoff para estudar o comportamento estacionário (que não varia com o tempo) de circuitos elétricos. Tal comportamento estacionário será examinado na Seção 12.3. Um outro problema importante envolve os circuitos que são transientes por natureza e em que ocorrem variações temporais súbitas. Tal situação ocorre depois que a chave na Figura 8.3 for fechada. Nesse caso, ocorrerá um período de ajuste após o fechamento da chave, até que um novo estado estacionário seja atingido. A duração desse período de ajustamento está intimamente ligada às propriedades de armazenamento do capacitor e do indutor. O armazenamento do circuito irá dissipar o módulo das oscilações.

O fluxo de corrente através do resistor causa uma queda de voltagem (V_R) dada por

$$V_R = iR$$

onde i é a corrente e R é a resistência do resistor. Quando R e i têm unidades de ohms e ampères, respectivamente, V_R tem unidades de volts.

Analogamente, um indutor resiste a variações na corrente, de modo que a queda de voltagem V_L através dele é

$$V_L = L\frac{di}{dt}$$

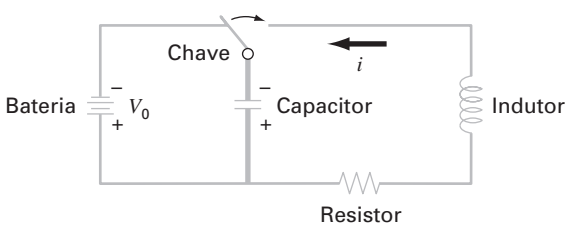
onde L é a indutância. Quando L e i têm unidades de henrys e amperes, respectivamente, V_L tem unidades de volts e t tem unidades de segundos.

A queda de voltagem no capacitor (V_C) depende da carga (q) nele:

$$V_C = \frac{q}{C}$$

onde C é a capacitância. Quando a carga estiver expressa em unidades de coulombs, a unidade de C é farad.

FIGURA 8.3
Um circuito elétrico. Quando a chave é fechada, a corrente sofre uma série de oscilações até que um novo estado estacionário seja atingido.



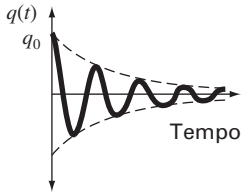


FIGURA 8.4
A carga em um capacitor como uma função do tempo logo após a chave da Figura 8.3 ter sido fechada.

A segunda lei de Kirchhoff afirma que a soma algébrica das quedas de voltagem em torno de um circuito fechado é zero. Depois que a chave é fechada, tem-se

$$L\frac{di}{dt} + Ri + \frac{q}{C} = 0$$

Entretanto, a corrente está relacionada com a carga por

$$i = \frac{dq}{dt}$$

Portanto,

$$L\frac{d^2q}{dt^2} + R\frac{dq}{dt} + \frac{1}{C}q = 0 \tag{8.18}$$

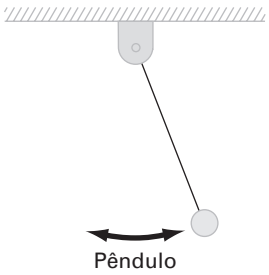
Essa é uma equação diferencial ordinária linear de segunda ordem que pode ser resolvida usando-se os métodos do cálculo (ver Seção 8.4). A solução é dada por

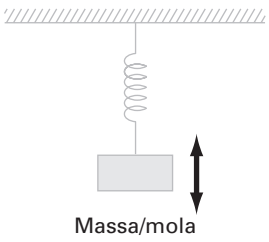
$$q(t) = q_0 e^{-Rt/(2L)} \cos\left[\sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}\, t\right] \tag{8.19}$$

na qual $t = 0$, $q = q_0 = V_0C$ e V_0 é a voltagem fornecida pela bateria. A Equação (8.19) descreve a variação no tempo da carga no capacitor. A solução $q(t)$ está traçada na Figura 8.4.

Um problema de projeto típico em engenharia elétrica poderia envolver a determinação do resistor apropriado para dissipar energia a uma taxa especificada, com valores conhecidos para L e C. Para esse problema, suponha que a carga deva ser dissipada a 1% de seu valor original ($q/q_0 = 0{,}01$) em $t = 0{,}05$ s, com $L = 5$ H e $C = 10^{-4}$ F.

Solução. É necessário resolver a Equação (8.19) para determinar R, com valores conhecidos de q, q_0, L e C. Entretanto, deve ser usada uma técnica de aproximação numérica, já que R é uma variável implícita na Equação (8.19). O método da bissecção será usado para esse propósito. Os outros métodos discutidos nos Capítulos 5 e 6 também





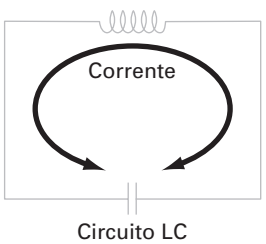
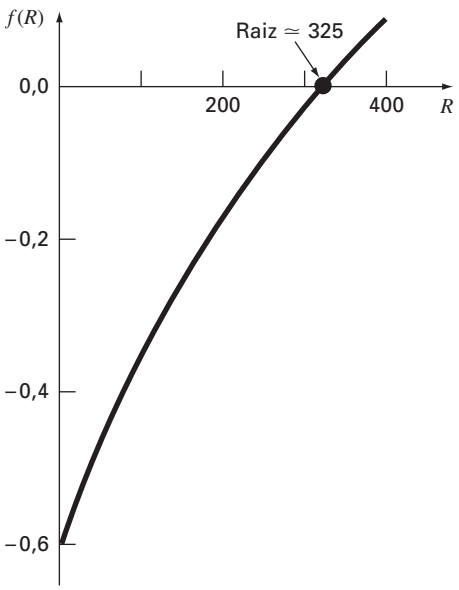


FIGURA 8.6
Três exemplos de osciladores harmônicos. A flecha de duas pontas ilustra as oscilações para cada sistema.

FIGURA 8.5
Gráfico da Equação (8.20) usada para obter aproximações iniciais para R que cercam a raiz.



são apropriados, embora o método de Newton-Raphson possa ser considerado inconveniente pelo fato de a derivada da Equação (8.19) ser um pouco trabalhosa. Reorganizando a Equação (8.19),

$$f(R) = e^{-Rt/(2L)} \cos\left[\sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2}\, t\right] - \frac{q}{q_0}$$

ou, usando os valores numéricos dados,

$$f(R) = e^{-0{,}005R} \cos[\sqrt{2000 - 0{,}01R^2}\ (0{,}05)] - 0{,}01 \tag{8.20}$$

O exame dessa equação sugere que um intervalo inicial razoável para R é de 0 a 400 Ω (pois $2000 - 0.01R^2$ deve ser maior que zero). A Figura 8.5, um gráfico da Equação (8.20), confirma isso. Vinte e uma iterações do método da bissecção dão $R = 328{,}1515$ Ω, com um erro de menos que 0,0001%.

Portanto, pode-se especificar um resistor com essa resistência para o circuito mostrado na Figura 8.6 e esperar atingir um comportamento de dissipação consistente com as exigências do problema. Esse problema de projeto não poderia ser resolvido eficientemente sem o uso dos métodos numéricos dos Capítulos 5 e 6.

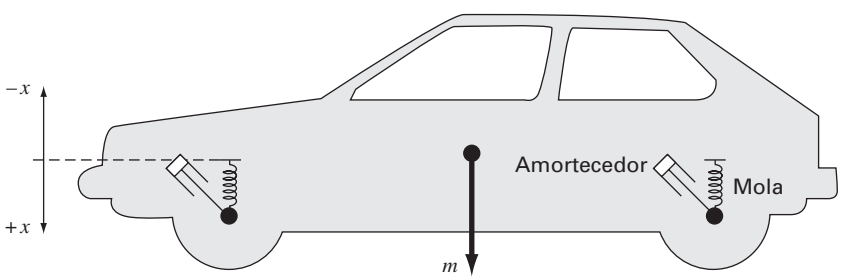
8.4 ANÁLISE DE VIBRAÇÃO (ENGENHARIA MECÂNICA/AEROESPACIAL)

Fundamentos. As equações diferenciais são freqüentemente usadas para modelar a vibração de sistemas em engenharia. Alguns exemplos (Figura 8.6) são um pêndulo simples, uma massa em uma mola e um circuito elétrico indutância-capacitância (lembre-se da Seção 8.3). A vibração desses sistemas pode ser amortecida por algum mecanismo que absorva energia. Além disso, a vibração pode ser livre ou sujeita a alguma perturbação periódica externa. No último caso, o movimento é dito *forçado*. Nesta seção, serão examinadas as vibrações livres e forçadas do automóvel mostrado na Figura 8.7. A abordagem geral é aplicável a vários outros problemas de engenharia.

Como mostrado na Figura 8.7, um carro de massa m é apoiado por molas e amortecedores. Os amortecedores oferecem uma resistência ao movimento que é proporcional à velocidade vertical (movimento para baixo e para cima). As vibrações livres ocorrem quando o carro é perturbado do equilíbrio, como depois de passar por um buraco. Em qualquer instante depois de atingir o buraco, as forças resultantes agindo em m são a resistência das molas e a força de amortecimento dos amortecedores. Essas forças tendem a fazer com que o carro volte ao seu estado de equilíbrio original. De acordo com a *lei de Hooke*, a resistência da mola é proporcional à constante da mola k e à distância da posição de equilíbrio, x. Portanto,

$$\text{Força da mola} = -kx$$

FIGURA 8.7
Um carro de massa m.



onde o sinal de menos indica que a força restauradora age na direção de restaurar a posição de equilíbrio do carro (isto é, na direção negativa de x). A força de amortecimento do amortecedor é dada por

$$\text{Força de amortecimento} = -c\frac{dx}{dt}$$

onde c é o coeficiente de amortecimento e dx/dt é a velocidade vertical. O sinal negativo indica que a força de amortecimento age na direção oposta à da velocidade.

As equações de movimento para o sistema são dadas pela segunda lei de Newton ($F = ma$), a qual para esse problema é expressa por

$$\underbrace{m}_{\text{Massa}} \times \underbrace{\frac{d^2x}{dt^2}}_{\text{aceleração}} = \underbrace{-c\frac{dx}{dt}}_{\text{força de amortecimento}} + \underbrace{(-kx)}_{\text{força da mola}}$$

ou

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = 0$$

Observe a similaridade com a Equação (8.18) que foi deduzida na Seção 8.3 para um circuito elétrico.

Supondo que a solução tenha a forma $x(t) = e^{rt}$, pode-se escrever a *equação característica*

$$mr^2 + cr + k = 0 \tag{8.21}$$

A incógnita, r, é a solução da equação característica quadrática que pode ser obtida tanto analítica quanto numericamente. Em problemas de projetos, primeiro se usa a solução analítica para obter uma visão geral da maneira pela qual o movimento do sistema é afetado pelos coeficientes do modelo — m, k e c. Também se usam vários métodos numéricos para obter soluções e verificar a acurácia dos resultados utilizando a solução analítica. Finalmente, prepara-se o terreno para problemas mais complexos que serão descritos mais tarde no texto, para os quais os resultados analíticos são difíceis ou impossíveis de obter.

A solução da Equação (8.21) para r é dada pela fórmula quadrática

$$\begin{matrix} r_1 \\ r_2 \end{matrix} = \frac{-c \pm \sqrt{c^2 - 4mk}}{2m} \tag{8.22}$$

Observe o significado do módulo de c comparado com $2\sqrt{km}$. Se $c > 2\sqrt{km}$, r_1 e r_2 são números reais negativos e a solução é da forma

$$x(t) = Ae^{r_1 t} + Be^{r_2 t} \tag{8.23}$$

onde A e B são constantes a serem determinadas a partir das condições iniciais para x e dx/dt. Tais sistemas são chamados *superamortecidos*.

Se $c < 2\sqrt{km}$, as raízes são complexas,

$$\begin{matrix} r_1 \\ r_2 \end{matrix} = \lambda \pm \mu i$$

onde

$$\mu = \frac{\sqrt{|c^2 - 4mk|}}{2m}$$

e a solução é da forma

$$x(t) = e^{-\lambda t}(A \cos \mu t + B \operatorname{sen} \mu t) \tag{8.24}$$

Tais sistemas são chamados de *subamortecidos*.

Finalmente, se $c = 2\sqrt{km}$, a equação característica tem uma raiz dupla e a solução é a da forma

$$x(t) = (A + Bt)e^{-\lambda t} \tag{8.25}$$

onde

$$\lambda = \frac{c}{2m}$$

Tais sistemas são chamados de *criticamente amortecidos.*

Em todos os três casos, *x (t)* tende a zero quando *t* tende a infinito. Isso significa que o carro sempre volta à sua posição de equilíbrio depois de passar pelo buraco (embora isso possa parecer pouco provável em algumas cidades que visitamos!). Esses casos estão ilustrados na Figura 8.8.

O *coeficiente de amortecimento crítico* c_c é o valor de c que torna o radical na Equação (8.22) igual a zero,

$$c_c = 2\sqrt{km} \qquad \text{ou} \qquad c_c = 2mp \tag{8.26}$$

onde

$$p = \sqrt{\frac{k}{m}} \tag{8.27}$$

A razão c/c_c é chamada de *fator de amortecimento*, e p é chamado de *freqüência natural* da vibração livre não amortecida.

Agora, considere o caso de um carro sujeito a uma força periódica dada por

$$P = P_m \operatorname{sen} \omega t \qquad \text{ou} \qquad d = d_m \operatorname{sen} \omega t$$

onde $d_m = P_m/k$ é a deflexão estática do carro sujeito à força P_m. A equação diferencial que rege esse caso é

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = P_m \operatorname{sen} \omega t$$

A solução geral dessa equação é obtida somando-se uma solução particular à solução de vibração livre dada pelas Equações (8.23) a (8.25). Considere um movimento estacionário do sistema forçado em que o movimento transiente inicial foi totalmente amortecido. Supondo que essa solução de estado estacionário tenha a forma

$$x_{ss}(t) = x_m \operatorname{sen}(\omega t - \phi)$$

pode ser mostrado que

$$\frac{x_m}{P_m/k} = \frac{x_m}{d_m} = \frac{1}{\sqrt{[1 - (\omega/p)^2]^2 + 4(c/c_c)^2(\omega/p)^2}} \tag{8.28}$$

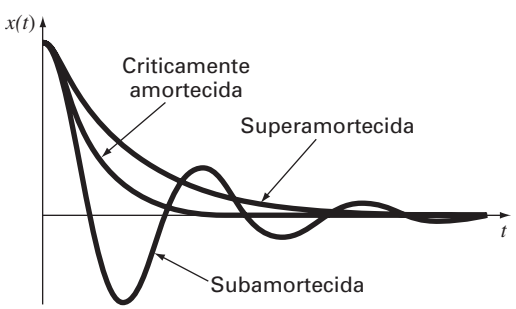


FIGURA 8.8
Oscilações (*a*) superamortecidas, (*b*) subamortecidas e (*c*) criticamente amortecidas.

A quantidade, x_m/d_m, chamada de *fator de aumento da amplitude*, depende apenas da razão do amortecimento real para o amortecimento crítico e da razão da freqüência forçante para a freqüência natural. Observe que, quando a freqüência forçante ω se aproxima de zero, o fator de aumento tende a 1. Além disso, se o sistema for levemente amortecido, ou seja, se c/c_c for pequeno, então o fator de aumento se torna infinito quando ω tende a p. Se o amortecimento for nulo, então o fator de aumento se torna infinito quando $\omega = p$ e diz-se que o termo forçante está em *ressonância* com o sistema. Finalmente, quando ω/p se torna muito grande, o fator de aumento se aproxima de zero. A Figura 8.9 mostra gráficos do fator de aumento como função de ω/p para diversos fatores de amortecimento.

Observe que o fator de amortecimento pode ser mantido pequeno escolhendo-se um fator de aumento grande ou mantendo-se as freqüências natural e forçada bem distantes.

O projeto da suspensão de um carro envolve os prós e contras entre o conforto e a estabilidade em todas as condições de condução e velocidades. Pediram que você determinasse a estabilidade de um projeto proposto que tem boas condições de conforto em estradas rugosas. A massa do carro é $m = 1{,}2 \times 10^6$ g e ele tem um amortecedor com coeficiente de amortecimento $c = 1 \times 10^7$ g/s.

Suponha que a expectativa de conforto do público esteja satisfeita se as vibrações livres do carro forem subamortecidas e o primeiro cruzamento com a posição de equilíbrio ocorra em 0,05 s. Se, em $t = 0$, o carro for subitamente deslocado da posição de equilíbrio, x_0, e a velocidade for zero ($dx/dt = 0$), a solução da equação de movimento é dada pela Equação (8.24), com $A = x_0$ e $B = x_0\lambda/\mu$. Portanto,

$$x(t) = x_0 e^{-\lambda t}\left(\cos \mu t + \frac{\lambda}{\mu} \operatorname{sen} \mu t\right)$$

As condições do projeto são satisfeitas se

$$x(t) = 0 = \cos(0{,}05\mu) + \frac{\lambda}{\mu} \operatorname{sen}(0{,}05\mu)$$

ou

$$0 = \cos\left(0{,}05\sqrt{\frac{k}{m} - \frac{c^2}{4m^2}}\right) + \frac{c}{\sqrt{4km - c^2}} \operatorname{sen}\left(0{,}05\sqrt{\frac{\mathrm{k}}{\mathrm{m}} - \frac{\mathrm{c}^2}{4\mathrm{m}^2}}\right) \tag{8.29}$$

Como c e m são dados, o problema de projeto se reduziu a encontrar um valor apropriado de k que satisfaça a Equação (8.29).

Solução. Isso pode ser feito usando os métodos da bissecção, falsa posição ou secante, pois eles não necessitam do cálculo da derivada da Equação (8.29), o que poderia ser considerado um pouco inconveniente nesse problema. A solução é $k = 1{,}397 \times 10^9$, com 12 iterações do método da bissecção com um intervalo inicial de $k = 1 \times 10^9$ a 2×10^9 $\varepsilon_a = 0{,}07305\%$.

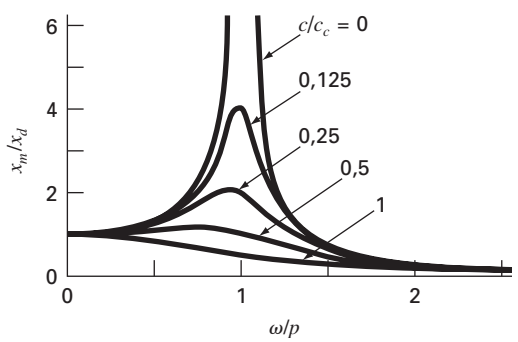


FIGURA 8.9
Gráfico do fator de aumento de amplitude x_m/x_d [Equação (8.28)] em função da freqüência ω dividida pela freqüência natural p para diversos valores do coeficiente de amortecimento c dividido pelo coeficiente de amortecimento crítico c_c.

Embora o projeto satisfaça as exigências de vibração livre (depois de atingir um buraco), ele também deve ser testado sob condições rugosas de estrada. A superfície da estrada pode ser aproximada por

$$d = d_m \operatorname{sen}\left(\frac{2\pi \mathrm{x}}{\mathrm{D}}\right)$$

onde d é a deflexão, d_m é a deflexão máxima de 0,1 m e D é a distância entre picos, igual a 20 m. Se a velocidade horizontal do carro for v (m/s), então a equação global do movimento para o sistema pode ser escrita como

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = kd_m \operatorname{sen}\left(\frac{2\pi \mathrm{v}}{\mathrm{D}}\mathrm{t}\right)$$

onde $\omega = 2\pi v/D$ é a freqüência forçante.

A estabilidade do carro é considerada satisfatória se em um estado estacionário a distância máxima x_m está abaixo de 0,2 m para todas as velocidades de condução. O fator de amortecimento é calculado de acordo com a Equação (8.26),

$$\frac{c}{c_c} = \frac{10}{2\sqrt{km}} = \frac{1 \times 10^7}{2\sqrt{1{,}397 \times 10^9(1{,}2 \times 10^6)}} = 0{,}1221$$

Agora, procuram-se valores de ω/p que satisfaçam a Equação (8.28),

$$2 = \frac{1}{\sqrt{[1-(\omega/p)^2]^2 + 4(0{,}1221)^2(\omega/p)^2}} \tag{8.30}$$

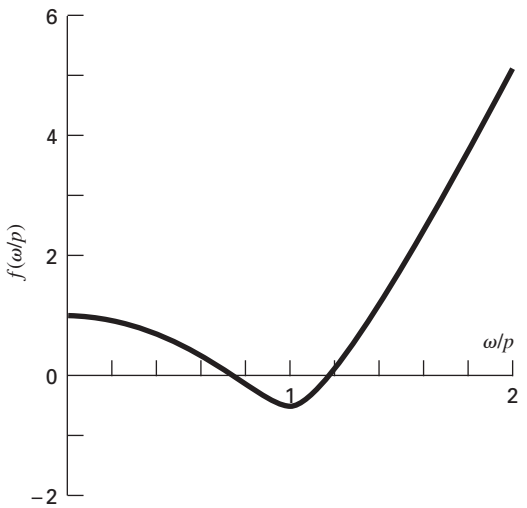
Quando a Equação (8.30) é expressa como um problema de raiz, tem-se

$$f(\omega/p) = 2\sqrt{[1-(\omega/p)^2]^2 + 4(0{,}1221)^2(\omega/p)^2} - 1 = 0 \tag{8.31}$$

Vê-se que os valores de ω/p podem ser determinados encontrando-se as raízes da Equação (8.31).

Um gráfico da Equação (8.31) é encontrado na Figura 8.10, e mostra que a Equação (8.31) tem duas raízes positivas que podem ser determinadas pelo método da bissecção. Em 18 iterações, obtém-se que o menor valor para ω/p é igual a 0,7300, com um erro estimado de 0,000525%, com aproximações inferior e superior de 0 e 1. Em 17 iterações,

FIGURA 8.10
Gráfico da Equação (8.31) indicando duas raízes positivas.



encontra-se que o maior valor para ω/p é igual a 1,1864, com um erro estimado de 0,00064%, com aproximações inferior e superior de 1 e 2.

Também é possível expressar a Equação (8.30) como um polinômio,

$$\left(\frac{\omega}{p}\right)^4 - 1{,}9404\left(\frac{\omega}{p}\right)^2 + 0{,}75 \tag{8.32}$$

e usar o software MATLAB para determinar as raízes, como a seguir:

```
>> a=[1 0 -1,9404 0 ,75];
>> roots (a)
ans =

    1,1864
   -1,1864
    0,7300
   -0,7300
```

Isso confirma o resultado obtido com a bissecção. E também sugere que, embora superficialmente pareça ser uma equação de quarto grau para ω/p, a Equação (8.32) é na realidade uma equação quadrática para $(\omega/p)^2$.

O valor da freqüência natural p é dado pela Equação (8.27),

$$p = \sqrt{\frac{1{,}397 \times 10^9}{1{,}2 \times 10^6}} = 34{,}12 \text{ s}^{-1}$$

As freqüências forçantes, para as quais a deflexão máxima é 0,2 m, são então calculadas por

$$\omega = 0{,}7300(34{,}12) = 24{,}91 \text{ s}^{-1}$$
$$\omega = 1{,}1864(34{,}12) = 40{,}48 \text{ s}^{-1}$$

o que fornece

$$v = \frac{\omega D}{2\pi} = \frac{24{,}91(20)}{2(3{,}14159)} = 79{,}29 \frac{\text{m}}{\text{s}} \times \frac{3600 \text{ s}}{\text{h}} \frac{\text{km}}{1000 \text{ m}} = 285 \text{ km/h} \ (= 177 \text{ mi/h})$$

$$v = \frac{\omega D}{2\pi} = \frac{40{,}48(20)}{2(3{,}14159)} = 128{,}85 \frac{\text{m}}{\text{s}} \times \frac{3600 \text{ s}}{\text{h}} \frac{\text{km}}{1000 \text{ m}} = 464 \text{ km/h} \ (= 288 \text{ mi/h})$$

Portanto, usando esses resultados e a Figura 8.10, encontra-se que o projeto proposto para o carro apresentará comportamento aceitável para as velocidades comuns de condução. Nesse ponto, o projetista deve estar ciente de que o projeto não satisfaz as condições de adequação para velocidades extremamente altas (por exemplo, em corridas).

Esse problema de projeto apresentou um exemplo extremamente simples que permitiu obter alguns resultados analíticos que foram usados para avaliar a acurácia dos métodos numéricos para encontrar as raízes. Os casos reais podem rapidamente se tornar tão complicados que as soluções só são obtidas usando-se métodos numéricos.

PROBLEMAS

Engenharia Química/Bioengenharia

8.1 Faça os mesmos cálculos que na Seção 8.1, mas para o álcool etílico ($a = 12{,}02$ e $b = 0{,}08407$) a uma temperatura de 400 K e p de 2,5 atm. Compare seus resultados com a lei dos gases ideais. Se possível, use um software de seu computador para determinar o volume molar. Caso contrário, use qualquer um dos métodos numéricos discutidos nos Capítulos 5 e 6 para fazer os cálculos. Justifique sua escolha da técnica.

8.2 Em engenharia química, os reatores tubulares de fluxo pistonado (isto é, aqueles nos quais o fluido escoa de uma extremidade para a outra com mistura mínima ao longo do eixo longitudinal) são usados com freqüência para converter reagentes em produtos. Foi determinado que a eficiência da conversão pode algumas vezes ser melhorada reciclando-se uma parte da corrente de produto de modo que ele volte para a entrada para uma passagem adicional pelo reator (Figura P8.2). A taxa de reciclagem é definida por

$$R = \frac{\text{volume do fluido que volta para a entrada}}{\text{volume deixando o sistema}}$$

Suponha que estamos processando uma substância química A para gerar um produto B. Para o caso em que A forma B por meio de uma reação autocatalisadora (isto é, na qual um dos produtos age como um catalisador ou um estímulo para a reação), pode ser mostrado que uma taxa de reciclagem ótima deve satisfazer

$$\ln \frac{1 + R(1 - X_{Af})}{R(1 - X_{Af})} = \frac{R + 1}{R[1 + R(1 - X_{Af})]}$$

onde X_{Af} é a fração de reagente A que é convertida para o produto B. A taxa de reciclagem ótima corresponde ao reator de tamanho mínimo necessário para se atingir o nível desejado de conversão. Use um método numérico para determinar as taxas de reciclagem necessárias para minimizar o tamanho do reator para uma fração de conversão de $X_{Af} = 0{,}95$.

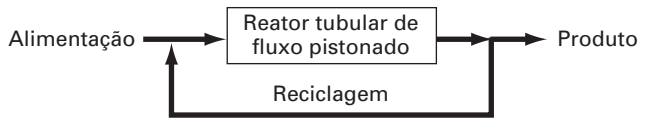


Figura P8.2
Representação esquemática de um reator tubular de fluxo pistonado com reciclagem.

8.3 Em um processo de engenharia química, vapor de água (H_2O) é aquecido a uma temperatura suficientemente alta para que uma parte significativa da água se dissocie, ou se quebre, para formar oxigênio (O_2) e hidrogênio (H_2):

$$H_2O \rightleftharpoons H_2 + \tfrac{1}{2}O_2$$

Se for suposto que essa é a única reação envolvida, a fração molar x de H_2O que se dissocia pode ser representada por

$$K = \frac{x}{1 - x}\sqrt{\frac{2p_t}{2 + x}} \tag{P8.3}$$

onde K é a constante de equilíbrio da reação e p_t é a pressão total da mistura. Se $p_t = 3{,}5$ atm e $K = 0{,}04$, determine o valor de x que satisfaz a Equação (P8.3).

8.4 A seguinte equação está ligada à concentração de uma substância química em um reator completamente misturado:

$$c = c_{\text{in}}(1 - e^{-0{,}04t}) + c_0 e^{-0{,}04t}$$

Se a concentração inicial for $c_0 = 5$ e a concentração do fluxo de entrada for $c_{\text{in}} = 12$, calcule o tempo necessário para que c seja 85% de c_{in}.

8.5 Uma reação química reversível

$$2A + B \rightleftharpoons C$$

pode ser caracterizada pela relação de equilíbrio

$$K = \frac{c_c}{c_a^2 c_b}$$

onde o símbolo c_i representa a concentração do constituinte i. Suponha que definamos uma variável x que represente o número de moles de C que são produzidos. A conservação da massa pode ser usada para reformular a relação de equilíbrio como

$$K = \frac{(c_{c,0} + x)}{(c_{a,0} - 2x)^2(c_{b,0} - x)}$$

onde o subscrito 0 indica a concentração inicial de cada constituinte. Se $K = 0{,}016$, $c_{a,0} = 42$, $c_{b,0} = 28$ e $c_{c,0} = 4$, determine o valor de x. **(a)** Obtenha a solução graficamente. **(b)** Com base em **(a)**, resolva para determinar a raiz com aproximações iniciais $x_l = 0$ e $x_u = 20$ até $\varepsilon_s = 0{,}5\%$. Escolha ou a bissecção ou a falsa posição para obter a solução. Justifique sua escolha.

8.6 As seguintes reações químicas ocorrem em um sistema fechado:

$$2A + B \rightleftharpoons C$$
$$A + D \rightleftharpoons C$$

No equilíbrio, elas podem ser caracterizadas por

$$K_1 = \frac{c_c}{c_a^2 c_b}$$
$$K_2 = \frac{c_c}{c_a c_d}$$

onde o símbolo c_i representa a concentração do constituinte i. Se x_1 e x_2 são o número de moles de C que são produzidos por causa da primeira e da segunda reações, respectivamente, use uma abordagem parecida com a do Problema 8.5 para reformular a relação de equilíbrio em termos das concentrações iniciais dos constituintes. A seguir, use o método de Newton-Raphson para resolver o par de equações não-lineares simultâneas para determinar x_1 e x_2 se $K_1 = 4 \times 10^{-4}$, $K_2 = 3{,}7 \times 10^{-2}$, $c_{a,0} = 50$, $c_{b,0} = 20$, $c_{c,0} = 5$ e $c_{d,0} = 10$. Utilize uma abordagem gráfica para encontrar suas aproximações iniciais.

8.7 A equação de estado de Redlich-Kwong é dada por

$$p = \frac{RT}{v - b} - \frac{a}{v(v + b)\sqrt{T}}$$

onde R é a constante universal dos gases [$= 0{,}518$ kJ/(kg K)], T é a temperatura absoluta (K), p é a pressão absoluta (kPa) e v é o volume de um kg de gás (m^3/kg). Os parâmetros a e b são calculados por

$$a = 0{,}427\frac{R^2 T_c^{2{,}5}}{p_c} \qquad b = 0{,}0866R\frac{T_c}{p_c}$$

onde p_c é a pressão crítica (kPa) e T_c é a temperatura crítica (K). Como engenheiro químico, foi-lhe pedido para determinar a quantidade de gás metano ($p_c = 4580$ kPa e $T_c = 191$ K) que pode ser mantido em um tanque de 3 m^3 a uma temperatura de -50°C com uma pressão de 65,000 kPa. Use um método para localizar raízes de sua escolha para calcular v e a seguir determinar a massa de metano contida no tanque.

8.8 O volume V de líquido em um cilindro horizontal raso de raio r e comprimento L se relaciona com a profundidade do líquido h por

$$V = \left[r^2 \cos^{-1}\left(\frac{r - h}{r}\right) - (r - h)\sqrt{2rh - h^2}\right] L$$

Determine h dado que $r = 2$ m, $L = 5$ m, e $V = 8{,}5$ m^3. Observe que, se você estiver usando uma linguagem de programação ou ferramenta de software que não seja rica em funções trigonométricas, o arco-co-seno pode ser calculado por

$$\cos^{-1} x = \frac{\pi}{2} - \text{tg}^{-1}\left(\frac{x}{\sqrt{1 - x^2}}\right)$$

8.9 O volume V de líquido em um tanque esférico de raio r está relacionado com a profundidade h do líquido por

$$V = \frac{\pi h^2(3r - h)}{3}$$

Determine h dado que $r = 1$ m e $V = 0{,}75$ m^3.

8.10 Para o tanque esférico do Problema 8.9, é possível desenvolver as seguintes fórmulas de dois pontos fixos:

$$h = \sqrt{\frac{h^3 + (3V/\pi)}{3r}}$$

e

$$h = \sqrt[3]{3\left(rh^2 - \frac{V}{\pi}\right)}$$

Se $r = 1$ m e $V = 0{,}75$ m^3, determine se alguma delas é estável e o intervalo de aproximações iniciais para o qual elas são estáveis.

8.11 A equação de Ergun, mostrada a seguir, é usada para descrever o escoamento de um fluido através de um leito compactado. ΔP é a queda de pressão, ρ é a densidade do fluido, G_o é a velocidade da massa (vazão em massa dividida pela área da seção transversal), D_p é o diâmetro das partículas dentro do leito, μ é a viscosidade do fluido, L é o comprimento do leito e ε é a fração vazia do leito.

$$\frac{\Delta P \rho}{G_o^2}\frac{D_p}{L}\frac{\varepsilon}{1-\varepsilon} = 150\frac{1-\varepsilon}{(D_p G_o/\mu)} + 1{,}75$$

Dados os valores dos parâmetros listados a seguir, encontre a fração vazia ε do leito.

$$\frac{D_p G_o}{\mu} = 1000$$

$$\frac{\Delta P \rho D_p}{G_o^2 L} = 20$$

8.12 A queda de pressão na seção de um tubo pode ser calculada por

$$\Delta p = f\frac{L\rho V^2}{2D}$$

onde Δp é a queda de pressão (Pa), f é o fator de atrito, L é o comprimento do tubo [m], ρ é a densidade (kg/m^3), V é a velocidade (m/s) e D é o diâmetro (m). Para escoamentos turbulentos, a *equação de Colebrook* fornece uma maneira de calcular o fator de atrito

$$\frac{1}{\sqrt{f}} = -2{,}0\log\left(\frac{\varepsilon}{3{,}7D} + \frac{2{,}51}{\mathrm{Re}\sqrt{f}}\right)$$

onde ε é a rugosidade (m) e Re é o *número de Reynolds*,

$$\mathrm{Re} = \frac{\rho V D}{\mu}$$

onde μ é a viscosidade dinâmica (N · s/m^2).

(a) Determine Δp para uma extensão horizontal de 0,2 m de comprimento de tubulação submersa lisa, dado que $\rho = 1{,}23$ kg/m^3, $\mu = 1{,}79 \times 10^{-5}$ N · s/m^2, $D = 0{,}005$ m, $V = 40$ m/s e $\varepsilon = 0{,}0015$ mm. Use um método numérico para determinar o fator de atrito. Observe que para os tubos lisos com Re $< 10^5$, uma boa aproximação inicial pode ser obtida usando a *fórmula Blasius,* $f = 0{,}316/\mathrm{Re}^{0{,}25}$.

(b) Repita os cálculos, agora para um tubo de aço comercial mais rugoso ($\varepsilon = 0{,}045$ mm).

8.13 O pH da água é muito importante para os engenheiros ambientalistas e químicos. Ele pode ser relacionado a processos variando de corrosão de tubos a chuva ácida. O pH está relacionado com a concentração de íon hidrogênio por

$$\mathrm{pH} = -\log_{10}[\mathrm{H}^+]$$

As seguintes cinco equações governam a concentração da mistura de dióxido de carbono e água em um sistema fechado:

$$K_1 = \frac{[\mathrm{H}^+][\mathrm{HCO}_3^-]}{[\mathrm{CO}_2]}$$

$$K_2 = \frac{[\mathrm{H}^+]\left[\mathrm{CO}_3^{2-}\right]}{[\mathrm{HCO}_3^-]}$$

$$K_w = [\mathrm{H}^+][\mathrm{OH}^-]$$

$$c_T = [\mathrm{CO}_2] + [\mathrm{HCO}_3^-] + \left[\mathrm{CO}_3^{2-}\right]$$

$$\mathrm{Alk} = [\mathrm{HCO}_3^-] + 2\left[\mathrm{CO}_3^{2-}\right] + [\mathrm{OH}^-] - [\mathrm{H}^+]$$

onde Alk é a alcalinidade, c_T é o carbono inorgânico total e os K's são os coeficientes de equilíbrio. As cinco incógnitas são $[\mathrm{CO}_2]$ para o dióxido de carbono, $[\mathrm{HCO}_3^-]$ para o bicarbonato, $[\mathrm{CO}_3^{2-}]$ para o carbonato, $[\mathrm{H}^+]$ para íon hidrogênio e $[\mathrm{OH}^-]$ para o íon hidroxila. Determine as cinco incógnitas dado que $\mathrm{Alk} = 2 \times 10^{-3}$, $c_T = 3 \times 10^{-3}$, $K_1 = 10^{-6{,}3}$, $K_2 = 10^{-10{,}3}$ e $K_w = 10^{-14}$. Além disso, calcule o pH da solução.

8.14 A operação de um reator tubular de fluxo pistonado com densidade constante para a produção de uma substância por meio de uma reação enzimática é descrita pela equação a seguir, onde V é o volume do reator, F é a taxa de escoamento do reagente C, C_{ent} e $C_{\mathrm{saída}}$ são as concentrações dos reagentes entrando e saindo do reator, respectivamente, e K e $k_{\mathrm{máx}}$ são constantes. Para um reator com 500 L, com uma concentração de entrada de $C_{\mathrm{ent}} = 0{,}5$ M, uma vazão de entrada de 40 L/s, $k_{\mathrm{máx}} = 5 \times 10^{-3}$ s^{-1} e $K = 0{,}1$ M, encontre a concentração de C na saída do reator.

$$\frac{V}{F} = -\int_{C_{\mathrm{ent}}}^{C_{\mathrm{saída}}} \frac{K}{k_{\max} C} + \frac{1}{k_{\max}} dC$$

Engenharia Civil e Ambiental

8.15 O deslocamento de uma estrutura é definido pela seguinte equação para uma oscilação amortecida:

$$y = 9e^{-kt}\cos\omega t$$

onde $k = 0{,}7$ e $\omega = 4$.

(a) Use o método gráfico para fazer uma estimativa inicial do tempo necessário para que o deslocamento diminua para 3,5.

(b) Use o método de Newton-Raphson para determinar a raiz até $\varepsilon_s = 0{,}01\%$.

(c) Use o método da secante para determinar a raiz até $\varepsilon_s = 0{,}01\%$.

8.16 Na engenharia estrutural, a *fórmula da secante* define a força por unidade de área P/A que causa uma tensão σ_m máxima em uma coluna de taxa de finura dada L/k:

$$\frac{P}{A} = \frac{\sigma_m}{1 + (ec/k^2)\sec[0{,}5\sqrt{P/(EA)}(L/k)]}$$

onde ec/k^2 é o raio de excentricidade e E é o módulo de elasticidade. Se para uma viga de aço $E = 200.000$ MPa, $ec/k^2 = 0{,}4$ e $\sigma_m = 250$ MPa, calcule P/A para $L/k = 50$. Lembre-se de que $\sec x = 1/\cos x$.

8.17 Um cabo catenário é um cabo que está pendurado entre dois pontos que não estão na mesma reta vertical. Como descrito na

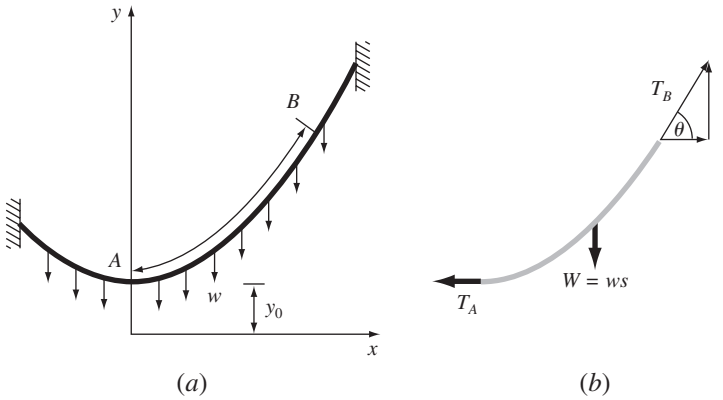


Figura P8.17
(*a*) As forças agindo em uma seção *AB* de um cabo flexível pendurado. A carga é uniforme ao longo do cabo (mas não é uniforme por distância horizontal *x*). (*b*) Um diagrama de corpo livre da seção *AB*.

Figura P8.17*a*, ele não está sujeito a nenhuma carga além do seu próprio peso. Portanto, seu peso (N/m) age como uma carga uniforme por unidade de comprimento ao longo do cabo. Um diagrama de corpo livre de uma seção *AB* é descrito na Figura P8.17*b*, onde T_A e T_B são as forças de tensão nas extremidades. Com base nos balanços de força horizontal e vertical, é possível deduzir a seguinte equação diferencial para modelar o cabo:

$$\frac{d^2y}{dx^2} = \frac{w}{T_A}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}$$

O cálculo pode ser usado para resolver essa equação para determinar a altura y do cabo como uma função da distância x,

$$y = \frac{T_A}{w}\cosh\left(\frac{w}{T_A}x\right) + y_0 - \frac{T_A}{w}$$

onde o cosseno hiperbólico pode ser calculado por

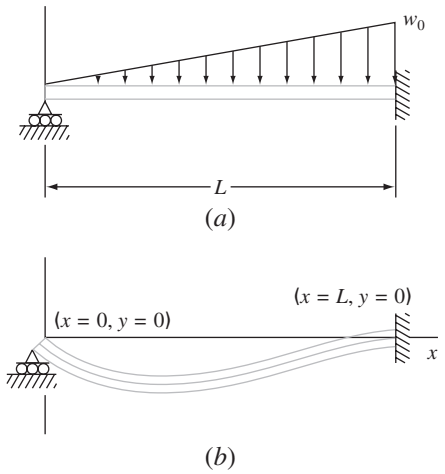
$$\cosh x = \frac{1}{2}(e^x + e^{-x})$$

Use um método numérico para calcular um valor para o parâmetro T_A dados os valores para os parâmetros $w = 12$ e $y_0 = 6$, tal que o cabo tenha uma altura $y = 15$ em $x = 50$.

8.18 A Figura P8.18*a* mostra uma viga uniforme sujeita a uma carga distribuída de forma linearmente crescente. A equação para a curva elástica resultante é (ver Figura P8.18*b*)

$$y = \frac{w_0}{120EIL}(-x^5 + 2L^2x^3 - L^4x) \qquad \text{(P8.18)}$$

Figura P8.18



Use a bissecção para determinar o ponto de deflexão máxima (isto é, o valor de x onde $dy/dx = 0$). A seguir, substitua esse valor na Equação (P8.18) para determinar o valor da deflexão máxima. Use os seguintes valores dos parâmetros nos seus cálculos: $L = 600$ cm, $E = 50.000$ kN/cm^2, $I = 30.000$ cm^4 e $w_0 = 2{,}5$ kN/cm.

8.19 Na engenharia ambiental (uma área especializada da engenharia civil), a seguinte equação pode ser usada para calcular o nível de oxigênio c (mg/L) em um rio a jusante de uma descarga de esgoto:

$$c = 10 - 20(e^{-0{,}15x} - e^{-0{,}5x})$$

onde x é a distância a jusante em quilômetros.

(a) Determine a distância a jusante onde o nível de oxigênio cai pela primeira vez até a leitura de 5 mg/L. (*Sugestão*: está a menos de 2 km da descarga). Determine sua resposta a menos de um erro de 1%. Observe que níveis de oxigênio abaixo de 5 mg/L, em geral, são prejudiciais a peixes tais como trutas e salmões.

(b) Determine a distância a jusante na qual o oxigênio está em um mínimo. Qual é a concentração nessa posição?

8.20 A concentração da bactéria poluente c em um lago diminui de acordo com

$$c = 75e^{-1{,}5t} + 20e^{-0{,}075t}$$

Determine o tempo necessário para que a concentração de bactéria seja reduzida a 15 usando **(a)** o método gráfico e **(b)** o método de Newton-Raphson com aproximação inicial $t = 6$ e critério de parada 0,5%. Verifique seu resultado.

8.21 Na engenharia oceanográfica, a equação para uma onda estacionária em um cais é dada por $\lambda = 16$, $t = 12$, $v = 48$:

$$h = h_0\left[\operatorname{sen}\left(\frac{2\pi x}{\lambda}\right)\cos\left(\frac{2\pi tv}{\lambda}\right) + e^{-x}\right]$$

Determine o menor valor positivo de x se $h = 0{,}5h_0$.

8.22 Você comprou uma peça de equipamento de $ 25.000 sem entrada e pagando $ 5.500 por ano por seis anos. Qual a taxa de juros que você está pagando? A fórmula que relaciona o valor atual P, os pagamentos anuais A, o número de anos n e a taxa de juros i é

$$A = P\frac{i(1+i)^n}{(1+i)^n - 1}$$

8.23 Muitos campos da engenharia necessitam de estimativas acuradas da população. Por exemplo, os engenheiros de transporte

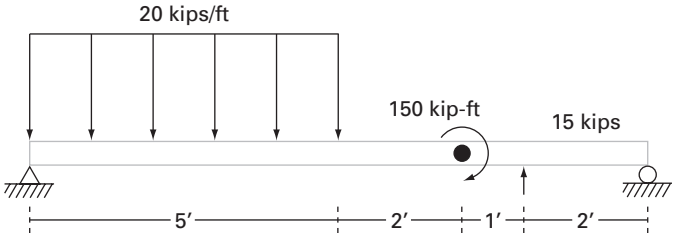


Figura P8.24

podem achar necessário determinar separadamente a tendência de crescimento da população de uma cidade e do subúrbio adjacente. A população da área urbana está diminuindo com o tempo de acordo com

$$P_u(t) = P_{u,\max} e^{-k_u t} + P_{u,\min}$$

enquanto a população suburbana está crescendo de acordo com

$$P_s(t) = \frac{P_{s,\max}}{1 + [P_{s,\max}/P_0 - 1]e^{-k_s t}}$$

onde $P_{u,\max}$, k_u, $P_{s,\max}$, P_0 e k_s são parâmetros determinados empiricamente. Determine o tempo e os valores correspondentes de $P_u(t)$ e $P_s(t)$ quando os subúrbios forem 20% maiores que a cidade. Os valores dos parâmetros são $P_{u,\max} = 75{,}000$, $k_u = 0{,}045$/ano, $P_{u,\text{mín}} = 100.000$ pessoas, $P_{s,\max} = 300.000$ pessoas, $P_0 = 10.000$ pessoas, $k_s = 0{,}08$/ano. Para obter suas soluções, use os métodos **(a)** gráfico, **(b)** falsa posição e **(c)** secante modificado.

8.24 Uma viga apoiada em apenas um ponto é carregada como mostra a Figura P8.24. Usando funções de singularidade, a tensão ao longo da viga pode ser expressa pela equação:

$$V(x) = 20[\langle x-0\rangle^1 - \langle x-5\rangle^1] - 15\langle x-8\rangle^0 - 57$$

Por definição, a função de singularidade pode ser expressa como:

$$\langle x-a\rangle^n = \begin{cases} (x-a)^n & \text{quando } x > a \\ 0 & \text{quando } x \le a \end{cases}$$

Use um método numérico para encontrar o(s) ponto(s) onde a tensão é igual a zero.

8.25 Para a mesma viga apoiada em um único ponto do Problema 8.24, o torque ao longo da viga é dado por:

$$M(x) = -10[\langle x-0\rangle^2 - \langle x-5\rangle^2] + 15\langle x-8\rangle^1 + 150\langle x-7\rangle^0 + 57x$$

Use um método numérico para encontrar o(s) ponto(s) onde o torque é igual a zero.

8.26 Para a mesma viga apoiada em um único ponto do Problema 8.24, a inclinação ao longo da viga é dada por:

$$\frac{du_y}{dx}(x) = \frac{-10}{3}[\langle x-0\rangle^3 - \langle x-5\rangle^3] + \frac{15}{2}\langle x-8\rangle^2 + 150\langle x-7\rangle^1 + \frac{57}{2}x^2 - 238{,}25$$

Use um método numérico para encontrar o(s) ponto(s) onde a inclinação é igual a zero.

8.27 Para a mesma viga apoiada em um único ponto do Problema 8.24, o deslocamento ao longo da viga é dado por:

$$u_y(x) = \frac{-5}{6}[\langle x-0\rangle^4 - \langle x-5\rangle^4] + \frac{15}{6}\langle x-8\rangle^2 + 75\langle x-7\rangle^2 + \frac{57}{6}x^3 - 238{,}25x$$

(a) Encontre o(s) ponto(s) onde o deslocamento é igual a zero.
(b) Como você usaria uma técnica de localização de raízes para determinar a posição de deslocamento mínimo?

Engenharia Elétrica

8.28 Faça os mesmos cálculos que na Seção 8.3 para determinar o valor C necessário para que o circuito dissipe até atingir 1% do seu valor original em $t = 0{,}05$ s, dado que $R = 280\,\Omega$ e $L = 7{,}5$ H. Use **(a)** uma abordagem gráfica, **(b)** bissecção e **(c)** um software de localização de raízes tal como o Excel Solver ou a função `fzero` do MATLAB.

8.29 Uma corrente oscilante em um circuito elétrico é descrita por $i = 9e^{-t}\cos(2\pi t)$, onde t está em segundos. Determine todos os valores t tais que $i = 3$.

8.30 A resistividade ρ do silício dopado é baseada na carga q de um elétron, na densidade eletrônica n e na mobilidade do elétron μ. A densidade eletrônica é dada em termos da densidade da dopagem N e da densidade de transporte intrínseca n_i. A mobilidade do elétron é descrita pela temperatura T, pela temperatura de referência T_0 e pela mobilidade de referência μ_0. As equações necessárias para o cálculo da resistividade são

$$\rho = \frac{1}{qn\mu}$$

onde

$$n = \frac{1}{2}\left(N + \sqrt{N^2 + 4n_i^2}\right) \qquad \text{e} \qquad \mu = \mu_0\left(\frac{T}{T_0}\right)^{-2{,}42}$$

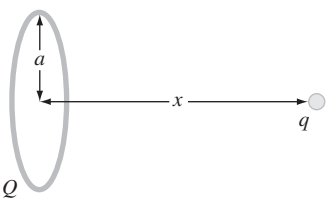
Determine N dado que $T_0 = 300$ K, $T = 1000$ K, $\mu_0 = 1.350$ cm^2 (V s)$^{-1}$, $q = 1{,}7 \times 10^{-19}$ C, $n_i = 6{,}21 \times 10^9$ cm^{-3} e uma resistividade desejada de $\rho = 6{,}5 \times 10^6$ V s cm/C. Use o método **(a)** da bissecção e **(b)** secante modificado.

8.31 Uma carga total Q está uniformemente distribuída ao redor de um condutor circular de raio a. Uma carga q está localizada a uma distância x do centro do anel (Figura P8.31). A força exercida na carga pelo anel é dada por

$$F = \frac{1}{4\pi e_0}\frac{qQx}{(x^2 + a^2)^{3/2}}$$

onde $e_0 = 8{,}85 \times 10^{-12}$ C^2/(N m^2). Encontre a distância x onde a força é 1,25 N se q e Q são 2×10^{-5} C para um anel de raio 0,9 m.

Figura P8.31



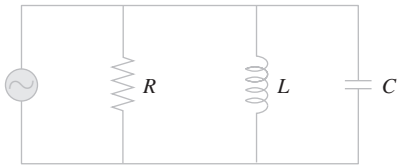


Figura P8.32

8.32 A Figura 8.32 mostra um circuito com um resistor, um indutor e um capacitor em paralelo. As regras de Kirchhoff podem ser usadas para exprimir a impedância do sistema como

$$\frac{1}{Z} = \sqrt{\frac{1}{R^2} + \left(\omega C - \frac{1}{\omega L}\right)^2}$$

onde Z é a impedância (Ω) e ω é a freqüência angular. Encontre a freqüência ω que resulta em uma impedância de 75 Ω usando tanto a bissecção quanto a falsa posição com aproximações iniciais de 1 e 1.000 para os seguintes parâmetros: $R = 225\ \Omega$, $C = 0{,}6 \times 10^{-6}$ F e $L = 0{,}5$ H. Determine quantas interações de cada técnica são necessárias para determinar a resposta até $\varepsilon_s = 0{,}1\%$. Use a abordagem gráfica para explicar quaisquer dificuldades que apareçam.

Engenharia Mecânica e Aeroespacial

8.33 Para o escoamento de fluidos em tubos, o atrito é descrito por um número adimensional, o *fator de atrito de Fanning* f. O fator de atrito de Fanning depende de diversos parâmetros relacionados ao tamanho do tubo e ao fluido, os quais podem todos ser representados por uma outra quantidade adimensional, o *número de Reynolds* Re. Uma fórmula que prevê f dado Re é a *equação de von Karman*,

$$\frac{1}{\sqrt{f}} = 4\log_{10}(\mathrm{Re}\sqrt{f}) - 0{,}4$$

Para o escoamento turbulento, os valores típicos do número de Reynolds estão entre 10.000 e 500.000 e os do fator de atrito de Fanning estão entre 0,001 e 0,01. Desenvolva uma função que use a bissecção para determinar f dado um valor fornecido pelo usuário de Re entre 2.500 e 1.000.000. Desenvolva a função para que ela garanta que o erro absoluto no resultado seja $E_{a,d} < 0{,}000005$.

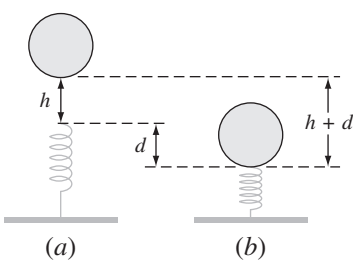
8.34 Os sistemas mecânicos reais podem envolver a deflexão de molas não-lineares. Na Figura 8.34, uma massa m é abandonada a uma distância h acima de uma mola não-linear. A força de resistência F da mola é dada por

$$F = -\left(k_1 d + k_2 d^{3/2}\right)$$

A conservação da energia pode ser usada para mostrar que

$$0 = \frac{2k_2 d^{5/2}}{5} + \frac{1}{2}k_1 d^2 - mgd - mgh$$

Figura P8.34



Determine d dados os seguintes valores dos parâmetros: $k_1 = 50.000\text{g/s}^2$, $k_2 = 40\text{g/(s}^2\ \text{m}^{0,5})$, $m = 90$ g, $g = 9{,}81\ \text{m/s}^2$ e $h = 0{,}45$ m.

8.35 Os engenheiros mecânicos, da mesma forma que a maioria dos outros engenheiros, utilizam a termodinâmica freqüentemente nos seus trabalhos. O seguinte polinômio pode ser usado para relacionar o calor específico a pressão nula do ar seco, c_p kJ/(kg K), à temperatura (K):

$$c_p = 0{,}99403 + 1{,}671 \times 10^{-4}T + 9{,}7215 \times 10^{-8}T^2 - 9{,}5838 \times 10^{-11}T^3 + 1{,}9520 \times 10^{-14}T^4$$

Determine a temperatura que corresponde a um calor específico de 1,1 kJ/(kg K).

8.36 Os engenheiros aeroespaciais algumas vezes calculam a trajetória de projéteis como foguetes. Um problema relacionado trata da trajetória de uma bola lançada. A trajetória da bola é definida pelas coordenadas (x, y) como mostrado na Figura P8.36. A trajetória pode ser modelada por

$$y = (\text{tg}\ \theta_0)x - \frac{g}{2v_0^2 \cos^2\theta_0}x^2 + y_0$$

Encontre o ângulo inicial apropriado θ_0, se a velocidade inicial for $v_0 = 20$ m/s e a distância do receptor x for 35 m. Observe que a bola deixa a mão do lançador a uma elevação de $y_0 = 2$ m e o receptor a recebe a 1 m. Expresse o resultado final em graus. Use um valor de 9,81 m/s^2 para g e utilize um método gráfico para encontrar suas aproximações iniciais.

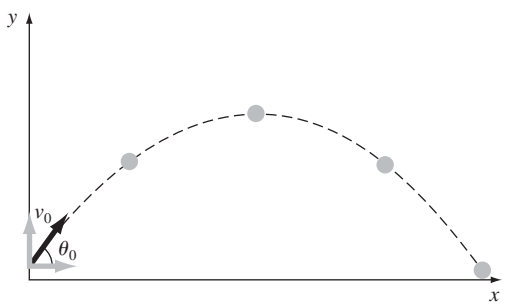


Figura P8.36

8.37 A velocidade para cima de um foguete pode ser calculada pela seguinte fórmula:

$$v = u\ln\frac{m_0}{m_0 - qt} - gt$$

onde v é a velocidade para cima, u é a velocidade na qual o combustível é expelido com relação ao foguete, m_0 é a massa inicial do foguete no instante $t = 0$, q é a taxa de consumo de combustível, e g é a aceleração da gravidade para baixo (suposta constante e igual a $= 9{,}81\ \text{m/s}^2$). Se $u = 2000$ m/s, $m_0 = 150.000$ kg, e $q = 2700$ kg/s, calcule o instante no qual a velocidade é igual a 750 m/s. (*Sugestão*: t está em algum ponto entre 10 e 50 s). Determine seu resultado de modo que ele esteja a menos de 1% do valor verdadeiro. Verifique sua resposta.

8.38 Na Seção 8.4, o ângulo de fase ϕ entre a vibração forçada causada pela rugosidade da estrada e o movimento do carro é dado por

$$\text{tg}\ \phi = \frac{2(c/c_c)(\omega/p)}{1 - (\omega/p)^2}$$

Como um engenheiro mecânico, você gostaria de saber se existem casos em que $\phi = \omega/3 - 1$. Use os outros parâmetros da seção para escrever a equação como um problema de raízes e determinar ω.

8.39 Dois fluidos em temperaturas diferentes entram em um misturador e saem à mesma temperatura. A capacidade térmica do fluido A é dada por

$$c_p = 3{,}381 + 1{,}804 \times 10^{-2}T - 4{,}300 \times 10^{-6}T^2$$

e a capacidade do fluido B é dada por:

$$c_p = 8{,}592 + 1{,}290 \times 10^{-1}T - 4{,}078 \times 10^{-5}T^2$$

onde c_p está em unidades de cal/mol K, e T está em unidades de K. Observe que

$$\Delta H = \int_{T_1}^{T_2} c_p dT$$

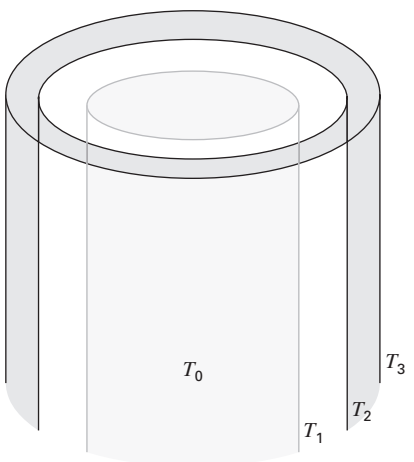
A entra no misturador a 400° C. B entra no misturador a 700° C. Há duas vezes mais de A do que de B entrando no misturador. A qual temperatura os dois fluidos saem do misturador?

8.40 Um compressor está operando à taxa de compressão R_c de 3,0 (a pressão de gás na saída é três vezes maior do que a pressão de gás na entrada). Os requisitos de potência do compressor H_p podem ser determinados a partir da equação a seguir. Supondo que os requisitos de potência do compressor são exatamente iguais a zRT_1/MW, encontre a eficiência politrópica n do compressor. O parâmetro z é a compressibilidade do gás nas condições de operação do compressor, R é a constante do gás, T_1 é a temperatura do gás na entrada do compressor e MW é o peso molecular do gás.

$$\text{HP} = \frac{zRT_1}{\text{MW}} \frac{n}{n-1}\left(R_c^{(n-1)/n} - 1\right)$$

8.41 Na garrafa térmica mostrada na Figura P8.41, o compartimento mais interno é separado do compartimento do meio por vácuo. Existe uma camada final ao redor da garrafa térmica. Essa camada final é separada da camada do meio por uma fina camada de ar. O lado de fora da última camada entra em contato com o ar ambiente. A transferência de calor do compartimento interno para a próxima camada q_1 se dá apenas por radiação (já que existe vácuo nesse espaço). A transferência de calor entre a camada do meio e a camada externa q_2 se dá por convecção em um pequeno espaço. A transferência da camada externa para o ar q_3 se dá por convecção natural. O fluxo de calor em cada região da garrafa térmica deve ser igual — isto é, $q_1 = q_2 = q_3$. Encontre as temperaturas T_1 e T_2 no estado estacionário. T_0 é 450° C e $T_3 = 25$° C.

Figura P8.41



$$q_1 = 10^{-9}[(T_0 + 273)^4 - (T_1 + 273)^4]$$
$$q_2 = 4(T_1 - T_2)$$
$$q_3 = 1{,}3(T_2 - T_3)^{4/3}$$

8.42 A forma geral para um campo de tensões tridimensional é dada por

$$\begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix}$$

onde os termos da diagonal representam as tensões de compressão ou dilatação e os termos fora da diagonal representam a tensão de cisalhamento. Um campo de tensões (em MPa) é dado por

$$\begin{bmatrix} 10 & 14 & 25 \\ 14 & 7 & 15 \\ 25 & 15 & 16 \end{bmatrix}$$

Para determinar as tensões principais, é necessário construir a seguinte matriz (novamente, em MPa):

$$\begin{bmatrix} 10-\sigma & 14 & 25 \\ 14 & 7-\sigma & 15 \\ 25 & 15 & 16-\sigma \end{bmatrix}$$

σ_1, σ_2 e σ_3 podem ser determinados a partir da equação

$$\sigma^3 - I\sigma^2 + II\sigma - III = 0$$

onde

$$I = \sigma_{xx} + \sigma_{yy} + \sigma_{zz}$$
$$II = \sigma_{xx}\sigma_{yy} + \sigma_{xx}\sigma_{zz} + \sigma_{yy}\sigma_{zz} - \sigma_{xy}^2 - \sigma_{xz}^2 - \sigma_{yz}^2$$
$$II = \sigma_{xx}\sigma_{yy}\sigma_{zz} - \sigma_{xx}\sigma_{yz}^2 - \sigma_{yy}\sigma_{xz}^2 - \sigma_{zz}\sigma_{xy}^2 + 2\sigma_{xy}\sigma_{xz}\sigma_{yz}$$

I, *II*, e *III* são conhecidos como invariantes da tensão. Encontre σ_1, σ_2 e σ_3 usando uma técnica para determinar raízes.

8.43 A Figura P8.43 mostra três reservatórios ligados por tubos circulares. Os tubos, que são feitos de ferro fundido recoberto por asfalto ($\varepsilon = 0{,}0012$ m), têm as seguintes características:

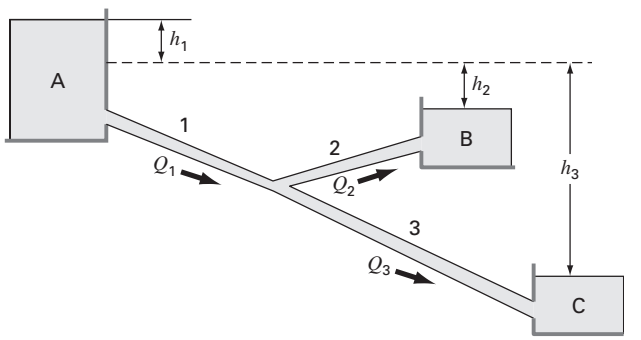


Figura P8.43

Tubo	1	2	3
Comprimento, m	1800	500	1400
Diâmetro, m	0,4	0,25	0,2
Escoamento, m^3/s	?	0,1	?

Se as elevações da superfície da água nos reservatórios A e C são 200 e 172,5 m, respectivamente, determine a elevação no reservatório B e o escoamento nos tubos 1 e 3. Observe que a viscosidade cinemática da água é 1×10^{-6} m^2/s e use a equação de Colebrook para determinar o fator de atrito (lembre-se do Problema 8.12).

8.44 Um fluido é bombeado na rede de tubos mostrada na Figura P8.44. No estado estacionário, os seguintes balanços de escoamento precisam ser satisfeitos:

$$Q_1 = Q_2 + Q_3$$
$$Q_3 = Q_4 + Q_5$$
$$Q_5 = Q_6 + Q_7$$

onde Q_i é o escoamento no tubo i [m^3/s]. Além disso, as quedas de pressão em torno dos três laços à direita devem ser nulas. A queda de pressão em cada seção de tubo circular pode ser calculada por

$$\Delta P = \frac{16}{\pi^2}\frac{fL\rho}{2D^5}Q^2$$

onde ΔP é a queda de pressão [Pa], f é o fator de atrito (adimensional), L é o comprimento do tubo [m], ρ é a densidade do fluido [kg/m^3] e D é o diâmetro do tubo (m). Escreva um programa (ou desenvolva um algoritmo em um pacote de software matemático) que lhe permita calcular o escoamento em todas as seções de tubo dado que $Q_1 = 1$ m^3/s e $\rho = 1{,}23$ kg/m^3. Todos os tubos têm $D = 500$ mm e $f = 0{,}005$. Os comprimentos dos tubos são $L_3 = L_5 = L_8 = L_9 = 2$ m; $L_2 = L_4 = L_6 = 4$ m e $L_7 = 8$ m.

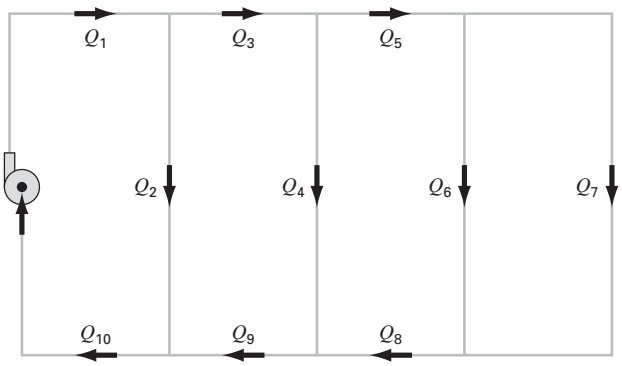


Figura P8.44

8.45 Repita o Problema 8.44, mas incorpore o dado de que o fator de atrito pode ser calculado com a *equação de von Karman,*

$$\frac{1}{\sqrt{f}} = 4\log_{10}(\mathrm{Re}\sqrt{f}) - 0{,}4$$

onde Re é o número de Reynolds

$$\mathrm{Re} = \frac{\rho V D}{\mu}$$

onde V é a velocidade do fluido no tubo [m/s] e μ é a viscosidade dinâmica (N · s/m^2). Observe que, para um tubo circular, $V = 4Q/\pi D^2$. Além disso, suponha que o fluido tem viscosidade de $1{,}79 \times 10^{-5}$ N · s/m^2.

8.46 O ônibus espacial, no lançamento a partir de sua plataforma de lançamento, tem quatro forças agindo sobre ele, as quais são mostradas no diagrama de corpo livre (Figura P8.46). O peso combinado dos dois foguetes auxiliares sólidos e do tanque de combustível externo é $W_B = 1{,}663 \times 10^6$ lb. O peso da nave espacial com carga total é $W_S = 0{,}23 \times 10^6$ lb. O empuxo combinado dos dois foguetes auxiliares sólidos é $T_B = 5{,}30 \times 10^6$ lb. O empuxo combinado dos três motores de combustível líquido do módulo orbital é $T_S = 1{,}125 \times 10^6$ lb.

No lançamento, o empuxo do motor do módulo orbital está direcionado a um ângulo θ para fazer com que o torque resultante na montagem total da espaçonave (tanque externo, foguetes auxiliares sólidos e módulo orbital) seja igual a zero. Com o torque resultante igual a zero, a espaçonave não irá girar em torno de seu centro de massa G no lançamento. Com essas forças, a espaçonave terá uma força resultante com componentes tanto na direção vertical quanto na horizontal. A componente vertical da força resultante é a que permite que a espaçonave seja lançada de sua plataforma de lançamento e que voe verticalmente. A componente horizontal da força resultante faz com que a espaçonave voe horizontalmente. O torque resultante agindo na espaçonave será nulo quando θ for ajustado no valor adequado. Se esse ângulo não for ajustado adequadamente e existir algum torque resultante agindo na espaçonave, ela tenderá a girar em torno de seu centro de massa.

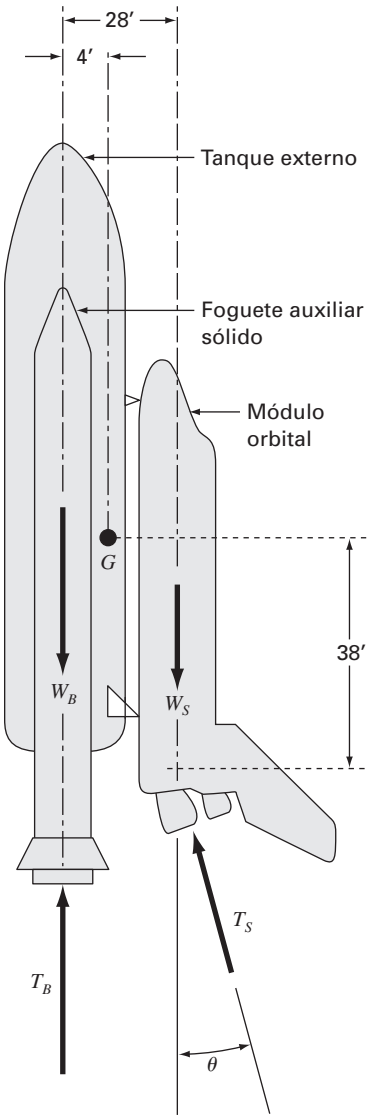


Figura P8.46

(a) Decomponha o empuxo do módulo orbital T_S em suas componentes horizontal e vertical e então some os torques relativos ao ponto G, o centro de massa da espaçonave. Iguale a equação do torque resultante a zero. Pode-se agora resolver essa equação para determinar o valor de θ necessário para o lançamento.

(b) Deduza uma equação para o torque resultante agindo na espaçonave em termos do ângulo θ. Trace o gráfico do momento resultante como função do ângulo θ sobre um intervalo de -5 radianos a $+5$ radianos.

(c) Escreva um programa computacional para determinar o ângulo θ usando o método de Newton para encontrar a raiz da equação do momento resultante. Escolha uma primeira aproximação inicial para a raiz de interesse a partir do gráfico. Pare suas iterações quando o valor de θ tiver mais que cinco algarismos significativos.

(d) Repita o programa para um peso mínimo de carga extra no módulo orbital de $W_S = 195.000$lb.



EPÍLOGO: PARTE DOIS

PT2.4 PRÓS E CONTRAS

A Tabela PT2.3 fornece um resumo dos prós e contras envolvidos na determinação de raízes de equações algébricas e transcendentais. Embora os métodos gráficos sejam demorados, eles fornecem uma visualização do comportamento da função e são úteis na identificação de aproximações iniciais e possíveis problemas como raízes múltiplas. Portanto, se o tempo permitir, um esboço rápido (ou, melhor ainda, um gráfico computadorizado) pode fornecer informações valiosas relativas ao comportamento da função.

Os próprios métodos numéricos estão divididos em duas categorias gerais: os métodos intervalares e os métodos abertos. Os primeiros exigem duas aproximações iniciais que estão uma de cada lado de uma raiz.

Esse fato de delimitar a raiz é mantido à medida que a solução prossegue, e portanto, essas técnicas são sempre convergentes. Entretanto, paga-se um preço por essa propriedade, já que a taxa de convergência é relativamente lenta.

As técnicas abertas diferem dos métodos intervalares no fato de que usam informação em um único ponto (ou dois valores que não precisam delimitar a raiz para extrapolar para uma nova estimativa da raiz). Tal propriedade é uma faca de dois gumes,

TABELA PT2.3 Comparação dos métodos alternativos para determinar raízes de equações algébricas e transcendentais. As comparações são baseadas na experiência geral e não levam em conta o comportamento de funções específicas.

Método	Aproximações Iniciais	Taxa de Convergência	Estabilidade	Acurácia	Amplitude de Aplicação	Esforço de Programação	Comentários
Direto	—	—	—	—	Limitada		
Gráfico	—	—	—	Ruim	Raízes reais	—	Pode levar mais tempo do que os métodos numéricos
Bissecção	2	Lenta	Sempre	Boa	Raízes reais	Fácil	
Falsa posição	2	Lenta/média	Sempre	Boa	Raízes reais	Fácil	
FP modificado	2	Média	Sempre	Boa	Raízes reais	Fácil	
Iteração de ponto fixo	1	Lenta	Possível divergência	Boa	Geral	Fácil	
Newton-Raphson	1	Rápida	Possível divergência	Boa	Geral	Fácil	Exige o cálculo de $f'(x)$
Newton-Raphson modificado	1	Rápida para raízes múltiplas; média para simples	Possível divergência	Boa	Geral	Fácil	Exige o cálculo de $f''(x)$ e $f'(x)$
Secante	2	Média a rápida	Possível divergência	Boa	Geral	Fácil	As aproximações iniciais não precisam cercar a raiz
Secante modificado	1	Média a rápida	Possível divergência	Boa	Geral	Fácil	
Müller	2	Média a rápida	Possível divergência	Boa	Polinômios	Moderado	
Bairstow	2	Rápida	Possível divergência	Boa	Polinômios	Moderado	

pois, embora leve a uma convergência mais rápida, também permite que a solução possa divergir. Em geral, a convergência nas técnicas abertas depende, em parte, da qualidade da aproximação inicial e da natureza da função. Quanto mais próxima da raiz verdadeira a aproximação estiver, mais provável será que os métodos convirjam.

Das técnicas abertas, o método de Newton-Raphson padrão é em geral usado por causa de sua propriedade de convergência quadrática. Entretanto, sua maior deficiência é que ele exige que a derivada da função seja calculada analiticamente. Para algumas funções, isso não é prático. Nesses casos, o método da secante, que usa uma representação por diferença finita da derivada, fornece uma alternativa viável. Por causa da aproximação por diferença finita, a taxa de convergência do método da secante é inicialmente mais lenta do que a do método de Newton-Raphson. Entretanto, à medida que a estimativa da raiz é refinada, a aproximação por diferença se torna uma representação melhor da derivada verdadeira e a convergência acelera rapidamente. A técnica de Newton-Raphson modificada pode ser usada para atingir uma convergência rápida no caso de raízes múltiplas. Mas tal técnica exige a expressão analítica para a primeira e a segunda derivadas.

Todos os métodos numéricos são fáceis de programar no computador e gastam um tempo mínimo para determinar uma única raiz. Com base nisso, você poderia concluir que um método simples como a bissecção seria suficientemente bom para os propósitos práticos. Isso seria verdade se você estivesse interessado exclusivamente em determinar a raiz de uma equação uma única vez. Entretanto, há muitos casos em engenharia nos quais muitas localizações de raízes são necessárias e a velocidade se torna importante. Nesses casos, os métodos lentos gastam muito tempo e são, portanto, caros. Por outro lado, os métodos abertos rápidos podem divergir e os conseqüentes atrasos também podem ser caros. Alguns algoritmos computacionais tentam aproveitar as vantagens de ambas as classes de técnicas, utilizando inicialmente um método intervalar para se aproximar dela e então mudando para um método aberto para refinar rapidamente a estimativa. Se uma única abordagem ou uma combinação for usada, os prós e contras entre a convergência e a velocidade estão no cerne da escolha da técnica de localização de raízes.

PT2.5 RELAÇÕES E FÓRMULAS IMPORTANTES

A Tabela PT2.4 resume informações importantes que foram apresentadas na Parte Dois. Ela pode ser consultada para acessar rapidamente relações e fórmulas importantes.

PT2.6 MÉTODOS AVANÇADOS E REFERÊNCIAS ADICIONAIS

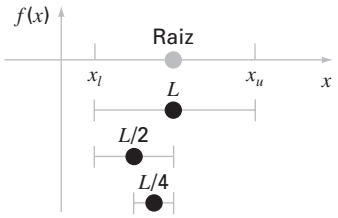
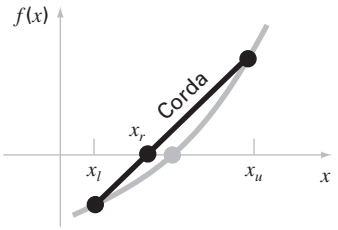
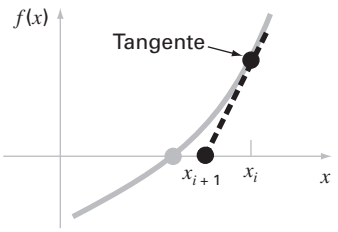
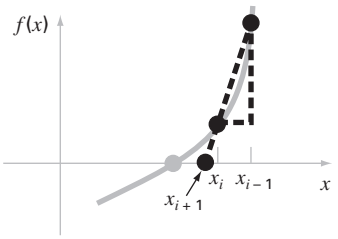
Os métodos deste texto se concentraram na determinação de uma única raiz de uma equação algébrica ou transcendental com base no conhecimento prévio de sua localização aproximada. Além disso, também foram descritos métodos construídos especificamente para determinar as raízes reais e complexas de polinômios. Referências adicionais sobre o assunto são Ralston e Rabinowitz (1978) e Carnahan, Luther e Wilkes (1969).

Além dos métodos de Müller e Bairstow, diversas técnicas estão disponíveis para determinar todas as raízes de polinômios. Em particular, o *algoritmo da diferença do quociente* (Henrici, 1964, e Gerald e Wheatley, 1989) determina todas as raízes sem aproximações iniciais. Ralston e Rabinowitz (1978) e Carnahan, Luther e Wilkes (1969) contêm discussões desse método, bem como de outras técnicas para localizar raízes de polinômios. Como discutido no texto, os métodos de *Jenkins-Traub* e *Laguerre* são amplamente usados.

Em resumo, a discussão anterior tem o objetivo de mostrar um caminho para uma exploração mais profunda do assunto. Além disso, todas as referências citadas fornecem descrições das técnicas básicas cobertas na Parte Dois. Nós o encorajamos a consultar essas fontes alternativas para ampliar seu entendimento dos métodos numéricos para a localização de raízes.[1]

[1] Os livros são citados aqui apenas pelos autores, uma bibliografia completa é encontrada no final do texto.

TABELA PT2.4 Resumo das informações importantes apresentadas na Parte Dois.

Método	Formulação	Interpretação Gráfica	Erros e Critério de Parada
		Métodos intervalares:	
Bissecção	$x_r = \frac{x_l + x_u}{2}$ Se $f(x_l)f(x_r) < 0$, $x_u = x_r$ $f(x_l)f(x_r) > 0$, $x_l = x_r$		Critério de parada: $\left\lvert \frac{x_r^{\text{novo}} - x_r^{\text{velho}}}{x_r^{\text{novo}}} \right\rvert 100\% \leq \epsilon_s$
Falsa posição	$x_r = x_u - \frac{f(x_u)(x_l - x_u)}{f(x_l) - f(x_u)}$ Se $f(x_l)f(x_r) < 0$, $x_u = x_r$ $f(x_l)f(x_r) > 0$, $x_l = x_r$		Critério de parada: $\left\lvert \frac{x_r^{\text{novo}} - x_r^{\text{velho}}}{x_r^{\text{novo}}} \right\rvert 100\% \leq \epsilon_s$
Newton-Raphson	$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$		Critério de parada: $\left\lvert \frac{x_{i+1} - x_i}{x_{i+1}} \right\rvert 100\% \leq \epsilon_s$ Erro: $E_{i+1} = 0(E_i^2)$
Secante	$x_{i+1} = x_i - \frac{f(x_i)(x_{i-1} - x_i)}{f(x_{i-1}) - f(x_i)}$		Critério de parada: $\left\lvert \frac{x_{i+1} - x_i}{x_{i+1}} \right\rvert 100\% \leq \epsilon_s$

PARTE TRÊS



EQUAÇÕES ALGÉBRICAS LINEARES

PT3.1 MOTIVAÇÃO

Na Parte Dois, foi determinado o valor de x que satisfaz uma única equação, $f(x) = 0$. Agora, será abordado o caso de determinar os valores $x_1, x_2, \ldots, x_n$ que satisfazem simultaneamente um conjunto de equações

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0 \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}$$

Tais sistemas podem ser ou lineares ou não-lineares. Na Parte Três, tratamos das *equações algébricas lineares* que têm a forma geral

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \qquad \text{(PT3.1)}$$

onde os a's são coeficientes constantes, os b's são constantes e n é o número de equações. Todas as outras equações são não-lineares. Os sistemas não-lineares foram discutidos no Capítulo 6 e serão de novo abordados brevemente no Capítulo 9.

PT 3.1.1 Métodos Não-computacionais para Resolver Sistemas de Equações

Para um número pequeno de equações ($n \leq 3$), as equações lineares (e, às vezes, as não-lineares) podem ser resolvidas rapidamente por técnicas simples. Alguns desses métodos serão revistos no Capítulo 9. Entretanto, para quatro ou mais equações, as soluções se tornam trabalhosas e os computadores devem ser utilizados. Historicamente, a impossibilidade de resolver todos os conjuntos de equações — exceto os menores — à mão limitou o alcance de problemas tratados em várias aplicações da engenharia.

Antes dos computadores, as técnicas para resolver equações algébricas lineares eram lentas e desconfortáveis. Essas abordagens restringem a criatividade, pois os métodos são, em geral, difíceis de implementar e entender. Conseqüentemente, as técnicas eram, algumas vezes, superenfatizadas à custa de outros aspectos do processo de resolução do problema, como a formulação e a interpretação (lembre-se da Figura PT1.1 e da discussão que a acompanha).

O advento de computadores facilmente acessíveis tornou possível e prática a resolução de grandes conjuntos de equações algébricas lineares simultâneas. E, portanto, você pode abordar exemplos e problemas mais realistas e complexos. Além disso, você terá mais tempo para testar suas habilidades criativas, pois poderá pôr mais ênfase na formulação do problema e na interpretação da solução.

PT3.1.2 Equações Algébricas Lineares e a Prática da Engenharia

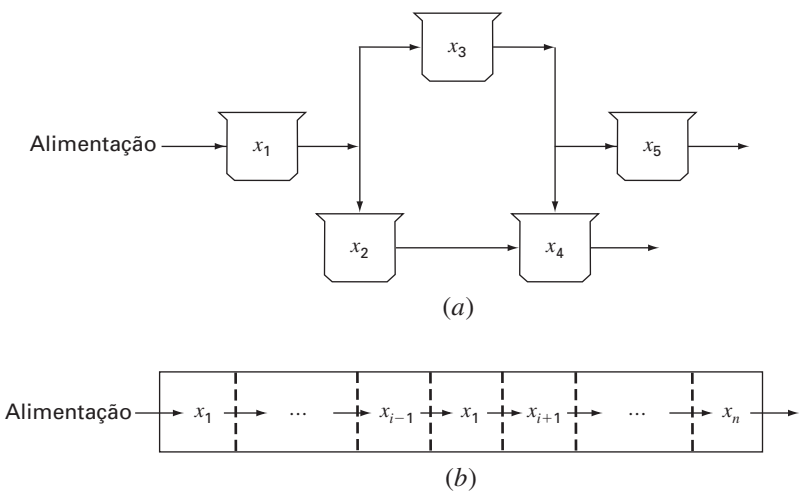
Muitas das equações fundamentais da engenharia são baseadas em leis de conservação (lembre-se da Tabela 1.1). Algumas quantidades familiares que se submetem a tais leis são massa, energia e momento. Em termos matemáticos, esses princípios levam a equações de continuidade ou de balanço que relacionam o *comportamento* do sistema, representado por seus *níveis* ou *respostas* das quantidades que estão sendo modeladas, às *propriedades* ou *características* do sistema e aos *estímulos* externos ou *termos forçantes* agindo sobre o sistema.

Como um exemplo, o princípio de conservação de massa pode ser usado para formular um modelo para uma série de reatores químicos (Figura PT3.1*a*). Para esse caso, a quantidade sendo modelada é a massa do composto químico em cada reator. As propriedades do sistema são as características da reação do composto químico e o tamanho dos reatores e as vazões. Os termos forçantes são as taxas de alimentação dos compostos químicos no sistema.

Na Parte Dois, você viu como os sistemas com uma única componente resultam em uma única equação que pode ser resolvida usando-se as técnicas de localização de raízes. Sistemas com componentes múltiplas resultam em um conjunto de equações matemáticas acopladas, que devem ser resolvidas simultaneamente. As equações são acopladas porque as partes individuais do sistema são influenciadas pelas outras partes. Por exemplo, na Figura PT3.1*a* o reator 4 recebe entradas dos reatores 2 e 3. Conseqüentemente, sua resposta depende das quantidades dos compostos químicos nesses outros reatores.

Quando essas dependências são expressas matematicamente, as equações resultantes são, muitas vezes, da forma algébrica linear da Equação (PT3.1). Os x's são em geral medidas das intensidades das respostas de componentes individuais. Usando a Figura PT3.1*a* como um exemplo, x_1 poderia indicar a quantidade de massa no primeiro reator, x_2 poderia indicar a quantidade no segundo e assim por diante. Os a's representam tipicamente as propriedades e características que produzem as interações entre as componentes. Por exemplo, os a's para a Figura PT3.1*a* poderiam refletir as vazões em massa entre os reatores. Finalmente, os b's representam os termos forçantes agindo sobre o sistema, como a taxa de alimentação na Figura PT3.1*a*. As aplicações no Capítulo 12 fornecem outros exemplos de tais equações vindas da prática da engenharia.

FIGURA PT3.1
Dois tipos de sistemas que podem ser modelados usando equações algébricas lineares: (*a*) sistemas de variáveis agrupadas que envolvem componentes finitas acopladas e (*b*) sistemas de variáveis distribuídas que envolvem um contínuo.



Problemas com várias componentes do tipo citado surgem de modelos matemáticos de variáveis tanto *agrupadas* (macro) quanto *distribuídas* (micro) (Figura PT3.1). Problemas de variáveis agrupadas envolvem componentes finitas acopladas. Os exemplos incluem treliças (Seção 12.2), reatores (Figura PT3.1*a* e Seção 12.1) e circuitos elétricos (Seção 12.3). Esses tipos de problemas usam modelos que fornecem pouco ou nenhum detalhe espacial.

Reciprocamente, problemas de variáveis distribuídas tentam descrever detalhes espaciais de sistemas em uma base contínua ou semicontínua. A distribuição de compostos químicos ao longo do comprimento de um reator retangular alongado (Figura PT3.1*b*) é um exemplo de um modelo de variável contínua. As equações diferenciais deduzidas das leis de conservação especificam a distribuição das variáveis dependentes para tais sistemas. Essas equações diferenciais podem ser resolvidas numericamente transformando-as em um sistema equivalente de equações algébricas simultâneas. A solução de tais conjuntos de equações representa uma área de aplicação importante na engenharia para os métodos nos capítulos seguintes. Essas equações são acopladas, pois as variáveis em uma posição dependem das variáveis nas regiões adjacentes. Por exemplo, a concentração no meio do reator é uma função da concentração nas regiões adjacentes. Exemplos similares podem ser desenvolvidos para distribuições espaciais de temperatura ou momento. Iremos tratar de tais problemas quando discutirmos as equações diferenciais mais adiante no livro.

Além dos sistemas físicos, as equações algébricas lineares simultâneas também aparecem em uma variedade de contextos de problemas matemáticos. Elas ocorrem quando exigimos que funções matemáticas satisfaçam várias condições simultâneas. Cada condição resulta em uma equação que tem coeficientes conhecidos e variáveis indeterminadas. As técnicas discutidas nesta parte podem ser utilizadas para determinar as incógnitas quando as equações forem lineares e algébricas. Algumas técnicas numéricas amplamente utilizadas baseadas em equações simultâneas são a análise de regressão (Capítulo 17) e a interpolação por splines (Capítulo 18).

PT3.2 FUNDAMENTOS MATEMÁTICOS

Todas as partes deste livro requerem algum fundamento matemático. Para a Parte Três, a notação e a álgebra matricial são úteis, pois fornecem uma forma concisa de representar e manipular equações algébricas lineares. Se você já estiver familiarizado com matrizes, sinta-se livre para pular para a Seção PT3.3. Para aqueles que não estão familiarizados ou querem uma revisão, o material a seguir fornece uma breve introdução ao assunto.

PT3.2.1 Notação Matricial

Uma *matriz* consiste em uma tabela retangular de elementos representados por um único símbolo. Como mostrado na Figura PT3.2, [A] é a notação abreviada para a matriz e a_{ij} designa um *elemento* individual da matriz.

Um conjunto horizontal de elementos é chamado de *linha* e um conjunto vertical é chamado de *coluna*. O primeiro índice i sempre representa o número da linha no qual o

FIGURA PT3.2
Uma matriz.

Coluna 3

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2m} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nm} \end{bmatrix}$$

Linha 2

elemento está. O segundo índice j designa a coluna. Por exemplo, o elemento a_{23} está na linha 2 e coluna 3.

A matriz na Figura PT3.2 tem n linhas e m colunas e dizemos que tem dimensão n por m (ou $n \times m$). Ela é chamada de uma matriz n por m.

As matrizes com número de linhas $n = 1$, tais como

$$[B] = [b_1 \quad b_2 \quad \cdots \quad b_m]$$

são chamadas *vetores linha.* Observe que, por simplicidade, o primeiro índice de cada elemento foi omitido. Também se deve mencionar que, alguma vezes, é desejável empregar uma notação abreviada especial para diferenciar uma matriz linha de outros tipos de matrizes. Uma forma de fazer isso é utilizar colchetes abertos em cima, como em $\lfloor B \rfloor$.

As matrizes com número de colunas $m = 1$, tais como

$$[C] = \begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{bmatrix}$$

são chamadas de *vetores coluna*. Por simplicidade, o segundo índice é omitido. Como nos vetores linha, há ocasiões nas quais é desejável empregar uma notação abreviada especial para diferenciar uma matriz coluna de outros tipos de matrizes. Uma forma de fazer isso é usar chaves, como em $\{C\}$.

As matrizes em que $n = m$ são chamadas *matrizes quadradas*. Por exemplo, uma matriz 4 por 4 é

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

A diagonal consistindo nos elementos a_{11}, a_{22}, a_{33} e a_{44} é chamada *diagonal principal* da matriz.

As matrizes quadradas são particularmente importantes quando resolvemos conjuntos de equações lineares simultâneas. Para tais sistemas, o número de equações (correspondentes às linhas) e o número de incógnitas (correspondentes às colunas) devem ser iguais para que seja possível uma solução única. Conseqüentemente, são encontradas matrizes de coeficientes quadradas quando lidamos com tais sistemas. Alguns tipos especiais de matrizes quadradas são descritas no Quadro PT3.1.

PT3.2.2 Regras de Operações Matriciais

Agora que foi especificado o que queremos dizer por matriz, pode-se definir certas operações que governam o seu uso. Duas matrizes n por m são iguais se, e somente se, todo elemento na primeira é igual a todo elemento na segunda, isto é, $[A] = [B]$ se $a_{ij} = b_{ij}$ para todo i e j.

A *adição* de duas matrizes, digamos $[A]$ e $[B]$, é feita somando-se os termos correspondentes em cada matriz. Os elementos da matriz resultante $[C]$ são calculados por

$$c_{ij} = a_{ij} + b_{ij}$$

para $i = 1, 2, \ldots, n$ e $j = 1, 2, \ldots, m$. Analogamente, a *subtração* de duas matrizes, digamos, $[E]$ menos $[F]$, é obtida subtraindo-se os termos correspondentes, como em

$$d_{ij} = e_{ij} - f_{ij}$$

para $i = 1, 2, \ldots, n$ e $j = 1, 2, \ldots, m$. Segue diretamente dessas definições que a adição e a subtração podem ser efetuadas apenas entre matrizes com as mesmas dimensões.

Quadro PT3.1 Tipos Especiais de Matrizes Quadradas

Há algumas formas especiais de matrizes quadradas que são importantes e devem ser observadas:

Uma ***matriz simétrica*** é aquela em que $a_{ij} = a_{ji}$ para todos os i's e j's. Por exemplo,

$$[A] = \begin{bmatrix} 5 & 1 & 2 \\ 1 & 3 & 7 \\ 2 & 7 & 8 \end{bmatrix}$$

é uma matriz 3 por 3 simétrica.

Uma *matriz diagonal* é uma matriz quadrada cujos elementos fora da diagonal principal são iguais a zero, como em

$$[A] = \begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & a_{33} & \\ & & & a_{44} \end{bmatrix}$$

Observe que, quando blocos grandes de elementos são nulos, eles são deixados em branco.

Uma *matriz-identidade* é uma matriz diagonal na qual todos os elementos da diagonal principal são iguais a 1, como em

$$[I] = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

O símbolo $[I]$ é usado para denotar a matriz identidade. A matriz identidade tem propriedades similares à unidade.

Uma *matriz triangular superior* é uma matriz em que todos os elementos abaixo da diagonal principal são zero, como em

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ & a_{22} & a_{23} & a_{24} \\ & & a_{33} & a_{34} \\ & & & a_{44} \end{bmatrix}$$

Uma *matriz triangular inferior* é uma matriz em que todos os elementos acima da diagonal principal são zero, como em

$$[A] = \begin{bmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \\ a_{31} & a_{32} & a_{33} & \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

Uma *matriz de banda* tem todos os elementos iguais a zero, com a possível exceção de uma faixa centrada na diagonal principal:

$$[A] = \begin{bmatrix} a_{11} & a_{12} & & \\ a_{21} & a_{22} & a_{23} & \\ & a_{32} & a_{33} & a_{34} \\ & & a_{43} & a_{44} \end{bmatrix}$$

Essa matriz possui largura da banda 3 e tem um nome especial — *matriz tridiagonal.*

Ambas, a adição e a subtração, são *comutativas*:

$$[A] + [B] = [B] + [A]$$

A adição e a subtração são também *associativas,* isto é,

$$([A] + [B]) + [C] = [A] + ([B] + [C])$$

A multiplicação de uma matriz $[A]$ por um escalar g é obtida multiplicando-se todo elemento de $[A]$ por g, como em

$$[D] = g[A] = \begin{bmatrix} ga_{11} & ga_{12} & \cdots & ga_{1m} \\ ga_{21} & ga_{22} & \cdots & ga_{2m} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ ga_{n1} & ga_{n2} & \cdots & ga_{nm} \end{bmatrix}$$

O *produto* de duas matrizes é representado por $[C] = [A][B]$, em que os elementos de $[C]$ são definidos por (ver Quadro PT3.2 para uma forma simples de visualizar a multiplicação de matrizes)

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{PT3.2}$$

onde n é o número de colunas de $[A]$ e o número de linhas de $[B]$. Isto é, o elemento c_{ij} é obtido somando-se o produto dos elementos individuais da i-ésima linha da primeira matriz, no caso $[A]$, com a j-ésima coluna da segunda matriz, $[B]$.

Quadro PT3.2 Um Método Simples para Multiplicar Duas Matrizes

Apesar de a Equação (PT3.2) ser bem apropriada para ser implementada em um computador, ela não é a forma mais simples de visualizar o mecanismo de multiplicação de duas matrizes. O que segue dá uma expressão mais concreta à operação.

Suponha que queremos multiplicar $[X]$ por $[Y]$ para obter $[Z]$,

$$[Z] = [X][Y] = \begin{bmatrix} 3 & 1 \\ 8 & 6 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 5 & 9 \\ 7 & 2 \end{bmatrix}$$

Uma forma simples de visualizar o cálculo de $[Z]$ é levantar $[Y]$, como em

$$\begin{bmatrix} & & \Uparrow & \\ & & \begin{bmatrix} 5 & 9 \\ 7 & 2 \end{bmatrix} & \leftarrow [Y] \\ [X] \rightarrow & \begin{bmatrix} 3 & 1 \\ 8 & 6 \\ 0 & 4 \end{bmatrix} & \begin{bmatrix} ? \end{bmatrix} & \leftarrow [Z] \end{bmatrix}$$

Agora, a resposta $[Z]$ pode ser calculada no espaço vago deixado por $[Y]$. Essa forma tem utilidade já que alinha as linhas e as colunas apropriadas que devem ser multiplicadas. Por exemplo, de acordo com a Equação (PT3.2), o elemento z_{11} é obtido multiplicando-se a primeira linha de $[X]$ pela primeira coluna de $[Y]$. Isso significa somar o produto de x_{11} e y_{11} ao produto de x_{12} e y_{21}, como em

$$\begin{bmatrix} & & \begin{bmatrix} \mathbf{5} & 9 \\ \mathbf{7} & 2 \end{bmatrix} \\ & & \downarrow \\ \begin{bmatrix} \mathbf{3} & \mathbf{1} \\ 8 & 6 \\ 0 & 4 \end{bmatrix} & \rightarrow & \begin{bmatrix} \mathbf{3 \times 5 + 1 \times 7 = 22} \\ \\ \end{bmatrix} \end{bmatrix}$$

Então, z_{11} é igual a 22. O elemento z_{21} pode ser calculado de maneira similar, como em

$$\begin{bmatrix} & & \begin{bmatrix} \mathbf{5} & 9 \\ \mathbf{7} & 2 \end{bmatrix} \\ & & \downarrow \\ & & 22 \\ \begin{bmatrix} \mathbf{3} & \mathbf{1} \\ 8 & 6 \\ 0 & 4 \end{bmatrix} & \rightarrow & \begin{bmatrix} \mathbf{8 \times 5 + 6 \times 7 = 82} \\ \\ \end{bmatrix} \end{bmatrix}$$

O cálculo continua dessa forma, seguindo o alinhamento das linhas e colunas, fornecendo o resultado

$$[Z] = \begin{bmatrix} 22 & 29 \\ 82 & 84 \\ 28 & 8 \end{bmatrix}$$

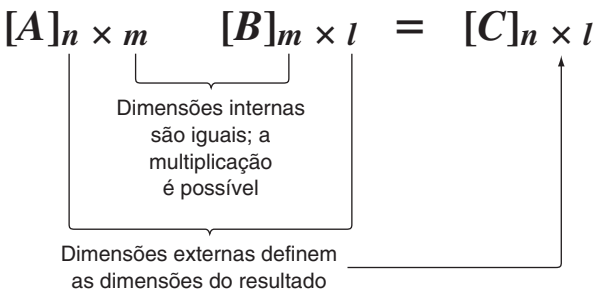
Observe como esse método simples torna claro por que é impossível multiplicar duas matrizes se o número de colunas da primeira não for igual ao número de linhas da segunda. Além disso, observe como ele ilustra que a ordem da multiplicação é importante (isto é, a multiplicação de matrizes não é comutativa).

De acordo com essa definição, a multiplicação de matrizes pode ser efetuada apenas quando a primeira matriz tem tantas colunas quanto o número de linhas da segunda matriz. Então, se $[A]$ é uma matriz n por m, $[B]$ pode ser uma matriz m por l. Nesse caso, a matriz resultante [C] terá dimensão n por l. Entretanto, se $[B]$ fosse uma matriz l por m, a multiplicação não poderia ser efetuada. A Figura PT3.3 fornece uma forma fácil de verificar quando duas matrizes podem ser multiplicadas.

Se as dimensões das matrizes forem adequadas, a multiplicação de matrizes será *associativa,*

$$([A][B])[C] = [A]([B][C])$$

FIGURA PT3.3



e *distributiva,*

$$[A]([B] + [C]) = [A][B] + [A][C]$$

ou

$$([A] + [B])[C] = [A][C] + [B][C]$$

Porém, em geral, a multiplicação não é *comutativa:*

$$[A][B] \neq [B][A]$$

Isto é, a ordem na multiplicação é importante.

A Figura PT3.4 mostra um pseudocódigo para multiplicar uma matriz [*A*] *n* por *m* por uma matriz [*B*] *m* por *l* e armazenar o resultado em uma matriz [C] *n* por *l.* Observe que, em vez de o produto interno ser diretamente acumulado em [C], ele é guardado em uma variável temporária, soma, o que é feito por duas razões. Primeiro, isso é um pouco mais eficiente, pois o computador precisa determinar a posição de $c_{i,j}$ apenas $n \times l$ vezes em vez de $n \times l \times m$ vezes. Segundo, a precisão da multiplicação pode ser bastante melhorada declarando-se a soma como uma variável de dupla precisão (lembre-se da discussão de produtos internos na Seção 3.4.2).

Apesar de a multiplicação ser possível, a divisão de matrizes não é uma operação definida. Porém, se a matriz [*A*] for quadrada e não singular, existe uma outra matriz $[A]^{-1}$, chamada a *inversa* de [*A*], tal que

$$[A][A]^{-1} = [A]^{-1}[A] = [I] \tag{PT3.3}$$

Assim, a multiplicação de uma matriz pela sua inversa é o análogo da divisão, no sentido de que um número dividido por si mesmo é igual a 1. Isto é, a multiplicação de uma matriz pela sua inversa resulta na matriz identidade (lembre-se do Quadro PT3.1).

A inversa de uma matriz quadrada de dimensão 2 pode ser representada simplesmente por

$$[A]^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{PT3.4}$$

Fórmulas semelhantes para matrizes de dimensão maior são muito mais complexas. Algumas seções nos Capítulos 10 e 11 serão dedicadas a técnicas para usar métodos numéricos e o computador para calcular a inversa de tais matrizes.

Duas outras manipulações de matrizes que terão utilidade na nossa discussão são a transposta e o traço de uma matriz. A transposta de uma matriz envolve transformar suas linhas em colunas e suas colunas em linhas. Por exemplo, para a matriz 4×4,

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

FIGURA PT3.4

```
SUBROUTINE Mmult (a, b, c, m, n, l)
DOFOR i = 1, n
  DOFOR j = 1, l
    soma = 0.
    DOFOR k = 1, m
      soma = soma + a(i,k) · b(k,j)
    END DO
    c(i,j) = soma
  END DO
END DO
```

a transposta, denotada por $[A]^T$, é definida por

$$[A]^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ a_{13} & a_{23} & a_{33} & a_{43} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix}$$

Em outras palavras, o elemento a_{ij} da transposta é igual ao elemento a_{ji} da matriz original.

A *transposta* tem uma grande variedade de funções na álgebra das matrizes. Uma vantagem simples é que ela permite que uma coluna seja escrita como uma linha. Por exemplo, se

$$\{C\} = \begin{Bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{Bmatrix}$$

então

$$\{C\}^T = \lfloor c_1 \quad c_2 \quad c_3 \quad c_4 \rfloor$$

onde o sobrescrito T designa a transposta. Por exemplo, isso pode economizar espaço quando escrevemos um vetor coluna em um texto. Além disso, a transposta tem inúmeras aplicações matemáticas.

O *traço* de uma matriz é a soma dos elementos da diagonal principal. Ele é denotado por tr $[A]$ e é calculado por

$$\text{tr}\,[A] = \sum_{i=1}^{n} a_{ii}$$

O traço será utilizado na discussão sobre autovalores no Capítulo 27.

A última manipulação de matrizes que terá utilidade na nossa discussão é o *aumento*. A matriz é aumentada pela adição de uma coluna (ou colunas) à matriz original. Por exemplo, suponha que temos uma matriz de coeficientes:

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Você poderia querer aumentar essa matriz $[A]$ com a matriz identidade (lembre-se do Quadro PT3.1) para produzir uma matriz de dimensão 3 por 6:

$$[A] = \left[\begin{array}{ccc:ccc} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{array}\right]$$

Tal expressão tem utilidade quando é preciso efetuar um conjunto de operações idênticas em duas matrizes. Assim, pode-se efetuar as operações na matriz aumentada em vez de fazê-lo nas duas matrizes individuais.

PT3.2.3 Representando Equações Algébricas Lineares na Forma Matricial

Deve ficar claro que as matrizes fornecem uma notação concisa para representar equações lineares simultâneas. Por exemplo, a Equação (PT3.1) pode ser expressa por

$$[A]\{X\} = \{B\} \tag{PT3.5}$$

onde $[A]$ é a matriz quadrada n por n de coeficientes,

$$[A] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$\{B\}$ é o vetor coluna $n \times 1$ de constantes,

$$\{B\}^T = \lfloor b_1 \quad b_2 \quad \cdots \quad b_n \rfloor$$

e $\{X\}$ é o vetor coluna $n \times 1$ de incógnitas:

$$\{X\}^T = \lfloor x_1 \quad x_2 \quad \cdots \quad x_n \rfloor$$

Recorde a definição de multiplicação de matrizes [Equação (PT3.2) ou Quadro PT3.2] para se convencer de que as Equações (PT3.1) e (PT3.5) são equivalentes. Além disso, observe que a Equação (PT3.5) é uma multiplicação de matrizes válida, pois o número de colunas, n, da primeira matriz $[A]$ é igual ao número de linhas, n, da segunda matriz $\{X\}$.

Esta parte do livro é dedicada a resolver a Equação (PT3.5), para determinar $\{X\}$. Uma maneira formal de obter a solução usando álgebra de matrizes é multiplicar cada lado da equação pela inversa de $[A]$ para obter

$$[A]^{-1}[A]\{X\} = [A]^{-1}\{B\}$$

Como $[A]^{-1}[A]$ é igual à matriz-identidade, a equação se torna

$$\{X\} = [A]^{-1}\{B\} \tag{PT3.6}$$

Então, a equação foi resolvida e $\{X\}$ foi determinado. Esse é um outro exemplo de como a inversa desempenha um papel na álgebra de matrizes que é semelhante à divisão. Devemos observar que esse não é um método muito eficiente para resolver um sistema de equações. Portanto, outras abordagens são empregadas nos algoritmos numéricos. Entretanto, como discutido no Capítulo 10, a própria matriz inversa tem grande importância na análise de tais sistemas.

Finalmente, algumas vezes, será útil aumentar $[A]$ com $\{B\}$. Por exemplo, se $n = 3$, isso resulta na matriz de dimensões 3 por 4:

$$[A] = \left[\begin{array}{ccc:c} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array}\right] \tag{PT3.7}$$

Expressar as equações nessa forma é útil, pois diversas técnicas para resolver sistemas lineares executam operações idênticas na linha de coeficientes e na constante correspondente do lado direito da equação. Como mostrado na Equação (PT3.7), pode-se efetuar a manipulação uma única vez na linha da matriz aumentada em vez de efetuá-la na matriz dos coeficientes e no vetor coluna do lado direito.

PT3.3 ORIENTAÇÃO

Antes de prosseguir para os métodos numéricos, mais um pouco de orientação pode ser útil. A seção a seguir tem a intenção de fornecer uma visão geral do material discutido na Parte Três. Além disso, formulamos alguns objetivos para ajudá-lo a concentrar seus esforços quando estudar este material.

PT3.3.1 Escopo e Visão Geral

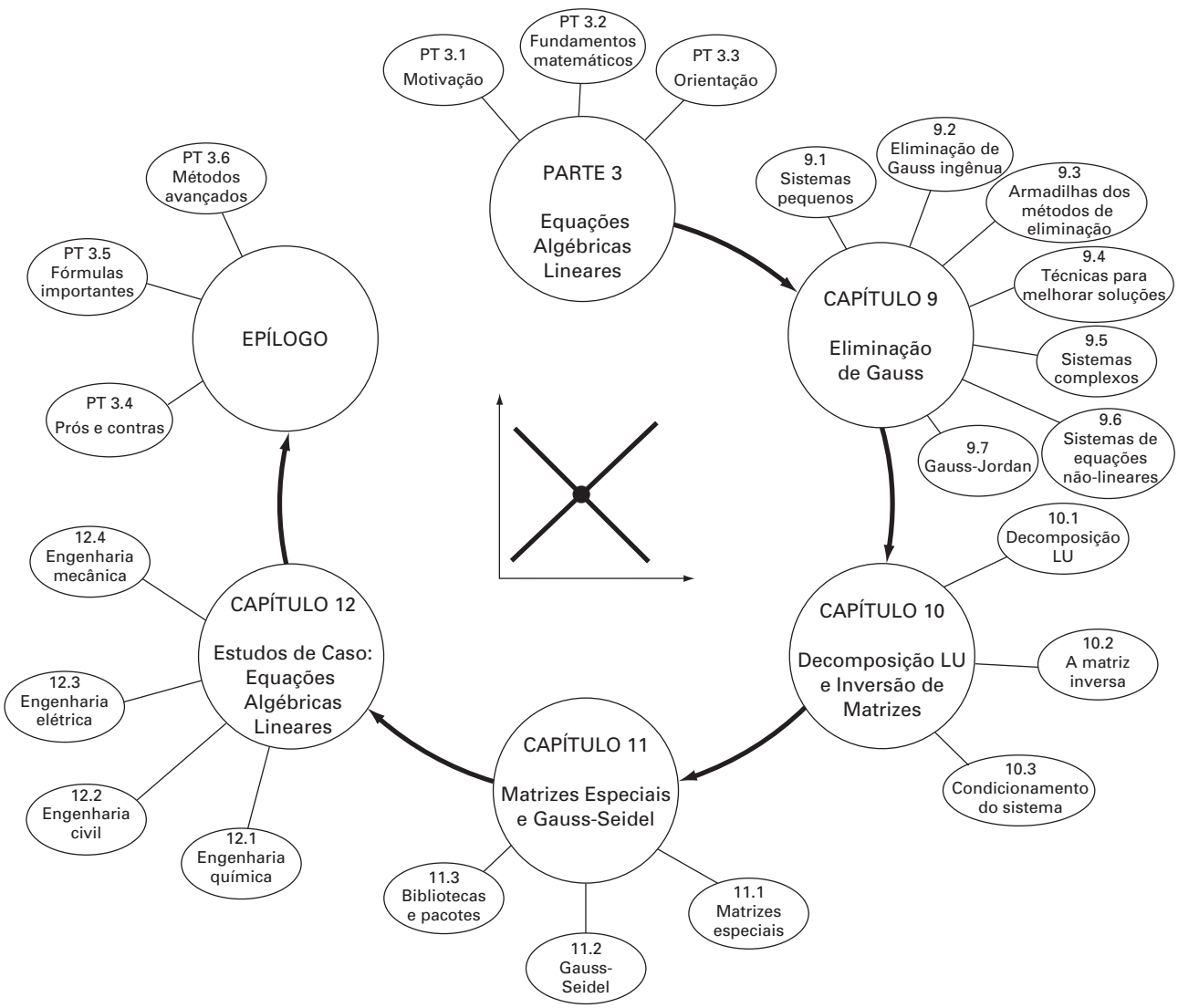
A Figura PT3.5 fornece uma visão geral da Parte Três. O **Capítulo 9** é dedicado à técnica mais fundamental para resolver sistemas algébricos lineares: *eliminação de Gauss.* Antes de uma discussão detalhada dessa técnica, uma seção preliminar trata dos métodos simples para resolver sistemas pequenos. Essas técnicas são apresentadas para fornecer uma

percepção visual e porque um dos métodos — a eliminação de incógnitas — representa a base da eliminação de Gauss.

Depois do material preliminar, a eliminação de Gauss "ingênua" é discutida. Começamos com essa versão simplificada porque ela permite que a técnica fundamental seja elaborada, sem detalhes complicados. Então, nas seções subseqüentes, serão discutidos problemas em potencial da abordagem ingênua e apresentadas diversas modificações pra minimizar e contornar esses problemas. O foco da discussão será o processo de permutar linhas, ou o *pivotamento parcial.*

O **Capítulo 10** começa ilustrando como a eliminação de Gauss pode ser formulada como uma *decomposição LU.* Tais técnicas de solução são valiosas para os casos em que muitos vetores do lado direito precisam sem calculados. É mostrado como esse atributo permite o cálculo eficiente da *matriz inversa,* o que tem tremenda utilidade na prática da engenharia. O capítulo termina com uma discussão do condicionamento de matrizes. O *número de condicionamento* é apresentado como uma medida da perda de algarismos significativos de acurácia que pode ocorrer ao se resolver matrizes mal condicionadas.

FIGURA PT3.5
Esquema da organização do material na Parte Três: Sistemas de Equações Algébricas Lineares.



O começo do **Capítulo 11** se concentra em tipos especiais de sistemas de equações que têm ampla gama de aplicações na engenharia. Em particular, são apresentadas técnicas eficientes para resolver *sistemas tridiagonais*. Então, o resto do capítulo se concentra em uma alternativa para os métodos de eliminação chamada *método de Gauss-Seidel.* Essa técnica é semelhante aos métodos de aproximação de raízes de equações que foram discutidos no Capítulo 6. Isto é, a técnica envolve escolher uma solução aproximada e então iterar para obter uma estimativa refinada. O capítulo termina com informações relacionadas à resolução de equações algébricas lineares com pacotes e bibliotecas de software.

O **Capítulo 12** mostra como os métodos podem de fato ser aplicados para resolver problemas. Como em outras partes do livro, as aplicações são extraídas de todos os campos da engenharia.

Finalmente, um epílogo é incluído no final da Parte Três. Essa revisão inclui uma discussão dos prós e contras relevantes para a implementação dos métodos na prática da engenharia. A seção também resume as fórmulas importantes e os métodos avançados relacionados às equações algébricas lineares. Assim, ela pode ser usada antes dos exames ou como lembrete depois que você se formar e precisar voltar às equações algébricas lineares como profissional.

PT3.3.2 Metas e Objetivos

Objetivos de Estudo. Após completar a Parte Três, você deve ser capaz de resolver problemas envolvendo equações algébricas lineares e apreciar a aplicação dessas equações em várias áreas da engenharia. Deve tentar dominar as diversas técnicas e avaliar sua confiabilidade. E deve entender os prós e contras envolvidos na escolha do "melhor" método (ou métodos) para qualquer problema particular. Além desses objetivos gerais, os conceitos específicos listados na Tabela PT3.1 precisam ser assimilados e dominados.

Objetivos Computacionais. Seu objetivo computacional mais importante é ser capaz de resolver um sistema de equações algébricas lineares e calcular a matriz inversa. Você certamente vai querer ter subprogramas desenvolvidos para decomposição *LU* de matrizes cheias ou tridiagonais. E também pode querer ter seu próprio software para implementar o método de Gauss-Seidel.

É necessário saber como usar pacotes para resolver equações algébricas lineares e determinar a matriz inversa. E se familiarizar com a forma pela qual os mesmos cálculos podem ser implementados com pacotes populares tais como o Excel ou o MATLAB, assim como com bibliotecas de software.

TABELA PT3.1 Objetivos específicos de estudo para a Parte Três.

1. Entender a interpretação gráfica dos sistemas mal condicionados e como isso se relaciona com o determinante.
2. Estar familiarizado com a terminologia: eliminação progressiva, substituição regressiva, equação pivô, coeficiente pivô.
3. Entender os problemas provenientes de divisão por zero, erros de arredondamento e mal condicionamento.
4. Saber calcular o determinante usando a eliminação de Gauss.
5. Entender as vantagens do pivotamento; perceber a diferença entre o pivotamento parcial e o completo.
6. Saber a diferença fundamental entre a eliminação de Gauss e o método de Gauss-Jordan e qual é o mais eficiente.
7. Reconhecer como a eliminação de Gauss pode ser formulada como uma decomposição *LU*.
8. Saber como incorporar o pivotamento e a inversão de matrizes em um algoritmo de decomposição *LU*.
9. Saber como interpretar os elementos da matriz inversa no cálculo de respostas a estímulos em engenharia.
10. Perceber como usar a inversa e a norma de uma matriz para avaliar o condicionamento do sistema.
11. Entender como sistemas de banda e simétricos podem ser decompostos e resolvidos eficientemente.
12. Entender por que o método de Gauss-Seidel é particularmente apropriado para sistemas de equações grandes e esparsos.
13. Saber como avaliar a dominância diagonal de um sistema de equações e como isso se relaciona com o fato de esse sistema poder ou não ser resolvido pelo método de Gauss-Seidel.
14. Entender o argumento por trás do relaxamento; saber onde sub-relaxamento e sobre-relaxamento são apropriados.

CAPÍTULO 9

Eliminação de Gauss

Este capítulo trata das equações algébricas lineares simultâneas que podem ser representadas no caso geral por

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdot \qquad\qquad & \cdot \\ \cdot \qquad\qquad & \cdot \\ \cdot \qquad\qquad & \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \tag{9.1}$$

onde os a's são coeficientes constantes e os b's são constantes.

A técnica descrita neste capítulo é chamada de *eliminação de Gauss*, pois envolve combinar equações para eliminar variáveis. Apesar de ser um dos métodos mais antigos para resolver equações simultâneas, mantém-se como um dos algoritmos mais importantes em uso hoje em dia e é a base da resolução de equações lineares em muitos pacotes de software populares.

9.1 RESOLUÇÃO DE UM NÚMERO PEQUENO DE EQUAÇÕES

Antes de prosseguir para os métodos computacionais, serão descritos vários métodos que são apropriados para resolver um conjunto pequeno de equações simultâneas ($n \leq 3$) e que não exigem um computador: o método gráfico, a regra de Cramer e a eliminação de variáveis.

9.1.1 O Método Gráfico

Uma solução gráfica para duas equações pode ser obtida esboçando-se seus gráficos em coordenadas cartesianas, com um eixo correspondente a x_1 e o outro a x_2. Como estamos lidando com sistemas lineares, cada equação é uma reta. Isso pode ser facilmente ilustrado para as equações na forma geral

$$a_{11}x_1 + a_{12}x_2 = b_1$$

$$a_{21}x_1 + a_{22}x_2 = b_2$$

Pode-se isolar x_2 nas duas equações:

$$x_2 = -\left(\frac{a_{11}}{a_{12}}\right)x_1 + \frac{b_1}{a_{12}}$$

$$x_2 = -\left(\frac{a_{21}}{a_{22}}\right)x_1 + \frac{b_2}{a_{22}}$$

Assim, as equações estão agora na forma de equações de retas: isto é, x_2 = (inclinação) x_1 + intersecção com o eixo. Essas retas podem ser traçadas em coordenadas cartesianas com x_2 como a ordenada e x_1 como a abscissa. Os valores de x_1 e x_2 na intersecção das retas representam a solução.

EXEMPLO 9.1 O Método Gráfico para Duas Equações

Enunciado do Problema. Use o método gráfico para resolver o sistema

$$3x_1 + 2x_2 = 18 \tag{E9.1.1}$$

$$-x_1 + 2x_2 = 2 \tag{E9.1.2}$$

Solução. Seja x_1 a abscissa. Isole x_2 na Equação (E9.1.1):

$$x_2 = -\frac{3}{2}x_1 + 9$$

que, quando traçada na Figura 9.1, é uma reta que intercepta o eixo x_2 em 9 e cuja inclinação é $-3/2$.

Também se pode isolar x_2 na Equação (E9.1.2):

$$x_2 = \frac{1}{2}x_1 + 1$$

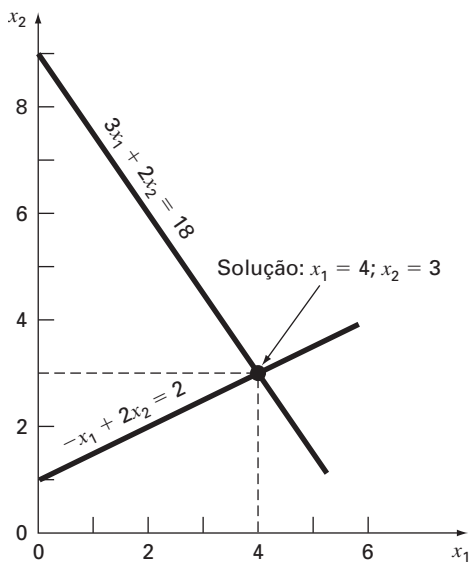
que também está traçada na Figura 9.1. A solução é a intersecção das duas retas em $x_1 = 4$ e $x_2 = 3$. Esses resultados podem ser verificados substituindo esses valores nas equações originais, o que leva a

$$3(4) + 2(3) = 18$$

$$-(4) + 2(3) = 2$$

Assim, os resultados coincidem com o lado direito das equações originais.

FIGURA 9.1
Solução gráfica de um conjunto de duas equações algébricas lineares simultâneas. A intersecção das retas representa a solução.



Para três equações simultâneas, cada equação será representada por um plano em um sistema de coordenadas tridimensional. O ponto no qual os planos se interceptam irá representar a solução. Para mais de três equações, os métodos gráficos não funcionam e, conseqüentemente, têm pouco valor prático para resolver equações simultâneas. Apesar disso, às vezes são úteis para visualizar propriedades das soluções. Por exemplo, a Figura 9.2

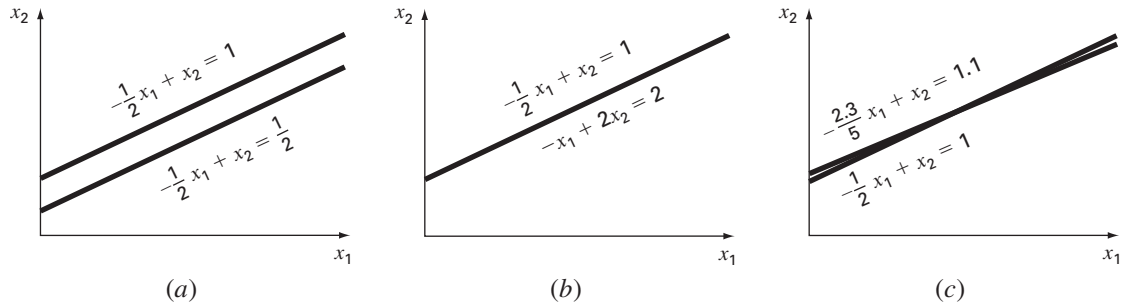


FIGURA 9.2
Descrição gráfica de sistemas singulares e mal condicionados: (*a*) nenhuma solução, (*b*) infinitas soluções e (*c*) sistemas mal condicionados, nos quais as inclinações são tão próximas que o ponto de intersecção é difícil de detectar visualmente.

descreve três casos que podem ser problemáticos quando se resolve um conjunto de equações lineares. A Figura 9.2*a* mostra o caso em que duas equações representam retas paralelas. Para tais situações, não há solução, pois as retas não se cruzam. A Figura 9.2*b* mostra o caso em que as duas retas são coincidentes. Para tais situações, há um número infinito de soluções. Ambos os tipos de sistemas são chamados de *singulares*. Além desses, sistemas que estão muito próximos de serem singulares (Figura 9.2*c*) também podem causar problemas. Esses sistemas são chamados de *mal condicionados*. Graficamente, isso corresponde ao fato de que é difícil identificar o ponto exato no qual as retas se interceptam. Os sistemas mal condicionados também causarão problemas quando forem encontrados durante as soluções numéricas de equações lineares, porque são extremamente sensíveis a erros de arredondamento (lembre-se da Seção 4.2.3).

9.1.2 Determinantes e a Regra de Cramer

A regra de Cramer é outra técnica de resolução que é mais bem adequada a um pequeno número de equações. Antes de descrever esse método, será apresentado brevemente o conceito de determinante, que é usado para implementar a regra de Cramer. Além disso, o determinante é relevante para avaliar o mal condicionamento de uma matriz.

Determinantes. O determinante pode ser ilustrado para um conjunto de três equações:

$$[A]\{X\} = \{B\}$$

onde $[A]$ é a matriz dos coeficientes:

$$[A] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

O determinante D desse sistema é formado a partir dos coeficientes da equação, como em

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \tag{9.2}$$

Apesar de o determinante D e a matriz dos coeficientes $[A]$ serem compostos dos mesmos elementos, são conceitos matemáticos completamente diferentes. Por essa razão, são distinguidos visualmente usando-se colchetes para delimitar a matriz e linhas retas para delimitar o determinante. Em contraste com a matriz, o determinante é um único número. Por exemplo, o valor do determinante de segunda ordem

$$D = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

é calculado por

$$D = a_{11}a_{22} - a_{12}a_{21} \tag{9.3}$$

Para o caso de terceira ordem [Equação (9.2)], o valor numérico do determinante pode ser calculado por

$$D = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \tag{9.4}$$

onde os determinantes dois por dois são chamados de **menores**.

EXEMPLO 9.2 Determinantes

Enunciado do Problema. Calcule os valores dos determinantes dos sistemas dados nas Figuras 9.1 e 9.2.

Solução. Para a Figura 9.1:

$$D = \begin{vmatrix} 3 & 2 \\ -1 & 2 \end{vmatrix} = 3(2) - 2(-1) = 8$$

Para a Figura 9.2*a*:

$$D = \begin{vmatrix} -1/2 & 1 \\ -1/2 & 1 \end{vmatrix} = \frac{-1}{2}(1) - 1\left(\frac{-1}{2}\right) = 0$$

Para a Figura 9.2*b*:

$$D = \begin{vmatrix} -1/2 & 1 \\ -1 & 2 \end{vmatrix} = \frac{-1}{2}(2) - 1(-1) = 0$$

Para a Figura 9.2*c*:

$$D = \begin{vmatrix} -1/2 & 1 \\ -2{,}3/5 & 1 \end{vmatrix} = \frac{-1}{2}(1) - 1\left(\frac{-2{,}3}{5}\right) = -0{,}04$$

No exemplo precedente, os sistemas singulares têm determinantes nulos. Além disso, os resultados sugerem que o sistema que é quase singular (Figura 9.2*c*) tem um determinante próximo de zero. Essas idéias serão examinadas mais a fundo na discussão subseqüente de mal condicionamento (Seção 9.3.3).

Regra de Cramer. Essa regra estabelece que cada incógnita em um sistema de equações algébricas lineares pode ser expressa como uma fração de dois determinantes, com denominador D e com o numerador obtido a partir de D trocando-se a coluna de coeficientes da incógnita em questão pelas constantes $b_1, b_2, \ldots, b_n$. Por exemplo, x_1 é calculada por

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}}{D} \tag{9.5}$$

EXEMPLO 9.3 Regra de Cramer

Enunciado do Problema. Use a regra de Cramer para resolver

$$0{,}3x_1 + 0{,}52x_2 + x_3 = -0{,}01$$
$$0{,}5x_1 + x_2 + 1{,}9x_3 = 0{,}67$$
$$0{,}1x_1 + 0{,}3x_2 + 0{,}5x_3 = -0{,}44$$

Solução. O determinante D pode ser escrito como [Equação (9.2)]

$$D = \begin{vmatrix} 0,3 & 0,52 & 1 \\ 0,5 & 1 & 1,9 \\ 0,1 & 0,3 & 0,5 \end{vmatrix}$$

Os menores são [Equação (9.3)]

$$A_1 = \begin{vmatrix} 1 & 1.9 \\ 0,3 & 0,5 \end{vmatrix} = 1(0,5) - 1,9(0,3) = -0,07$$

$$A_2 = \begin{vmatrix} 0,5 & 1,9 \\ 0,1 & 0,5 \end{vmatrix} = 0,5(0,5) - 1,9(0,1) = 0,06$$

$$A_3 = \begin{vmatrix} 0,5 & 1 \\ 0,1 & 0,3 \end{vmatrix} = 0,5(0,3) - 1(0,1) = 0,05$$

Esses podem ser usados para calcular o determinante, como em [Equação (9.4)]

$$D = 0,3(-0,07) - 0,52(0,06) + 1(0,05) = -0,0022$$

Aplicando a Equação (9.5), a solução é

$$x_1 = \frac{\begin{vmatrix} -0,01 & 0,52 & 1 \\ 0,67 & 1 & 1,9 \\ -0,44 & 0,3 & 0,5 \end{vmatrix}}{-0,0022} = \frac{0,03278}{-0,0022} = -14,9$$

$$x_2 = \frac{\begin{vmatrix} 0,3 & -0,01 & 1 \\ 0,5 & 0,67 & 1,9 \\ 0,1 & -0,44 & 0,5 \end{vmatrix}}{-0,0022} = \frac{0,0649}{-0,0022} = -29,5$$

$$x_3 = \frac{\begin{vmatrix} 0,3 & 0.52 & -0,01 \\ 0,5 & 1 & 0,67 \\ 0,1 & 0,3 & -0,44 \end{vmatrix}}{-0,0022} = \frac{-0,04356}{-0,0022} = 19,8$$

Para mais de três equações, a regra de Cramer torna-se impraticável, pois à medida que o número de equações cresce, os determinantes demoram mais para ser calculados à mão (ou pelo computador). Conseqüentemente, são usadas alternativas mais eficientes. Algumas dessas alternativas são baseadas na última técnica de solução não-computacional, estudada na próxima seção — a eliminação de variáveis.

9.1.3 A Eliminação de Variáveis

A eliminação de variáveis pela combinação de equações é uma abordagem algébrica que pode ser ilustrada para um conjunto de duas equações:

$$a_{11}x_1 + a_{12}x_2 = b_1 \tag{9.6}$$

$$a_{21}x_1 + a_{22}x_2 = b_2 \tag{9.7}$$

A estratégia básica é multiplicar as equações por constantes de modo que uma das variáveis seja eliminada quando as duas equações são combinadas. O resultado é uma única equação que pode ser resolvida determinando a variável restante. Então, esse valor pode ser substituído em qualquer uma das equações originais para calcular a outra variável.

Por exemplo, a Equação (9.6) pode ser multiplicada por a_{21} e a Equação (9.7) por a_{11} para fornecer

$$a_{11}a_{21}x_1 + a_{12}a_{21}x_2 = b_1a_{21} \tag{9.8}$$

$$a_{21}a_{11}x_1 + a_{22}a_{11}x_2 = b_2a_{11} \tag{9.9}$$

Subtrair a Equação (9.8) da Equação (9.9) irá, portanto, eliminar o termo x_1 da equação, dando

$$a_{22}a_{11}x_2 - a_{12}a_{21}x_2 = b_2a_{11} - b_1a_{21}$$

que pode ser resolvida por

$$x_2 = \frac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}} \tag{9.10}$$

A Equação (9.10) pode então ser substituída na Equação (9.6), a qual é resolvida por

$$x_1 = \frac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}} \tag{9.11}$$

Observe que as Equações (9.10) e (9.11) seguem diretamente da regra de Cramer, a qual afirma que

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{b_1a_{22} - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}}$$

$$x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} = \frac{a_{11}b_2 - b_1a_{21}}{a_{11}a_{22} - a_{12}a_{21}}$$

EXEMPLO 9.4 Eliminação de Variáveis

Enunciado do Problema. Use a eliminação de variáveis para resolver (lembre-se do Exemplo 9.1)

$$3x_1 + 2x_2 = 18$$

$$-x_1 + 2x_2 = 2$$

Solução. Usando as Equações (9.11) e (9.10),

$$x_1 = \frac{2(18) - 2(2)}{3(2) - 2(-1)} = 4$$

$$x_2 = \frac{3(2) - (-1)18}{3(2) - 2(-1)} = 3$$

que é consistente com a solução gráfica (Figura 9.1).

A eliminação de variáveis pode ser estendida a sistemas com mais do que duas ou três equações. Porém, os numerosos cálculos necessários para sistemas maiores tornam o método extremamente tedioso para ser implementado à mão. Mas, como se verá na próxima seção, a técnica pode ser formalizada e rapidamente programada para o computador.

9.2 ELIMINAÇÃO DE GAUSS INGÊNUA

Na seção anterior, a eliminação de variáveis foi usada para resolver um par de equações simultâneas. O processo consiste em dois passos:

1. As equações são manipuladas para eliminar uma das variáveis das equações. O resultado desse passo de *eliminação* foi que se obteve uma equação com uma variável.
2. Conseqüentemente, essa equação pode ser resolvida diretamente e o resultado *substituído regressivamente* em uma das equações originais para determinar a variável remanescente.

Essa abordagem básica pode ser estendida a conjuntos maiores de equações desenvolvendo-se um esquema sistemático ou um algoritmo para eliminar variáveis e substituir regressivamente. A *eliminação de Gauss* é o mais básico desses esquemas.

Esta seção inclui técnicas sistemáticas para eliminação progressiva e substituição regressiva que compõem a eliminação de Gauss. Apesar de essas técnicas serem hipoteticamente adequadas para implementação em computadores, algumas modificações serão necessárias para se obter um algoritmo confiável. Em particular, o programa de computador deve evitar a divisão por zero. O método a seguir é chamado de *eliminação de Gauss "ingênua"* porque não evita esse problema. Seções subseqüentes irão tratar das características adicionais necessárias para um programa de computador efetivo.

A abordagem é planejada para resolver um conjunto arbitrário de n equações:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \tag{9.12a}$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \tag{9.12b}$$

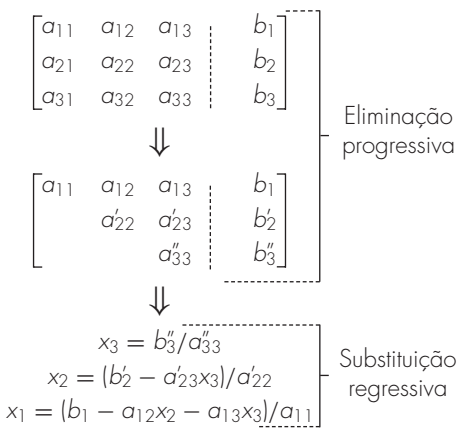
$$\begin{array}{cc} . & . \\ . & . \\ . & . \end{array}$$

$$a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \cdots + a_{nn}x_n = b_n \tag{9.12c}$$

Da mesma forma como no caso da solução de duas equações, a técnica para n equações consiste em duas fases: a eliminação de variáveis e a solução por meio da substituição regressiva.

Eliminação Progressiva de Variáveis. A primeira fase é projetada para reduzir o conjunto de equações a um sistema triangular superior (Figura 9.3). O passo inicial será eliminar a primeira variável, x_1, da segunda até a n-ésima equação. Para fazer isso, multiplique a Equação (9.12a) por a_{21}/a_{11} para obter

FIGURA 9.3
As duas fases da eliminação de Gauss: eliminação progressiva e substituição regressiva. As linhas indicam o número de vezes que os coeficientes e as constantes foram modificados.



$$a_{21}x_1 + \frac{a_{21}}{a_{11}}a_{12}x_2 + \cdots + \frac{a_{21}}{a_{11}}a_{1n}x_n = \frac{a_{21}}{a_{11}}b_1 \tag{9.13}$$

Agora, essa equação pode ser subtraída da Equação (9.12*b*) para se obter

$$\left(a_{22} - \frac{a_{21}}{a_{11}}a_{12}\right)x_2 + \cdots + \left(a_{2n} - \frac{a_{21}}{a_{11}}a_{1n}\right)x_n = b_2 - \frac{a_{21}}{a_{11}}b_1$$

ou

$$a'_{22}x_2 + \cdots + a'_{2n}x_n = b'_2$$

onde as linhas indicam que o elemento foi modificado de seu valor original.

O processo é repetido para as equações seguintes. Por exemplo, a Equação (9.12*a*) pode ser multiplicada por a_{31}/a_{11} e o resultado, subtraído da terceira equação. Repetindo o procedimento para as equações restantes, obtém-se o seguinte sistema modificado:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \tag{9.14a}$$

$$a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n = b'_2 \tag{9.14b}$$

$$a'_{32}x_2 + a'_{33}x_3 + \cdots + a'_{3n}x_n = b'_3 \tag{9.14c}$$

$$\begin{array}{cc} . & . \\ . & . \\ . & . \end{array}$$

$$a'_{n2}x_2 + a'_{n3}x_3 + \cdots + a'_{nn}x_n = b'_n \tag{9.14d}$$

Para os passos precedentes, a Equação (9.12*a*) é chamada de *equação pivô* e a_{11} é chamado de *coeficiente* ou *elemento pivô.* Observe que o processo de multiplicar a primeira linha por a_{21}/a_{11} é equivalente a dividi-la por a_{11} e multiplicá-la por a_{21}. Às vezes, essa operação de divisão é chamada de normalização. Fazemos essa distinção porque um elemento pivô nulo pode interferir na normalização, já que provoca uma divisão por zero. Voltaremos a esse assunto importante depois de completar a descrição da eliminação de Gauss ingênua.

Agora, repita o que foi feito para eliminar a segunda variável das Equações (9.14*c*) até (9.14*d*). Para fazer isso, multiplique a Equação (9.14*b*) por a'_{32}/a'_{22} e subtraia o resultado da Equação (9.14*c*). Faça uma eliminação semelhante nas equações restantes para obter

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n &= b'_2 \\ a''_{33}x_3 + \cdots + a''_{3n}x_n &= b''_3 \\ . \qquad\qquad . & \\ . \qquad\qquad . & \\ . \qquad\qquad . & \\ a''_{n3}x_3 + \cdots + a''_{nn}x_n &= b''_n \end{aligned}$$

onde as linhas duplas indicam que os elementos foram modificados duas vezes.

O processo pode continuar usando as equações pivô remanescentes. A última coisa a fazer é usar a $(n-1)$-ésima equação para eliminar o termo x_{n-1} da n-ésima equação. Nesse ponto, o sistema terá sido transformado em um sistema triangular superior (lembre-se do Quadro PT3.1):

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \tag{9.15a}$$

$$a'_{22}x_2 + a'_{23}x_3 + \cdots + a'_{2n}x_n = b'_2 \tag{9.15b}$$

$$a''_{33}x_3 + \cdots + a''_{3n}x_n = b''_3 \tag{9.15c}$$

$$\begin{array}{ccc} . & . & . \\ . & . & . \\ . & . & . \end{array}$$

$$a_{nn}^{(n-1)}x_n = b_n^{(n-1)} \tag{9.15d}$$

Um pseudocódigo para implementar a eliminação progressiva é apresentado na Figura 9.4*a*. Observe que três laços encaixados fornecem uma representação concisa do processo. O laço exterior move-se para baixo na matriz de uma linha pivô para a próxima. O laço do meio move-se para baixo da linha pivô, para cada uma das linhas subseqüentes onde a eliminação é efetuada. Finalmente, o laço interior progride ao longo das colunas para eliminar ou modificar os elementos de uma linha particular.

Substituição Regressiva. A Equação (9.15*d*) pode agora ser resolvida para determinar x_n:

$$x_n = \frac{b_n^{(n-1)}}{a_{nn}^{(n-1)}} \tag{9.16}$$

Esse resultado pode ser substituído de volta na $(n-1)$-ésima equação para determinar x_{n-1}. O processo, que é repetido para determinar os x's remanescentes, pode ser representado pela seguinte fórmula:

$$x_i = \frac{b_i^{(i-1)} - \displaystyle\sum_{j=i+1}^{n} a_{ij}^{(i-1)}x_j}{a_{ii}^{(i-1)}} \qquad \text{para } i = n-1, n-2, \ldots, 1 \tag{9.17}$$

Um pseudocódigo para implementar as Equações (9.16) e (9.17) é apresentado na Figura 9.4*b*. Observe a semelhança entre esse pseudocódigo e aquele na Figura PT3.4 para a multiplicação de matrizes. Como na Figura PT3.4, uma variável temporária, *soma*, é usada para acumular as somas da Equação (9.17). Isso resulta em uma execução um pouco mais rápida do que se as somas fossem acumuladas em b_i. Mais importante, permite um aperfeiçoamento eficiente se a variável, *soma,* for declarada em dupla precisão.

FIGURA 9.4
Pseudocódigo para efetuar (*a*) eliminação progressiva e (*b*) substituição regressiva.

(*a*)
```
DOFOR k = 1, n − 1
 DOFOR i = k + 1, n
  fator = a_{i,k} / a_{k,k}
  DOFOR j = k + 1 to n
   a_{i,j} = a_{i,j} − fator · a_{k,j}
  END DO
  b_i = b_i − fator · b_k
 END DO
END DO
```

(*b*)
```
x_n = b_n / a_{n,n}
DOFOR i = n − 1, 1, −1
  soma = b_i
  DOFOR j = i + 1, n
   soma = soma − a_{i,j} · x_j
  END DO
  x_i = b_i / a_{i,i}
END DO
```

EXEMPLO 9.5 Eliminação de Gauss Ingênua

Enunciado do Problema. Use eliminação de Gauss para resolver

$$3x_1 - 0,1x_2 - 0,2x_3 = 7,85 \quad \text{(E9.5.1)}$$
$$0,1x_1 + 7x_2 - 0,3x_3 = -19,3 \quad \text{(E9.5.2)}$$
$$0,3x_1 - 0,2x_2 + 10x_3 = 71,4 \quad \text{(E9.5.3)}$$

Use seis algarismos significativos durante os cálculos.

Solução. A primeira parte do processo é a eliminação progressiva. Multiplique a Equação (E9.5.1) por (0,1)/3 e subtraia o resultado da Equação (E9.5.2) para obter

$$7,00333x_2 - 0,293333x_3 = -19,5617$$

Então, multiplique a Equação (E9.5.1) por (0,3)/3 e subtraia da Equação (E9.5.3) para eliminar x_1. Depois dessas operações, o conjunto de equações é

$$3x_1 - 0,1x_2 - 0,2x_3 = 7,85 \quad \text{(E9.5.4)}$$
$$7,00333x_2 - 0,293333x_3 = -19,5617 \quad \text{(E9.5.5)}$$
$$-0,190000x_2 + 10,0200x_3 = 70,6150 \quad \text{(E9.5.6)}$$

Para completar a eliminação progressiva, x_2 deve ser removida da Equação (E9.5.6). Para fazer isso, multiplique a Equação (E9.5.5) por $-0,190000/7,00333$ e subtraia o resultado da Equação (E9.5.6). Isso elimina x_2 da terceira equação e reduz o sistema a um sistema triangular superior, como em

$$3x_1 - 0,1x_2 - 0,2x_3 = 7,85 \quad \text{(E9.5.7)}$$
$$7,00333x_2 - 0,293333x_3 = -19,5617 \quad \text{(E9.5.8)}$$
$$10,0120x_3 = 70,0843 \quad \text{(E9.5.9)}$$

Agora, é possível resolver essas equações por substituição regressiva. Primeiro, a Equação (E9.5.9) pode ser resolvida para determinar

$$x_3 = \frac{70,0843}{10,0120} = 7,0000 \quad \text{(E9.5.10)}$$

Esse resultado pode ser substituído na Equação (E9.5.8):

$$7,00333x_2 - 0,293333(7,0000) = -19,5617$$

que pode ser resolvida por

$$x_2 = \frac{-19,5617 + 0,293333(7,0000)}{7,00333} = -2,50000 \quad \text{(E9.5.11)}$$

Finalmente, as Equações (E9.5.10) e (E9.5.11) podem ser substituídas na Equação (E9.5.4):

$$3x_1 - 0,1(-2,50000) - 0,2(7,0000) = 7,85$$

a qual pode ser resolvida por

$$x_1 = \frac{7,85 + 0,1(-2,50000) + 0,2(7,0000)}{3} = 3,00000$$

Tais resultados são idênticos à solução exata $x_1 = 3$, $x_2 = -2,5$ e $x_3 = 7$. Isso pode ser verificado substituindo-se os resultados no conjunto original de equações

$$3(3) - 0,1(-2,5) - 0,2(7) = 7,85$$
$$0,1(3) + 7(-2,5) - 0,3(7) = -19,3$$
$$0,3(3) - 0,2(-2,5) + 10(7) = 71,4$$

9.2.1 Contando Operações

O tempo de execução da eliminação de Gauss depende da quantidade de *operações com ponto flutuante* (ou *flops)* envolvidas no algoritmo. Em computadores modernos usando co-processadores matemáticos, o tempo consumido para efetuar uma adição/subtração ou uma multiplicação/divisão é mais ou menos o mesmo. Portanto, a contagem dessas operações fornece informação sobre quais partes do algoritmo consomem mais tempo e sobre como o tempo de computação cresce à medida que o sistema fica maior.

Antes de analisar a eliminação de Gauss ingênua, primeiro é preciso definir algumas quantidades que facilitam a contagem de operações:

$$\sum_{i=1}^{m} cf(i) = c\sum_{i=1}^{m} f(i) \qquad \sum_{i=1}^{m} f(i)+g(i) = \sum_{i=1}^{m} f(i) + \sum_{i=1}^{m} g(i) \tag{9.18a,b}$$

$$\sum_{i=1}^{m} 1 = 1+1+1+\cdots+1 = m \qquad \sum_{i=k}^{m} 1 = m-k+1 \tag{9.18c,d}$$

$$\sum_{i=1}^{m} i = 1+2+3+\cdots+m = \frac{m(m+1)}{2} = \frac{m^2}{2} + O(m) \tag{9.18e}$$

$$\sum_{i=1}^{m} i^2 = 1^2+2^2+3^2+\cdots+m^2 = \frac{m(m+1)(2m+1)}{6} = \frac{m^3}{3} + O(m^2) \tag{9.18f}$$

onde $O(m^n)$ significa "termos de ordem m^n e menores".

Em seguida, o algoritmo de eliminação de Gauss ingênua (Figura 9.4*a*) é examinado em detalhes. Primeiro, contam-se os flops no estágio de eliminação. Na primeira passagem pelo laço exterior, $k = 1$. Então, os limites no laço do meio são de $i = 2$ até n. De acordo com a Equação (9.18*d*), isso significa que o número de iterações do laço do meio será

$$\sum_{i=2}^{n} 1 = n-2+1 = n-1 \tag{9.19}$$

Para cada uma dessas iterações há uma divisão para definir o fator. Então, o laço interior executa uma única multiplicação e subtração para cada iteração de $j = 2$ até n. Finalmente, há uma multiplicação e uma subtração adicionais para determinar o valor do lado direito. Assim, para cada iteração do laço do meio, o número de multiplicações é

$$1+[n-2+1]+1 = 1+n \tag{9.20}$$

O total de multiplicações para a primeira passagem pelo laço exterior é, portanto, obtida pela multiplicação da Equação (9.19) pela Equação (9.20), o que dá $[n-1](1+n)$. De modo semelhante, o número de subtrações é calculado como $[n-1](n)$.

Raciocínios semelhantes podem ser usados para estimar os flops para as iterações subseqüentes do laço exterior. Isso pode ser resumido por

Laço Exterior ***k***	**Laço do Meio** ***i***	**Flops Adição/Subtração**	**Flops Multiplicação/Divisão**
1	2, *n*	$(n-1)(n)$	$(n-1)(n+1)$
2	3, *n*	$(n-2)(n-1)$	$(n-2)(n)$
.	.		
.	.		
.	.		
k	$k+1$, *n*	$(n-k)(n+1-k)$	$(n-k)(n+2-k)$
.	.		
.	.		
.	.		
$n-1$	*n*, *n*	(1)(2)	(1) (3)

Então, o total de flops de adição/subtração para a eliminação pode ser calculado por

$$\sum_{k=1}^{n-1}(n-k)(n+1-k)=\sum_{k=1}^{n-1}[n(n+1)-k(2n+1)+k^2]$$

ou

$$n(n+1)\sum_{k=1}^{n-1}1-(2n+1)\sum_{k=1}^{n-1}k+\sum_{k=1}^{n-1}k^2$$

Aplicar algumas das relações da Equação (9.18) resulta em

$$[n^3+O(n)]-[n^3+O(n^2)]+\left[\frac{1}{3}n^3+O(n^2)\right]=\frac{n^3}{3}+O(n) \tag{9.21}$$

Uma análise parecida para os flops de multiplicação/divisão resulta em

$$[n^3+O(n^2)]-[n^3+O(n)]+\left[\frac{1}{3}n^3+O(n^2)\right]=\frac{n^3}{3}+O(n^2) \tag{9.22}$$

Somando os resultados, tem-se

$$\frac{2n^3}{3}+O(n^2)$$

Assim, o número total de flops é igual a $2n^3/3$ mais uma componente adicional proporcional a termos de ordem n^2 e menores. O resultado é escrito dessa forma porque, à medida que n fica maior, o $O(n^2)$ e termos menores ficam desprezíveis. Pode-se então concluir, justificadamente, que para n grande, o esforço envolvido na eliminação progressiva converge para $2n^3/3$.

Como apenas um único laço foi utilizado, a substituição regressiva é muito mais simples de ser avaliada. O número de flops de adição/subtração é igual a $n(n-1)/2$. Por causa da divisão extra anterior ao laço, o número de flops de multiplicação/divisão é $n(n+1)/2$. Isso pode ser somado para se obter um total de

$$n^2+O(n)$$

Assim, o esforço total na eliminação de Gauss ingênua pode ser representado por

$$\underbrace{\frac{2n^3}{3}+O(n^2)}_{\substack{\text{Eliminação}\\ \text{progressiva}}}+\underbrace{n^2+O(n)}_{\substack{\text{Substituição}\\ \text{regressiva}}}\xrightarrow{\text{quando } n \text{ cresce}}\frac{2n^3}{3}+O(n^2) \tag{9.23}$$

Duas conclusões gerais úteis podem ser tiradas dessa análise:

1. À medida que o sistema fica maior, o tempo de computação cresce bastante. Como na Tabela 9.1, a quantidade de flops cresce quase três ordens de grandeza para cada aumento na ordem de grandeza da dimensão.
2. A maior parte do esforço vem do passo de eliminação. Assim, os esforços para tornar o algoritmo mais eficiente devem provavelmente concentrar-se nesse passo.

TABELA 9.1 Número de flops para a eliminação de Gauss.

n	Eliminação	Substituição Regressiva	Total de Flops	$2n^3/3$	Percentagem Devida à Eliminação
10	705	100	805	667	87,58%
100	671550	10000	681550	666667	98,53%
1000	$6{,}67 \times 10^8$	1×10^6	$6{,}68 \times 10^8$	$6{,}67 \times 10^8$	99,85%

9.3 ARMADILHAS DOS MÉTODOS DE ELIMINAÇÃO

Embora existam muitos sistemas de equações que podem ser resolvidas pela eliminação de Gauss ingênua, há algumas armadilhas que precisam ser exploradas antes de escrever um programa de computador geral para implementar o método. Apesar de o material a seguir se relacionar diretamente ao método de eliminação de Gauss ingênua, a informação é relevante também para outras técnicas de eliminação.

9.3.1 Divisão por Zero

A principal razão para que a técnica anterior seja chamada de "ingênua" é que tanto durante a fase de eliminação quanto durante a de substituição é possível ocorrer uma divisão por zero. Por exemplo, se for usada a eliminação de Gauss ingênua para resolver

$$2x_2 + 3x_3 = 8$$

$$4x_1 + 6x_2 + 7x_3 = -3$$

$$2x_1 + x_2 + 6x_3 = 5$$

a normalização da primeira linha iria envolver a divisão por $a_{11} = 0$. Também podem surgir problemas quando um coeficiente está muito próximo de zero. A técnica de *pivotamento* foi desenvolvida para evitar parcialmente esses problemas. Ela será descrita na Seção 9.4.2.

9.3.2 Erros de Arredondamento

Embora a solução no Exemplo 9.5 estivesse próxima da resposta verdadeira, havia uma pequena discrepância no resultado de x_3 [Equação (E9.5.10)]. Essa discrepância, que corresponde a um erro relativo de $-0{,}00043\%$, decorreu do uso de seis algarismos significativos durante os cálculos. Se tivessem sido usados mais algarismos significativos, o erro no resultado seria ainda mais reduzido. Se tivessem sido usadas frações em vez de decimais (e, conseqüentemente, fossem evitados erros de arredondamento), as respostas seriam exatas. De qualquer forma, como os computadores usam apenas um número limitado de algarismos significativos (lembre-se da Seção 3.4.1), os erros de arredondamento podem ocorrer e devem ser considerados quando se avaliam os resultados.

O problema de erros de arredondamento pode tornar-se particularmente importante quando se resolve um número grande de equações, por causa do fato de que cada resultado depende dos resultados anteriores. Conseqüentemente, um erro nas etapas iniciais tende a se propagar — isto é, irá causar erros nas etapas subseqüentes.

Especificar o tamanho do sistema no qual os erros de arredondamento se tornam significativos é complicado pelo fato de que o tipo do computador e as propriedades das equações são fatores determinantes. Uma regra empírica é que os erros de arredondamento podem ser importantes quando se lida com 100 ou mais equações. De qualquer modo, você deve sempre substituir suas respostas no sistema original para verificar se erros substanciais ocorreram. Entretanto, como será discutido a seguir, as ordens de grandezas dos próprios coeficientes podem influenciar no fato de tal verificação de erros garantir ou não um resultado confiável.

9.3.3 Sistemas Mal Condicionados

A adequação da solução depende do condicionamento do sistema. Na Seção 9.1.1, foi desenvolvida uma descrição gráfica do condicionamento do sistema. Como discutido na Seção 4.2.3, *sistemas bem condicionados* são aqueles em que uma pequena mudança em um ou mais coeficientes resulta em uma pequena mudança similar nas soluções. *Sistemas mal condicionados* são aqueles para os quais pequenas mudanças nos coeficientes resultam em grandes mudanças nas soluções. Uma interpretação alternativa do mal condicionamento é que uma gama ampla de respostas pode satisfazer aproximadamente as equações. Como os erros de arredondamento podem induzir pequenas mudanças nos coeficientes, essas mudanças artificiais são capazes de conduzir a erros grandes nas soluções nos sistemas mal condicionados, como ilustrado no seguinte exemplo.

EXEMPLO 9.6 Sistemas Mal Condicionados

Enunciado do Problema. Resolva o seguinte sistema:

$$x_1 + 2x_2 = 10 \tag{E9.6.1}$$

$$1{,}1x_1 + 2x_2 = 10{,}4 \tag{E9.6.2}$$

A seguir, resolva novamente com o coeficiente de x_1 na segunda equação levemente modificado para 1,05.

Solução. Usando as Equações (9.10) e (9.11), a solução é

$$x_1 = \frac{2(10) - 2(10{,}4)}{1(2) - 2(1{,}1)} = 4$$

$$x_2 = \frac{1(10{,}4) - 1{,}1(10)}{1(2) - 2(1{,}1)} = 3$$

Entretanto, com a pequena variação do coeficiente a_{21} de 1,1 para 1,05, o resultado muda de forma dramática para

$$x_1 = \frac{2(10) - 2(10{,}4)}{1(2) - 2(1{,}05)} = 8$$

$$x_2 = \frac{1(10{,}4) - 1{,}1(10)}{1(2) - 2(1{,}05)} = 1$$

Observe que a razão principal para as discrepâncias entre os dois resultados é que o denominador representa a diferença de dois números quase iguais. Como ilustrado previamente na Seção 3.4.2, tais diferenças são altamente sensíveis a pequenas variações nos números manipulados.

Nesse ponto, você poderia sugerir que a substituição dos resultados nas equações originais deveria alertá-lo para o problema. Infelizmente, em geral esse não é o caso para sistemas mal condicionados. A substituição dos valores errados $x_1 = 8$ e $x_2 = 1$ nas Equações (E9.6.1) e (E9.6.2) leva a

$$8 + 2(1) = 10 = 10$$

$$1{,}1(8) + 2(1) = 10{,}8 \cong 10{,}4$$

Assim, embora $x_1 = 8$ e $x_2 = 1$ não sejam a solução verdadeira, a verificação de erro é suficientemente próxima para possivelmente levá-lo a concluir, de maneira errônea, que essa solução é adequada.

Como foi feito anteriormente na seção sobre métodos gráficos, uma representação visual do mal condicionamento pode ser desenvolvida traçando-se as Equações (E9.6.1) e (E9.6.2) (lembre-se da Figura 9.2). Como as inclinações das retas são quase iguais, é difícil determinar visualmente o ponto exato onde elas se interceptam. Essa dificuldade visual é refletida quantitativamente nos resultados nebulosos do Exemplo 9.6. Pode-se caracterizar matematicamente a situação escrevendo as duas equações na forma geral:

$$a_{11}x_1 + a_{12}x_2 = b_1 \tag{9.24}$$

$$a_{21}x_1 + a_{22}x_2 = b_2 \tag{9.25}$$

Dividindo a Equação (9.24) por a_{12} e a Equação (9.25) por a_{22} e reorganizando, teremos versões alternativas que estão no formato de retas [$x_2 =$ (inclinação) $x_1 +$ intersecção com o eixo]:

$$x_2 = -\frac{a_{11}}{a_{12}}x_1 + \frac{b_1}{a_{12}}$$

$$x_2 = -\frac{a_{21}}{a_{22}}x_1 + \frac{b_2}{a_{22}}$$

Conseqüentemente, se as inclinações forem quase iguais,

$$\frac{a_{11}}{a_{12}} \cong \frac{a_{21}}{a_{22}}$$

ou, multiplicando em cruz,

$$a_{11}a_{22} \cong a_{12}a_{21}$$

o que pode ser expresso por

$$a_{11}a_{22} - a_{12}a_{21} \cong 0 \tag{9.26}$$

Agora, lembrando que $a_{11}a_{22} - a_{12}a_{2l}$ é o determinante do sistema de duas equações [Equação (9.3)], chega-se à conclusão geral de que um sistema mal condicionado é aquele com determinante próximo de zero. De fato, se o determinante for exatamente zero, as duas inclinações serão idênticas, o que implica que o sistema não tem solução ou que tem infinitas soluções, como no caso dos sistemas singulares esboçados nas Figuras 9.2*a* e 9.2*b*.

É difícil especificar quão perto de zero o determinante deve estar para indicar mal condicionamento. Isso é complicado pelo fato de que o determinante pode ser alterado pela multiplicação de uma ou mais das equações por um fator escalar, sem mudar a solução. Conseqüentemente, o determinante é um valor relativo que é influenciado pela ordem de grandeza dos coeficientes.

EXEMPLO 9.7 Efeito da Escala no Determinante

Enunciado do Problema. Calcule o determinante dos seguintes sistemas:

(a) Do Exemplo 9.1:

$$3x_1 + 2x_2 = 18 \tag{E9.7.1}$$

$$-x_1 + 2x_2 = 2 \tag{E9.7.2}$$

(b) Do Exemplo 9.6:

$$x_1 + 2x_2 = 10 \tag{E9.7.3}$$

$$1{,}1x_1 + 2x_2 = 10{,}4 \tag{E9.7.4}$$

(c) Repita (*b*), com as equações multiplicadas por 10.

Solução.

(a) O determinante das Equações (E9.7.1) e (E9.7.2), que são bem condicionadas, é

$$D = 3(2) - 2(-1) = 8$$

(b) O determinante das Equações (E9.7.3) e (E9.7.4), que são mal condicionadas, é

$$D = 1(2) - 2(1{,}1) = -0{,}2$$

(c) Os resultados de (*a*) e (*b*) parecem levar à conclusão de que sistemas mal condicionados têm determinantes próximos de zero. Porém, suponha que o sistema mal condicionado em (*b*) seja multiplicado por 10 dando

$$10x_1 + 20x_2 = 100$$

$$11x_1 + 20x_2 = 104$$

A multiplicação de uma equação por uma constante não tem efeito na sua solução. Além disso, ele é ainda mal condicionado. Isso pode ser verificado pelo fato de que a multiplicação por uma constante não tem efeito na solução gráfica. Porém, o determinante é afetado de maneira dramática:

$$D = 10(20) - 20(11) = -20$$

Não apenas aumentou duas ordens de grandeza, como é agora mais que o dobro do determinante do sistema bem condicionado em (*a*).

Como ilustrado pelo exemplo anterior, o valor dos coeficientes introduz um efeito de escala que complica a relação entre condicionamento do sistema e o tamanho do determinante. Uma forma de evitar parcialmente essa dificuldade é mudar a escala das equações de modo que o elemento máximo em qualquer linha seja igual a 1.

EXEMPLO 9.8 Mudando a Escala

Enunciado do Problema. Faça uma mudança de escala nos sistemas de equações do Exemplo 9.7 para um valor máximo 1 e calcule novamente seus determinantes.

Solução.

(a) Para o sistema bem condicionado, a mudança de escala resulta em

$$\begin{aligned} x_1 + 0{,}667x_2 &= 6 \\ -0{,}5x_1 + \quad x_2 &= 1 \end{aligned}$$

e o determinante é

$$D = 1(1) - 0{,}667(-0{,}5) = 1{,}333$$

(b) Para o sistema mal condicionado, mudando a escala tem-se

$$\begin{aligned} 0{,}5x_1 + x_2 &= 5 \\ 0{,}55x_1 + x_2 &= 5{,}2 \end{aligned}$$

e o determinante é

$$D = 0{,}5(1) - 1(0{,}55) = -0{,}05$$

(c) No último caso, mudar a escala muda o sistema para a mesma forma obtida em (*b*) e o determinante é novamente $-0{,}05$. Assim, o efeito da escala foi removido.

Quadro 9.1 Cálculo de Determinantes Usando Eliminação de Gauss

Na Seção 9.1.2, afirmamos que o cálculo de determinantes por expansão em menores era pouco prático para conjuntos grandes de equações. Assim, concluímos que a regra de Cramer seria aplicável apenas para sistemas pequenos. No entanto, como foi mencionado na Seção 9.3.3, o determinante tem utilidade na avaliação do condicionamento do sistema. Seria, portanto, útil ter um método prático para calculá-lo.

Felizmente, a eliminação de Gauss fornece uma forma simples de fazer isso. O método se baseia no fato de que o determinante de uma matriz triangular pode ser calculado simplesmente como o produto dos elementos da diagonal:

$$D = a_{11}a_{22}a_{33} \cdots a_{nn} \tag{Q9.1.1}$$

A validade dessa afirmação pode ser ilustrada para um sistema 3 por 3:

$$D = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{vmatrix}$$

onde o determinante pode ser calculado por [lembre-se da Equação 9.4]

$$D = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} 0 & a_{23} \\ 0 & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} 0 & a_{22} \\ 0 & 0 \end{vmatrix}$$

ou, calculando os menores (isto é, os determinantes 2 por 2),

$$D = a_{11}a_{22}a_{33} - a_{12}(0) + a_{13}(0) = a_{11}a_{22}a_{33}$$

Lembre-se de que o passo de eliminação progressiva na eliminação de Gauss resulta em um sistema triangular superior. Como o valor do determinante não muda no processo de eliminação progressiva, o determinante pode ser calculado simplesmente no fim desse processo por

$$D = a_{11}a'_{22}a''_{33} \cdots a^{(n-1)}_{nn} \tag{Q9.1.2}$$

onde os subscritos significam o número de vezes que os elementos foram modificados pelo processo de eliminação. Assim, podemos aproveitar o esforço que já foi despendido para reduzir o sistema a uma forma triangular e, de brinde, obter uma estimativa simples do determinante.

Há uma pequena modificação nessa abordagem quando o programa utiliza pivotamento parcial (Seção 9.4.2). Nesse caso, o determinante troca de sinal toda vez que uma linha é pivotada. Uma forma de representar isso é modificar a Equação (Q9.1.2):

$$D = a_{11}a'_{22}a''_{33} \cdots a^{(n-1)}_{nn}(-1)^p \tag{Q9.1.3}$$

onde p indica o número de vezes que as linhas foram pivotadas. Essa modificação pode ser incorporada de modo simples em um programa; meramente mantenha controle do número de pivotamentos que ocorreram durante os cálculos e então use a Equação (Q9.1.3) para calcular o determinante.

Em uma seção anterior (Seção 9.1.2), foi sugerido que o determinante é difícil de calcular para mais do que três equações simultâneas. Assim sendo, pode parecer que não fornece um meio prático de avaliar o condicionamento do sistema. Porém, como se verá no Quadro 9.1, existe um algoritmo simples que vem da eliminação de Gauss que pode ser usado para calcular os determinantes.

Além da abordagem usada no exemplo anterior, há uma variedade de outras formas de avaliar o condicionamento do sistema. Por exemplo, existem métodos alternativos para normalizar os elementos (ver Stark, 1970). Além disso, como descrito no próximo capítulo (Seção 10.3), a inversa e as normas da matriz podem ser usadas para avaliar o condicionamento do sistema. Finalmente, um teste simples (mas que gasta muito tempo) é modificar levemente os coeficientes e repetir a solução. Se tais modificações levarem a resultados drasticamente diferentes, o sistema provavelmente é mal condicionado.

Como se pode ver da discussão precedente, sistemas mal condicionados são problemáticos. Felizmente, em sua maior parte as equações algébricas lineares originárias de problemas de engenharia são naturalmente bem condicionadas. Além disso, algumas técnicas esboçadas na Seção 9.4 ajudam a amenizar o problema.

9.3.4 Sistemas Singulares

Na seção anterior, aprendemos que uma forma pela qual um sistema de equações pode ser mal condicionado é quando duas ou mais equações são quase iguais. Obviamente, é ainda pior quando duas equações são idênticas. Em tais casos, perderíamos um grau de liberdade, e estaríamos tratando do caso impossível de $n - 1$ equações com n incógnitas. Tais casos podem não ser óbvios para você, particularmente ao lidar com um conjunto grande de equações. Conseqüentemente, seria ótimo ter alguma forma de detectar a singularidade automaticamente.

A resposta para esse problema é oferecida naturalmente pelo fato de o determinante de um sistema singular ser zero. Essa idéia pode, por sua vez, ser relacionada à eliminação de Gauss, reconhecendo-se que, após a etapa de eliminação, o determinante pode ser calculado como o produto dos elementos da diagonal (ver Quadro 9.1). Assim, um algoritmo computacional pode testar para detectar se foi criado um elemento nulo na diagonal durante o processo de eliminação. Se algum for descoberto, os cálculos podem ser interrompidos imediatamente e uma mensagem é mostrada para alertar o usuário. Serão mostrados detalhes de como isso é feito na apresentação de um algoritmo completo para a eliminação de Gauss mais adiante neste capítulo.

9.4 TÉCNICAS PARA MELHORAR SOLUÇÕES

As seguintes técnicas podem ser incorporadas ao algoritmo da eliminação de Gauss ingênua para evitar algumas das armadilhas discutidas na seção anterior.

9.4.1 Uso de Mais Algarismos Significativos

O remédio mais simples para o mal condicionamento é usar mais algarismos significativos nos cálculos. Se sua aplicação puder ser estendida para manipular palavras de tamanho maior, tal recurso vai reduzir bastante o problema. Entretanto, deve-se pagar um preço na forma de uso elevado de memória e recursos computacionais associado ao uso de precisão estendida (lembre-se da Seção 3.4.1).

9.4.2 Pivotamento

Como mencionado no começo da Seção 9.3, problemas óbvios ocorrem quando um elemento pivô é zero, pois o passo de normalização leva a uma divisão por zero. Problemas também podem surgir quando o elemento pivô é muito próximo de zero, em vez de exatamente igual a zero, porque, se a ordem de grandeza do elemento pivô é pequena comparada com a dos outros elementos, então podem ocorrer erros de arredondamento.

Assim, antes que cada linha seja normalizada, é vantajoso determinar o maior coeficiente disponível na coluna abaixo do elemento pivô. As linhas podem ser trocadas de modo que o maior coeficiente seja o elemento pivô. Isso é chamado de *pivotamento parcial.* Se procurarmos o maior elemento também nas colunas, além de nas linhas, e então trocarmos, o processo é chamado de *pivotamento completo*. O pivotamento completo é usado raramente, pois trocar colunas muda a ordem dos x's e, conseqüentemente, acrescenta uma complexidade significativa e normalmente injustificada ao programa de computador. O próximo exemplo ilustra as vantagens do pivotamento parcial. Além de evitar a divisão por zero, o pivotamento também minimiza os erros de arredondamento. Assim, serve como um remédio parcial para o mal condicionamento.

EXEMPLO 9.9 Pivotamento Parcial

Enunciado do Problema. Use eliminação de Gauss para resolver

$$0{,}0003x_1 + 3{,}0000x_2 = 2{,}0001$$
$$1{,}0000x_1 + 1{,}0000x_2 = 1{,}0000$$

Observe que dessa forma o primeiro elemento pivô, $a_{11} = 0{,}0003$, é muito próximo de zero. Então, repita os cálculos, mas use pivotamento parcial, trocando a ordem das equações. As soluções exatas são $x_1 = 1/3$ e $x_2 = 2/3$.

Solução. Multiplicando a primeira equação por 1/(0,0003), obtém-se

$$x_1 + 10{,}000x_2 = 6667$$

o que pode ser usado para eliminar x_1 da segunda equação:

$$-9999x_2 = -6666$$

a qual pode ser resolvida por

$$x_2 = \frac{2}{3}$$

Esse resultado pode ser substituído na primeira equação para calcular x_1:

$$x_1 = \frac{2{,}0001 - 3(2/3)}{0{,}0003} \tag{E9.9.1}$$

Entretanto, por causa do cancelamento na subtração, o resultado é muito sensível ao número de algarismos significativos nos cálculos:

Algarismos Significativos	x_2	x_1	Valor Absoluto do Erro Relativo Percentual para x_1
3	0,667	−3,33	1099
4	0,6667	0,0000	100
5	0,66667	0,30000	10
6	0,666667	0,330000	1
7	0,6666667	0,3330000	0,1

Observe como a solução para x_1 é altamente dependente do número de algarismos significativos. Isso ocorre porque, na Equação (E9.9.1), estamos subtraindo dois números quase iguais. Por outro lado, se as equações forem resolvidas na ordem inversa, a linha com elemento pivô maior é normalizada. As equações são

$$1{,}0000x_1 + 1{,}0000x_2 = 1{,}0000$$
$$0{,}0003x_1 + 3{,}0000x_2 = 2{,}0001$$

```
p = k
maior = |a_k,k|
DOFOR ii = k+1, n
  dummy = |a_ii,k|
  IF (dummy > big)
    maior = dummy
    p = ii
  END IF
END DO
IF (p ≠ k)
  DOFOR jj = k, n
    dummy = a_p,jj
    a_p,jj = a_k,jj
    a_k,jj = dummy
  END DO
  dummy = b_p
  b_p = b_k
  b_k = dummy
END IF
```

FIGURA 9.5
Pseudocódigo para implementar o pivotamento parcial.

A eliminação e a substituição resultam em $x_2 = 2/3$. Para números diferentes de algarismos significativos, x_1 pode ser calculado da primeira equação, como em

$$x_1 = \frac{1 - (2/3)}{1} \tag{E9.9.2}$$

Esse caso é muito menos sensível ao número de algarismos significativos nos cálculos:

Algarismos Significativos	x_2	x_1	Valor Absoluto do Erro Relativo Percentual para x_1
3	0,667	0,333	0,1
4	0,6667	0,3333	0,01
5	0,66667	0,33333	0,001
6	0,666667	0,333333	0,0001
7	0,6666667	0,3333333	0,00001

Assim, a estratégia com pivotamento é muito mais satisfatória.

Os programas de computador com propósitos gerais devem incluir uma estratégia de pivotamento. A Figura 9.5 fornece um algoritmo simples para implementar tal estratégia. Observe que o algoritmo consiste em dois laços principais. Depois de armazenar o elemento pivô atual e o número da linha nas variáveis, *maior* e *p*, o primeiro laço compara o elemento pivô com os elementos abaixo dele para verificar se algum deles é maior que o elemento pivô. Se isso ocorre, o novo maior elemento e sua linha são armazenados em *maior* e *p*. Então o segundo laço troca a linha do pivô original com aquela com o maior elemento para que essa última se torne a nova linha pivô. Esse pseudocódigo pode ser integrado a um programa baseado nos outros elementos da eliminação de Gauss esboçado na Figura 9.4. A melhor forma de fazer isso é empregar uma abordagem modular e escrever a Figura 9.5 como uma sub-rotina (ou um procedimento) que será chamado diretamente após o começo do primeiro laço na Figura 9.4*a*.

Observe que o segundo IF/THEN construído na Figura 9.5 troca fisicamente as linhas. Para matrizes maiores, isso pode tornar-se bastante demorado. Conseqüentemente, a maioria dos códigos não troca de fato as linhas, mas mantém um controle das linhas pivô armazenando os índices apropriados em um vetor. Então, esse vetor fornece uma base para especificar a ordenação de linhas apropriada durante a eliminação progressiva e a substituição regressiva. Assim, dizemos que as operações são implementadas *no lugar.*

9.4.3 Mudando a Escala

Na Seção 9.3.3, foi proposto que a mudança de escala é valiosa na padronização do tamanho do determinante. Além dessa aplicação, ela tem utilidade na minimização de erros de arredondamento para aqueles casos nos quais algumas equações do sistema têm coeficientes muito maiores que as outras. Tais situações são encontradas freqüentemente na prática da engenharia quando unidades muito diferentes são usadas na dedução de equações simultâneas. Por exemplo, em problemas de circuitos elétricos, as voltagens indeterminadas podem ser expressas em unidades variando de microvolts até quilovolts. Exemplos similares podem surgir em todos os campos da engenharia. Desde que cada equação seja consistente, o sistema será tecnicamente correto e solúvel. No entanto, o uso de unidades que difiram muito pode levar a coeficientes de ordens de grandeza muito diferentes. Isso, por sua vez, pode ter um impacto nos erros de arredondamento, já que afeta o pivotamento, como ilustrado no próximo exemplo.

```
SUB Gauss (a, b, n, x, tol, er)
  DIMENSION s(n)
  er = 0
  DOFOR i = 1, n
    s_i = ABS(a_i,1)
    DOFOR j = 2, n
      IF ABS(a_i,j)>s_i THEN s_i = ABS(a_i,j)
    END DO
  END DO
  CALL Eliminate(a, s, n, b, tol, er)
  IF er ≠ −1 THEN
     CALL Substitute(a, n, b, x)
  END IF
END Gauss


SUB Eliminate (a, s, n, b, tol, er)
  DOFOR k = 1, n − 1
    CALL Pivot (a, b, s, n, k)
    IF ABS (a_k,k/s_k) < tol THEN
      er = −1
      EXIT DO
    END IF
    DOFOR i = k + 1, n
      fator = a_i,k/a_k,k
      DOFOR j = k + 1, n
        a_i,j = a_i,j − fator*a_k,j
      END DO
      b_i = b_i − fator * b_k
    END DO
  END DO
  IF ABS(a_k,k/s_k) < tol THEN er = −1
END Eliminate
```

```
SUB Pivot (a, b, s, n, k)
  p = k
  maior = ABS(a_k,k/s_k)
  DOFOR ii = k + 1, n
    dummy = ABS(a_ii,k/s_ii)
    IF dummy > big THEN
       maior = dummy
       p = ii
    END IF
  END DO
  IF p ≠ k THEN
     DOFOR jj = k, n
       dummy = a_p,jj
       a_p,jj = a_k,jj
       a_k,jj = dummy
     END DO
     dummy = b_p
     b_p = b_k
     b_k = dummy
     dummy = s_p
     s_p = s_k
     s_k = dummy
  END IF
END pivot


SUB Substitute (a, n, b, x)
  x_n = b_n/a_n,n
  DOFOR i = n − 1, 1, −1
    soma = 0
    DOFOR j = i + 1, n
      soma = soma + a_i,j * x_j
    END DO
    x_i = (b_i − soma) / a_i,i
  END DO
END Substitute
```

FIGURA 9.6
Pseudocódigo para implementação da eliminação de Gauss com pivotamento parcial.

EXEMPLO 9.10

Efeito da Mudança de Escala no Pivotamento e nos Erros de Arredondamento

Enunciado do Problema.

(a) Resolva o seguinte conjunto de equações usando a eliminação de Gauss e a estratégia de pivotamento:

$$2x_1 + 100.000x_2 = 100.000$$
$$x_1 + \quad x_2 = 2$$

(b) Repita a solução depois de mudar a escala das equações de modo que o maior coeficiente em cada linha seja 1.

(c) Finalmente, use os coeficientes normalizados para determinar se o pivotamento é necessário. Entretanto, resolva de fato as equações com os valores originais dos coeficientes. Em todos os casos, trabalhe apenas com três algarismos significativos. Observe que as respostas corretas são $x_1 = 1{,}00002$ e $x_2 = 0{,}99998$ com três algarismos significativos, $x_1 = x_2 = 1{,}00$.

Solução.

(a) Sem mudar a escala, aplica-se a eliminação progressiva e obtém-se

$$\begin{aligned} 2x_1 + 100.000x_2 &= 100.000 \\ -50.000x_2 &= -50.000 \end{aligned}$$

o que pode ser resolvido por substituição regressiva, fornecendo

$$\begin{aligned} x_2 &= 1{,}00 \\ x_1 &= 0{,}00 \end{aligned}$$

Apesar de x_2 estar correto, x_1 está 100% errado, por causa dos erros de arredondamento.

(b) Mudando a escala das equações originais, obtém-se

$$\begin{aligned} 0{,}00002x_1 + x_2 &= 1 \\ x_1 + x_2 &= 2 \end{aligned}$$

Portanto, as linhas devem ser pivotadas para pôr o maior valor na diagonal.

$$\begin{aligned} x_1 + x_2 &= 2 \\ 0{,}00002x_1 + x_2 &= 1 \end{aligned}$$

A eliminação dá

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_2 &= 1{,}00 \end{aligned}$$

o que pode ser resolvido por

$$x_1 = x_2 = 1$$

Assim, a mudança de escala conduziu à resposta correta.

(c) Os coeficientes depois da mudança de escala indicam que o pivotamento é necessário. Então, pivotamos, mas mantemos os coeficientes originais obtendo

$$\begin{aligned} x_1 + \quad x_2 &= 2 \\ 2x_1 + 100.000x_2 &= 100.000 \end{aligned}$$

A eliminação dá

$$\begin{aligned} x_1 + \quad x_2 &= 2 \\ 100.000x_2 &= 100.000 \end{aligned}$$

o que pode ser resolvido dando a resposta correta: $x_1 = x_2 = 1$. Assim, a mudança de escala foi útil para determinar quando o pivotamento é necessário, mas não foi necessária uma mudança de escala nas próprias equações para se chegar a uma resposta correta.

Como no exemplo anterior, a mudança de escala tem utilidade na minimização de erros de arredondamento. No entanto, deve ser observado que a própria mudança de escala também leva a erros de arredondamento. Por exemplo, dada a equação

$$2x_1 + 300.000x_2 = 1$$

e usando três algarismos significativos, a mudança de escala leva a

$$0{,}00000667x_1 + x_2 = 0{,}00000333$$

Assim, a mudança de escala introduz um erro de arredondamento no primeiro coeficiente e na constante do lado direito. Por essa razão, às vezes recomendamos que a mudança de escala seja utilizada apenas como na parte (*c*) do exemplo anterior. Ou seja, é usada para calcular os valores dos coeficientes na nova escala simplesmente como um critério para pivotamento, mas os coeficientes originais são mantidos nos cálculos da eliminação e substituição propriamente ditos. Isso envolve prós e contras se o determinante estiver sendo calculado como parte do programa. Isto é, o determinante resultante não estará na nova escala. Entretanto, como muitas aplicações da eliminação de Gauss não exigem o cálculo do determinante, essa é a abordagem mais comum e será usada no algoritmo da próxima seção.

9.4.4 Algoritmo Computacional para a Eliminação de Gauss

Os algoritmos das Figuras 9.4 e 9.5 podem ser combinados agora em um algoritmo maior para implementar o algoritmo completo de eliminação de Gauss. A Figura 9.6 mostra um algoritmo para uma sub-rotina geral que implemente a eliminação de Gauss.

Observe que o programa inclui módulos para as três operações principais do algoritmo da eliminação de Gauss: eliminação progressiva, substituição regressiva e pivotamento. Além disso, temos diversos aspectos do código que diferem e são aperfeiçoamentos dos pseudocódigos das Figuras 9.4 e 9.5. São eles:

- Não são feitas mudanças de escala nas equações, mas os valores na nova escala dos elementos são usados para determinar quando o pivotamento deve ser implementado.
- O termo na diagonal é monitorado durante a fase de pivotamento para se detectar a ocorrência de valores próximos de zero e identificar sistemas singulares. Se o algoritmo devolver um valor $er = -1$, uma matriz singular foi detectada e os cálculos devem ser interrompidos. Um valor pequeno para o parâmetro *tol* é escolhido pelo usuário, para detectar ocorrências próximas de zero.

EXEMPLO 9.11 Solução de Equações Algébricas Lineares Usando o Computador

Enunciado do Problema. Um programa de computador para resolver equações algébricas lineares, como o baseado na Figura 9.6, pode ser usado para resolver o problema associado ao exemplo do pára-quedista em queda livre discutido no Capítulo 1. Suponha que um time de três pára-quedistas está ligado por uma corda sem peso enquanto cai, em queda livre, a uma velocidade de 5 m/s (Figura 9.7). Calcule a tensão em cada seção da corda e a aceleração do time, dado o seguinte:

Páraquedista	Massa, kg	Coeficiente de Arrasto, kg/s
1	70	10
2	60	14
3	40	17

Solução. Diagramas de corpo livre foram esboçados na Figura 9.8. Somando as forças na vertical e usando a segunda lei de Newton, tem-se um conjunto de três equações lineares simultâneas:

$$\begin{aligned} m_1 g - T - c_1 v &= m_1 a \\ m_2 g + T - c_2 v - R &= m_2 a \\ m_3 g - c_3 v + R &= m_3 a \end{aligned}$$



FIGURA 9.7
Três pára-quedistas em queda livre ligados por uma corda sem peso.

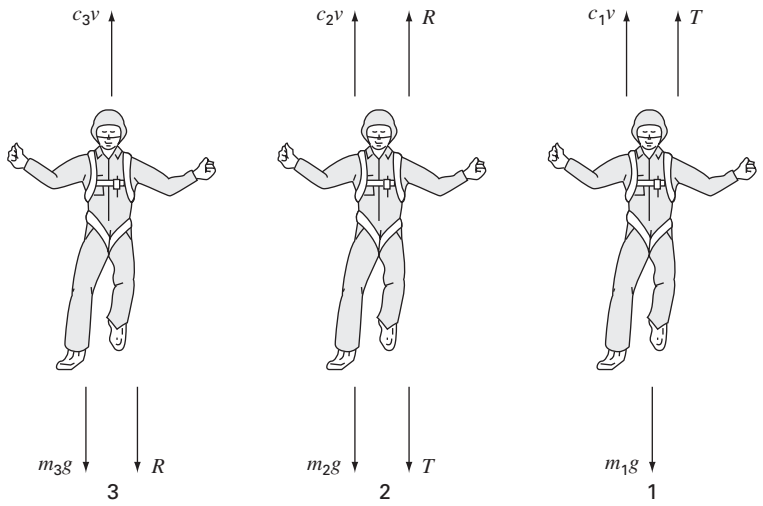


FIGURA 9.8
Diagramas de corpo livre para cada um dos três pára-quedistas em queda livre.

Essas equações possuem três incógnitas: *a, T* e *R*. Depois de substituir os valores conhecidos, as equações podem ser expressas na forma matricial ($g = 9{,}8$ m/s^2),

$$\begin{bmatrix} 70 & 1 & 0 \\ 60 & -1 & 1 \\ 40 & 0 & -1 \end{bmatrix} \begin{Bmatrix} a \\ T \\ R \end{Bmatrix} = \begin{Bmatrix} 636 \\ 518 \\ 307 \end{Bmatrix}$$

Esse sistema pode ser resolvido usando seu próprio software. O resultado é $a = 8{,}5941$ m/s^2; $T = 34{,}4118$ N e $R = 36{,}7647$ N.

9.5 SISTEMAS COMPLEXOS

Em alguns problemas, é possível obter um sistema de equações complexas

$$[C]\{Z\} = \{W\} \tag{9.27}$$

onde

$$\begin{aligned} [C] &= [A] + i[B] \\ \{Z\} &= \{X\} + i\{Y\} \\ \{W\} &= \{U\} + i\{V\} \end{aligned} \tag{9.28}$$

onde $i = \sqrt{-1}$.

A forma mais direta de resolver tal sistema é usar um dos algoritmos descritos nesta parte do livro, trocando as operações reais por operações complexas. É claro que isso só é possível para aquelas linguagens, tal como o Fortran, que permitem variáveis complexas.

Para linguagens que não permitem a declaração de variáveis complexas, é possível escrever um código para transformar operações reais em complexas. Contudo, essa não é uma tarefa trivial. Uma alternativa é transformar o sistema complexo em outro equivalente envolvendo apenas variáveis reais, o que pode ser feito substituindo-se a Equação (9.28) na Equação (9.27) e igualando as partes reais e imaginárias da equação resultante para obter

$$[A]\{X\} - [B]\{Y\} = \{U\} \tag{9.29}$$

e

$$[B]\{X\} + [A]\{Y\} = \{V\} \tag{9.30}$$

Assim, o sistema de n equações complexas se transforma em um conjunto de $2n$ equações reais. Isso significa que o espaço de armazenamento e o tempo de execução vão crescer significativamente. Conseqüentemente, há prós e contras relativos a essa opção. Se você calcula sistemas complexos raramente, é preferível usar as Equações (9.29) e (9.30) por causa de sua conveniência. Por outro lado, se você os usa muitas vezes e deseja utilizar uma linguagem que não permite dados do tipo complexo, então pode valer a pena o esforço de escrever um programa adaptado para resolver equações que transforme operações reais em complexas.

9.6 SISTEMAS DE EQUAÇÕES NÃO-LINEARES

Lembre-se de que no final do Capítulo 6 foi apresentada uma abordagem para resolver duas equações não-lineares com duas variáveis. Essa abordagem pode ser estendida ao caso geral de resolver n equações não-lineares simultâneas.

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0 \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \ldots, x_n) &= 0 \end{aligned} \tag{9.31}$$

A solução desse sistema consiste no conjunto dos valores de x para os quais todas as equações sejam simultaneamente nulas.

Como descrito na Seção 6.5.2, uma abordagem para resolver tais sistemas se baseia em uma versão multidimensional do método de Newton-Raphson. Assim, é escrita uma expansão em série de Taylor para cada equação. Por exemplo, para a k-ésima equação,

$$f_{k,i+1} = f_{k,i} + (x_{1,i+1} - x_{1,i})\frac{\partial f_{k,i}}{\partial x_1} + (x_{2,i+1} - x_{2,i})\frac{\partial f_{k,i}}{\partial x_2} + \cdots + (x_{n,i+1} - x_{n,i})\frac{\partial f_{k,i}}{\partial x_n} \tag{9.32}$$

onde o primeiro índice, k, representa a equação ou a variável e o segundo índice denota se o valor ou função em questão é o valor atual (i) ou o valor seguinte ($i + 1$).

Equações da forma de (9.32) são escritas para cada uma das equações não-lineares originais. Então, do mesmo modo como foi feito para deduzir a Equação (6.20) da Equação (6.19), todos os termos $f_{k,i+1}$ são igualados a zero, como seria o caso na raiz, e a Equação (9.32) pode ser escrita como

$$\begin{aligned} &-f_{k,i} + x_{1,i}\frac{\partial f_{k,i}}{\partial x_1} + x_{2,i}\frac{\partial f_{k,i}}{\partial x_2} + \cdots + x_{n,i}\frac{\partial f_{k,i}}{\partial x_n} \\ &\quad = x_{1,i+1}\frac{\partial f_{k,i}}{\partial x_1} + x_{2,i+1}\frac{\partial f_{k,i}}{\partial x_2} + \cdots + x_{n,i+1}\frac{\partial f_{k,i}}{\partial x_n} \end{aligned} \tag{9.33}$$

Observe que as únicas incógnitas na Equação (9.33) são os termos $x_{k,i+1}$ do lado direito da equação. Todas as outras quantidades são localizadas no valor atual (i) e, portanto, são conhecidas em qualquer iteração. Conseqüentemente, o conjunto de equações genericamente representado pela Equação (9.33) (isto é, com $k = 1, 2, \ldots, n$) constitui um conjunto de equações lineares simultâneas, que pode ser resolvido pelos métodos tratados nesta parte do livro.

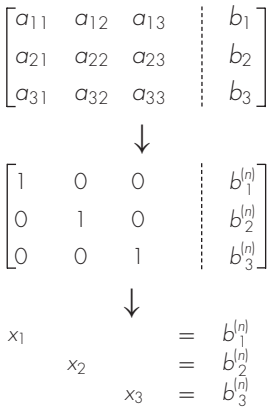


FIGURA 9.9
Descrição gráfica do método de Gauss-Jordan. Compare com a Figura 9.3 para observar as diferenças entre essa técnica e a eliminação de Gauss. O sobrescrito (*n*) significa que o elemento do lado direito da equação foi modificado *n* vezes (para esse caso, $n = 3$).

A notação matricial pode ser usada para expressar a Equação (9.33) de forma concisa. As derivadas parciais podem ser expressas por

$$[Z] = \begin{bmatrix} \dfrac{\partial f_{1,i}}{\partial x_1} & \dfrac{\partial f_{1,i}}{\partial x_2} & \cdots & \dfrac{\partial f_{1,i}}{\partial x_n} \\ \dfrac{\partial f_{2,i}}{\partial x_1} & \dfrac{\partial f_{2,i}}{\partial x_2} & \cdots & \dfrac{\partial f_{2,i}}{\partial x_n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \dfrac{\partial f_{n,i}}{\partial x_1} & \dfrac{\partial f_{n,i}}{\partial x_2} & \cdots & \dfrac{\partial f_{n,i}}{\partial x_n} \end{bmatrix} \tag{9.34}$$

Os valores inicial e final se escrevem na forma de vetores

$$\{X_i\}^T = \lfloor x_{1,i} \quad x_{2,i} \quad \cdots \quad x_{n,i} \rfloor$$

e

$$\{X_{i+1}\}^T = \lfloor x_{1,i+1} \quad x_{2,i+1} \quad \cdots \quad x_{n,i+1} \rfloor$$

Finalmente, os valores da função em i podem ser expressos por

$$\{F_i\}^T = \lfloor f_{1,i} \quad f_{2,i} \quad \cdots \quad f_{n,i} \rfloor$$

Usando essas relações, a Equação (9.33) é escrita de forma concisa como

$$[Z]\{X_{i+1}\} = -\{F_i\} + [Z]\{X_i\} \tag{9.35}$$

A Equação (9.35) pode ser resolvida usando-se uma técnica como a eliminação de Gauss. Esse processo pode ser repetido iterativamente para se obter estimativas refinadas, de forma parecida com o caso das duas equações tratado na Seção 6.5.2.

Deve-se observar que há dois grandes defeitos nessa abordagem anterior. Primeiro, a Equação (9.34) é, em geral, inconveniente de ser calculada. Assim, variações do método de Newton-Raphson foram desenvolvidas para contornar esse problema. Como seria de esperar, a maioria delas é baseada em aproximações por diferenças finitas das derivadas parciais que compõem [Z].

O segundo defeito do método de Newton-Raphson para equações múltiplas é que, em geral, são necessárias estimativas excelentes para os valores iniciais para garantir a convergência. Como esses valores são geralmente difíceis de obter, foram desenvolvidas outras abordagens alternativas que são mais lentas do que o método de Newton-Raphson, mas que têm um comportamento de convergência melhor. Uma abordagem comum é reformular o sistema não-linear como uma única função

$$F(x) = \sum_{i=1}^{n} [f_i(x_1, x_2, \ldots, x_n)]^2 \tag{9.36}$$

onde $f_i(x_1, x_2, \ldots, x_n)$ é o i-ésimo membro do sistema original da Equação (9.31). Os valores de x que minimizam essa função também representam a solução do sistema não-linear. Como veremos no Capítulo 17, essa reformulação pertence à classe de problemas chamada de *regressão não-linear*. Assim, poderá ser tratada com várias técnicas de otimização como as que serão descritas mais tarde neste texto (Parte Quatro e especificamente Capítulo 14).

9.7 GAUSS-JORDAN

O método de Gauss-Jordan é uma variação da eliminação de Gauss. A maior diferença é que, quando uma variável é eliminada no método de Gauss-Jordan, ela é eliminada de todas as outras equações, não só das posteriores. Além disso, todas as linhas são norma-

lizadas pela divisão pelo seu elemento pivô. Então, o passo de eliminação resulta na matriz identidade e não em uma matriz triangular (Figura 9.9). Conseqüentemente, não é necessário usar a substituição regressiva para obter o resultado. O método é mais bem ilustrado por um exemplo.

EXEMPLO 9.12 Método de Gauss-Jordan

Enunciado do Problema. Use a técnica de Gauss-Jordan para resolver o mesmo sistema do Exemplo 9.5:

$$3x_1 - 0{,}1x_2 - 0{,}2x_3 = 7{,}85$$
$$0{,}1x_1 + 7x_2 - 0{,}3x_3 = -19{,}3$$
$$0{,}3x_1 - 0{,}2x_2 + 10x_3 = 71{,}4$$

Solução. Primeiro, expresse os coeficientes e as constantes do lado direito da equação como uma matriz aumentada:

$$\begin{bmatrix} 3 & -0{,}1 & -0{,}2 & 7{,}85 \\ 0{,}1 & 7 & -0{,}3 & -19{,}3 \\ 0{,}3 & -0{,}2 & 10 & 71{,}4 \end{bmatrix}$$

Então, normalize a primeira linha dividindo-a pelo elemento pivô, 3, para obter

$$\begin{bmatrix} 1 & -0{,}0333333 & -0{,}066667 & 2{,}61667 \\ 0{,}1 & 7 & -0{,}3 & -19{,}3 \\ 0{,}3 & -0{,}2 & 10 & 71{,}4 \end{bmatrix}$$

O termo x_1 pode ser eliminado da segunda linha subtraindo 0,1 vez a primeira linha da segunda linha. De maneira semelhante, subtraindo 0,3 vez a primeira linha da terceira linha, iremos eliminar o termo x_1 da terceira linha:

$$\begin{bmatrix} 1 & -0{,}0333333 & -0{,}066667 & 2{,}61667 \\ 0 & 7{,}00333 & -0{,}293333 & -19{,}5617 \\ 0 & -0{,}190000 & 10{,}0200 & 70{,}6150 \end{bmatrix}$$

Em seguida, normalizamos a segunda linha, dividindo-a por 7,00333:

$$\begin{bmatrix} 1 & -0{,}0333333 & -0{,}066667 & 2{,}61667 \\ 0 & 1 & -0{,}0418848 & -2{,}79320 \\ 0 & -0{,}190000 & 10{,}0200 & 70{,}6150 \end{bmatrix}$$

A redução do termo x_2 da primeira e da terceira linhas resulta em

$$\begin{bmatrix} 1 & 0 & -0{,}0680629 & 2{,}52356 \\ 0 & 1 & -0{,}0418848 & -2{,}79320 \\ 0 & 0 & 10{,}01200 & 70{,}0843 \end{bmatrix}$$

A terceira linha é então normalizada dividindo-se por 10,0120:

$$\begin{bmatrix} 1 & 0 & -0{,}0680629 & 2{,}52356 \\ 0 & 1 & -0{,}0418848 & -2{,}79320 \\ 0 & 0 & 1 & 7{,}0000 \end{bmatrix}$$

Finalmente, o termo x_3 pode ser reduzido da primeira e da segunda equações dando

$$\begin{bmatrix} 1 & 0 & 0 & 3{,}0000 \\ 0 & 1 & 0 & -2{,}5000 \\ 0 & 0 & 1 & 7{,}0000 \end{bmatrix}$$

Assim, como descrito na Figura 9.8, a matriz dos coeficientes foi transformada na matriz identidade e a solução é obtida no vetor do lado direito. Observe que não foi necessária a substituição regressiva para se obter a solução.

Todo o material deste capítulo relativo a armadilhas e aprimoramentos do método de eliminação de Gauss também se aplica ao método de Gauss-Jordan. Por exemplo, uma estratégia de pivotamento parecida pode ser usada para evitar a divisão por zero e reduzir erros de arredondamento.

Apesar de a técnica de Gauss-Jordan e a eliminação de Gauss poderem parecer quase idênticas, a primeira requer mais trabalho. Usando uma abordagem semelhante à da Seção 9.2.1, pode-se determinar que o número de flops de multiplicação/divisão envolvidos no Gauss-Jordan ingênuo é

$$\frac{n^3}{2} + n^2 - \frac{n}{2} \xrightarrow{\text{quando } n \text{ cresce}} \frac{n^3}{2} + O(n^2) \tag{9.37}$$

Assim, o Gauss-Jordan envolve aproximadamente 50% mais operações do que a eliminação de Gauss [compare com a Equação (9.23)]. Portanto, a eliminação de Gauss é o método de eliminação simples preferido para se obter soluções de equações algébricas lineares. Entretanto, uma das razões principais para introduzirmos o método de Gauss-Jordan é que ele ainda é usado tanto em engenharia quanto em alguns algoritmos numéricos.

9.8 RESUMO

Em resumo, devotamos a maior parte deste capítulo à eliminação de Gauss, o método mais fundamental para resolver equações algébricas lineares simultâneas. Apesar de ser uma das técnicas mais antigas desenvolvidas para esse propósito, é, no entanto, um algoritmo extremamente efetivo para se obter soluções para muitos problemas da engenharia. Além dessa utilidade prática, este capítulo também fornece um contexto para a discussão de questões gerais como arredondamento, mudança de escala e condicionamento. Também, foi apresentado brevemente algum material sobre o método de Gauss-Jordan, assim como sobre sistemas complexos e não-lineares.

As respostas obtidas pela eliminação de Gauss podem ser verificadas por substituição nas equações originais. Entretanto, isso nem sempre representa uma verificação confiável para sistemas mal condicionados. Desse modo, deve ser calculada alguma medida do condicionamento, tal como o determinante do sistema depois da mudança de escala, se houver suspeita de erros de arredondamento. Usar pivotamento parcial e mais algarismos significativos nos cálculos são duas opções para diminuir os erros de arredondamento. No próximo capítulo, retornaremos ao tópico de condicionamento do sistema quando discutirmos a matriz inversa.

PROBLEMAS

9.1

(a) Escreva o seguinte conjunto de equações na forma matricial:

$$50 = 5x_3 + 2x_2$$

$$10 - x_1 = x_3$$

$$3x_2 + 8x_1 = 20$$

(b) Escreva a transposta da matriz dos coeficientes.

9.2 Diversas matrizes são definidas a seguir

$$[A] = \begin{bmatrix} 4 & 7 \\ 1 & 2 \\ 5 & 6 \end{bmatrix} \qquad [B] = \begin{bmatrix} 4 & 3 & 7 \\ 1 & 2 & 7 \\ 2 & 0 & 4 \end{bmatrix}$$

$$\{C\} = \begin{Bmatrix} 3 \\ 6 \\ 1 \end{Bmatrix} \qquad [D] = \begin{bmatrix} 9 & 4 & 3 & -6 \\ 2 & -1 & 7 & 5 \end{bmatrix}$$

$$[E] = \begin{bmatrix} 1 & 5 & 8 \\ 7 & 2 & 3 \\ 4 & 0 & 6 \end{bmatrix}$$

$$[F] = \begin{bmatrix} 3 & 0 & 1 \\ 1 & 7 & 3 \end{bmatrix} \qquad \lfloor G \rfloor = \lfloor 7 \quad 6 \quad 4 \rfloor$$

Responda às seguintes questões em relação a essas matrizes:

(a) Quais as dimensões dessas matrizes?

(b) Identifique as matrizes quadrada, linha e coluna.

(c) Qual o valor dos elementos: a_{12}, b_{23}, d_{32}, e_{22}, f_{12}, g_{12}?

(d) Efetue as seguintes operações:

(1) $[E] + [B]$ (7) $[B] \times [A]$

(2) $[A] + [F]$ (8) $[D]^T$

(3) $[B] - [E]$ (9) $[A] \times \{C\}$

(4) $7 \times [B]$ (10) $[I] \times [B]$

(5) $[E] \times [B]$ (11) $[E]^T[E]$

(6) $\{C\}^T$ (12) $\{C\}^T\{C\}$

9.3 Três matrizes são dadas por

$$[A] = \begin{bmatrix} 1 & 6 \\ 3 & 10 \\ 7 & 4 \end{bmatrix} \quad [B] = \begin{bmatrix} 1 & 3 \\ 0{,}5 & 2 \end{bmatrix} \quad [C] = \begin{bmatrix} 2 & -2 \\ -3 & 1 \end{bmatrix}$$

(a) Efetue todas as multiplicações possíveis entre pares dessas matrizes.
(b) Use o método do Quadro PT3.2 para justificar por que os outros pares não podem ser multiplicados.
(c) Use os resultados de **(a)** para ilustrar por que a ordem da multiplicação é importante.

9.4 Use o método gráfico para resolver

$$4x_1 - 8x_2 = -24$$
$$x_1 + 6x_2 = 34$$

Verifique seus resultados substituindo as soluções no sistema.

9.5 Dado o sistema de equações

$$-1{,}1x_1 + 10x_2 = 120$$
$$-2x_1 + 17{,}4x_2 = 174$$

(a) Resolva graficamente e verifique os resultados substituindo-os nas equações.
(b) Baseado no resultado gráfico, o que você espera em relação ao condicionamento do sistema?
(c) Calcule o determinante.
(d) Resolva por eliminação de variáveis.

9.6 Para o conjunto de equações

$$2x_2 + 5x_3 = 9$$
$$2x_1 + x_2 + x_3 = 9$$
$$3x_1 + x_2 = 10$$

(a) Calcule o determinante.
(b) Use a regra de Cramer para determinar os x's.
(c) Substitua seus resultados nas equações originais para verificá-los.

9.7 Dadas as equações

$$0{,}5x_1 - x_2 = -9{,}5$$
$$1{,}02x_1 - 2x_2 = -18{,}8$$

(a) Resolva graficamente.
(b) Calcule o determinante.
(c) Com base em **(a)** e **(b)**, o que você espera em relação ao condicionamento do sistema?
(d) Resolva por eliminação de variáveis.
(e) Resolva novamente com a_{11} levemente modificado para 0,52. Interprete seus resultados.

9.8 Dadas as equações

$$10x_1 + 2x_2 - x_3 = 27$$
$$-3x_1 - 6x_2 + 2x_3 = -61{,}5$$
$$x_1 + x_2 + 5x_3 = -21{,}5$$

(a) Resolva por eliminação de Gauss ingênua. Mostre todos os passos dos cálculos.
(b) Substitua seus resultados nas equações originais para verificar suas respostas.

9.9 Use eliminação de Gauss para resolver:

$$8x_1 + 2x_2 - 2x_3 = -2$$
$$10x_1 + 2x_2 + 4x_3 = 4$$
$$12x_1 + 2x_2 + 2x_3 = 6$$

Utilize pivotamento parcial e verifique suas respostas, substituindo-as nas equações originais.

9.10 Dado o sistema de equações

$$-3x_2 + 7x_3 = 2$$
$$x_1 + 2x_2 - x_3 = 3$$
$$5x_1 - 2x_2 = 2$$

(a) Calcule o determinante.
(b) Use a regra de Cramer para determinar os x's.
(c) Use eliminação de Gauss com pivotamento parcial para determinar os x's.
(d) Substitua seus resultados nas equações originais para verificar suas respostas.

9.11 Dadas as equações

$$2x_1 - 6x_2 - x_3 = -38$$
$$-3x_1 - x_2 + 7x_3 = -34$$
$$-8x_1 + x_2 - 2x_3 = -20$$

(a) Resolva por eliminação de Gauss com pivotamento parcial. Mostre todos os passos dos cálculos.
(b) Substitua seus resultados nas equações originais para verificar suas respostas.

9.12 Use eliminação de Gauss-Jordan para resolver:

$$2x_1 + x_2 - x_3 = 1$$
$$5x_1 + 2x_2 + 2x_3 = -4$$
$$3x_1 + x_2 + x_3 = 5$$

Não utilize pivotamento. Substitua seus resultados nas equações originais para verificar suas respostas.

9.13 Resolva:

$$x_1 + x_2 - x_3 = -3$$
$$6x_1 + 2x_2 + 2x_3 = 2$$
$$-3x_1 + 4x_2 + x_3 = 1$$

com **(a)** eliminação de Gauss ingênua, **(b)** eliminação de Gauss com pivotamento parcial e **(c)** Gauss-Jordan sem pivotamento parcial.

9.14 Efetue os mesmos cálculos do Exemplo 9.11, mas use cinco pára-quedistas com as seguintes características:

Pára-quedista	Massa, kg	Coeficiente de Arrasto, kg/s
1	55	10
2	75	12
3	60	15
4	75	16
5	90	10

Os pára-quedistas têm velocidade de 9 m/s.

9.15 Resolva

$$\begin{bmatrix} 3+2i & 4 \\ -i & 1 \end{bmatrix} \begin{Bmatrix} z_1 \\ z_2 \end{Bmatrix} = \begin{Bmatrix} 2+i \\ 3 \end{Bmatrix}$$

9.16 Desenvolva, *debug* e teste um programa em linguagem de alto nível ou em macro a sua escolha para multiplicar duas matrizes — isto é, $[X] = [Y][Z]$, onde $[Y]$ é uma matriz m por n e $[Z]$ é uma n por p. Teste seu programa usando as matrizes do Problema 9.3.

9.17 Desenvolva, *debug* e teste um programa em linguagem de alto nível ou em macro a sua escolha para gerar a transposta de uma matriz. Teste seu programa usando as matrizes do Problema 9.3.

9.18 Desenvolva, *debug* e teste um programa em linguagem de alto nível ou em macro a sua escolha para resolver um sistema de equações com eliminação de Gauss e com pivotamento parcial. Use como base o pseudocódigo da Figura 9.6. Teste seu programa utilizando o seguinte sistema (cuja resposta é $x_1 = x_2 = x_3 = 1$),

$$x_1 + 2x_2 - x_3 = 2$$
$$5x_1 + 2x_2 + 2x_3 = 9$$
$$-3x_1 + 5x_2 - x_3 = 1$$

CAPÍTULO 10

Decomposição *LU* e Inversão de Matrizes

Este capítulo trata de uma classe de métodos de eliminação chamados de técnicas de decomposição *LU*. O interesse principal da decomposição *LU* é que o passo de eliminação, que consome muito tempo, pode ser formulado de modo que envolva apenas operações na matriz dos coeficientes, $[A]$. Assim, ele se adapta bem àquelas situações nas quais muitos vetores à direita, $\{B\}$, devem ser calculados para um único valor de $[A]$. Apesar de haver uma variedade de maneiras pelas quais isso pode ser feito, vamo-nos concentrar em mostrar como o método de eliminação de Gauss pode ser implementado como uma decomposição *LU*.

Um motivo para introduzir a decomposição *LU* é que ela fornece uma maneira eficiente de calcular a matriz inversa, a qual tem um grande número de aplicações valiosas na prática da engenharia. A decomposição *LU* também fornece um meio de avaliar o condicionamento do sistema.

10.1 DECOMPOSIÇÃO *LU*

Como descrito no capítulo anterior, a eliminação de Gauss é projetada para resolver sistemas de equações algébricas lineares,

$$[A]\{X\} = \{B\} \tag{10.1}$$

Embora certamente represente um método seguro de resolver tais sistemas, ele se torna ineficiente ao se resolver equações com os mesmos coeficientes $[A]$, mas com diferentes constantes à direita (os b's).

Lembre-se que a eliminação de Gauss envolve dois passos: a eliminação progressiva e a substituição regressiva (Figura 9.3). Desses, a eliminação progressiva compreende o maior volume de esforço computacional (lembre-se da Tabela 9.1). Isso é particularmente verdadeiro para grandes sistemas de equações.

Os métodos de decomposição LU separam a eliminação da matriz $[A]$, que consome tempo, das manipulações do lado direito $\{B\}$. Assim, uma vez que $[A]$ tenha sido "decomposta", vetores múltiplos à direita podem ser calculados de maneira eficiente.

É interessante observar que a própria eliminação de Gauss pode ser escrita como uma decomposição *LU*. Antes de mostrar como isso pode ser feito, inicialmente será fornecida uma visão matemática geral da estratégia da decomposição.

10.1.1 Uma Visão Geral da Decomposição *LU*

Do mesmo modo como no caso da eliminação de Gauss, a decomposição *LU* requer pivotamento para se evitar divisões por zero. Entretanto, para simplificar a descrição a seguir, a questão do pivotamento será adiada até que a abordagem principal seja elaborada. Além disso, a explicação seguinte está limitada a um conjunto de três equações simultâneas. Os resultados podem ser estendidos diretamente a sistemas de n dimensões.

A Equação (10.1) pode ser reorganizada como

$$[A]\{X\} - \{B\} = 0 \tag{10.2}$$

Suponha que a Equação (10.2) possa ser escrita como um sistema triangular superior:

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} \tag{10.3}$$

Perceba que isso é parecido com os cálculos que ocorrem no primeiro passo da eliminação de Gauss. Ou seja, a eliminação é usada para reduzir o sistema a uma forma triangular superior. A Equação (10.3) também pode ser escrita em notação matricial e reorganizada para dar

$$[U]\{X\} - \{D\} = 0 \tag{10.4}$$

Agora, suponha que exista uma matriz triangular inferior com 1's na diagonal,

$$[L] = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \tag{10.5}$$

que tenha a propriedade de que, quando o primeiro membro da Equação (10.4) for multiplicado à esquerda por ela, o resultado seja o primeiro membro da Equação (10.2). Isto é,

$$[L]\{[U]\{X\} - \{D\}\} = [A]\{X\} - \{B\} \tag{10.6}$$

Se essa equação for válida para todo {X}, segue das regras de multiplicação de matrizes que

$$[L][U] = [A] \tag{10.7}$$

e

$$[L]\{D\} = \{B\} \tag{10.8}$$

Uma estratégia de dois passos (ver Figura 10.1) para obter soluções pode basear-se nas Equações (10.4), (10.7) e (10.8):

1. *Passo da decomposição LU*. [*A*] é fatorada ou "decomposta" em matrizes triangulares inferior [*L*] e superior [*U*].
2. *Passo de substituição.* [*L*] e [*U*] são usadas para determinar a solução {*X*} para um lado direito {*B*}. Esse passo consiste em duas etapas. Primeiro, a Equação (10.8) é usada para gerar um vetor intermediário {*D*} por substituição progressiva. A seguir, o resultado é substituído na Equação (10.4), que pode ser resolvida por substituição regressiva, determinando-se {*X*}.

Agora, vejamos como a eliminação de Gauss pode ser implementada dessa forma.

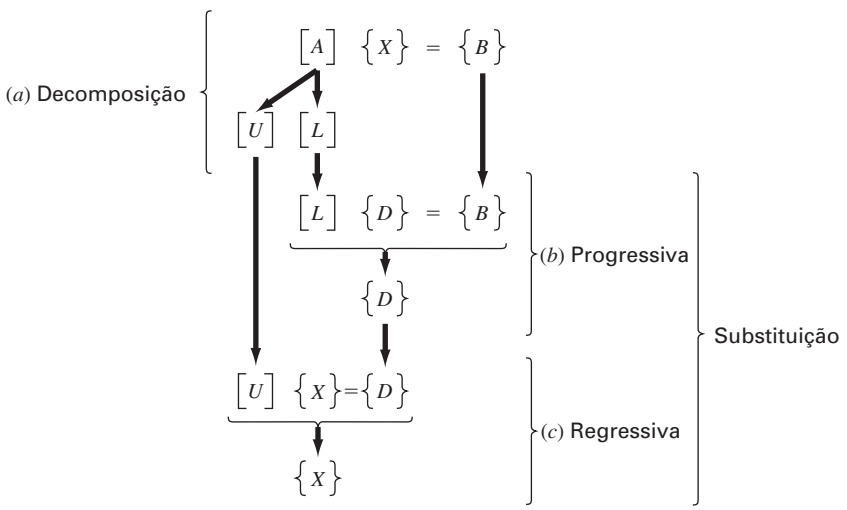


FIGURA 10.1
Os passos da decomposição *LU*.

10.1.2 A Versão da Eliminação de Gauss por Decomposição *LU*

Embora à primeira vista pareça não estar relacionada com a decomposição *LU*, a eliminação de Gauss pode ser usada para decompor [*A*] em [*L*] e [*U*]. Isso pode ser visto facilmente para [*U*], que é um produto direto da eliminação progressiva. Lembre-se que o passo de eliminação progressiva tem por objetivo reduzir a matriz original dos coeficientes [*A*] para a forma

$$[U] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a'_{22} & a'_{23} \\ 0 & 0 & a''_{33} \end{bmatrix} \tag{10.9}$$

que está na forma triangular superior procurada.

Embora possa não ser tão aparente, a matriz [*L*] também é produzida durante essa etapa. Isso pode ser prontamente ilustrado para um sistema de três equações,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} b_1 \\ b_2 \\ b_3 \end{Bmatrix}$$

A primeira etapa da eliminação de Gauss é multiplicar a linha 1 pelo fator [lembre-se da Equação (9.13)]

$$f_{21} = \frac{a_{21}}{a_{11}}$$

e subtrair o resultado da segunda linha para eliminar a_{21}. Analogamente, a linha 1 é multiplicada por

$$f_{31} = \frac{a_{31}}{a_{11}}$$

e o resultado é subtraído da terceira linha para eliminar a_{31}. O passo final é multiplicar a segunda linha modificada por

$$f_{32} = \frac{a'_{32}}{a'_{22}}$$

e subtrair o resultado da terceira linha para eliminar a'_{32}.

Agora, suponha que foram meramente efetuadas todas essas manipulações na matriz [*A*]. Claramente, se não se quer modificar a equação, também se deve fazer o mesmo com o lado direito {*B*}. Mas não há nenhuma razão que obrigue a efetuar essas operações simultaneamente. Assim, pode-se salvar os f's e manipular {*B*} depois.

Onde serão armazenados os fatores f_{21}, f_{31} e f_{32}? Lembre-se que a idéia por trás da eliminação era criar zeros em a_{21}, a_{31} e a_{32}. Assim, é possível armazenar f_{21} em a_{21}, f_{31} em a_{31} e f_{32} em a_{32}. Depois da eliminação, a matriz [*A*] poderia, portanto, ser escrita como

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ f_{21} & a'_{22} & a'_{23} \\ f_{31} & f_{32} & a''_{33} \end{bmatrix} \tag{10.10}$$

Essa matriz, de fato, representa um armazenamento eficiente da decomposição *LU* de [*A*],

$$[A] \rightarrow [L][U] \tag{10.11}$$

onde

$$[U] = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a'_{22} & a'_{23} \\ 0 & 0 & a''_{33} \end{bmatrix}$$

e

$$[L] = \begin{bmatrix} 1 & 0 & 0 \\ f_{21} & 1 & 0 \\ f_{31} & f_{32} & 1 \end{bmatrix}$$

O exemplo a seguir confirma que $[A] = [L][U]$.

EXEMPLO 10.1

Decomposição *LU* com Eliminação de Gauss

Enunciado do Problema. Obtenha uma decomposição *LU* baseada na eliminação de Gauss efetuada no Exemplo 9.5.

Solução. No Exemplo 9.5, foi resolvido um sistema com a matriz

$$[A] = \begin{bmatrix} 3 & -0,1 & -0,2 \\ 0,1 & 7 & -0,3 \\ 0,3 & -0,2 & 10 \end{bmatrix}$$

Após a eliminação progressiva, foi obtida a seguinte matriz triangular superior:

$$[U] = \begin{bmatrix} 3 & -0,1 & -0,2 \\ 0 & 7,00333 & -0,293333 \\ 0 & 0 & 10,0120 \end{bmatrix}$$

Os fatores empregados para obter a matriz triangular superior podem ser reunidos em uma matriz triangular inferior. Os elementos a_{21} e a_{31} foram eliminados usando-se os fatores

$$f_{21} = \frac{0,1}{3} = 0,03333333 \qquad f_{31} = \frac{0,3}{3} = 0,1000000$$

e o elemento a'_{32} foi eliminado usando-se o fator

$$f_{32} = \frac{-0,19}{7,00333} = -0,0271300$$

Assim, a matriz triangular inferior é

$$[L] = \begin{bmatrix} 1 & 0 & 0 \\ 0,0333333 & 1 & 0 \\ 0,100000 & -0,0271300 & 1 \end{bmatrix}$$

Conseqüentemente, a decomposição *LU* é

$$[A] = [L][U] = \begin{bmatrix} 1 & 0 & 0 \\ 0,0333333 & 1 & 0 \\ 0,100000 & -0,0271300 & 1 \end{bmatrix} \begin{bmatrix} 3 & -0,1 & -0,2 \\ 0 & 7,00333 & -0,293333 \\ 0 & 0 & 10,0120 \end{bmatrix}$$

Esse resultado pode ser verificado efetuando-se a multiplicação de $[L][U]$ para obter

$$[L][U] = \begin{bmatrix} 3 & -0,1 & -0,2 \\ 0,0999999 & 7 & -0,3 \\ 0,3 & -0,2 & 9,99996 \end{bmatrix}$$

cujas pequenas discrepâncias se devem a erros de arredondamento.

A seguir tem-se um pseudocódigo para uma sub-rotina que implemente a fase de decomposição:

```
SUB Decompose (a, n)
  DOFOR k = 1, n − 1
    DOFOR i = k + 1, n
      fator = a_{i,k}/a_{k,k}
      a_{i,k} = fator
      DOFOR j = k + 1, n
        a_{i,j} = a_{i,j} − fator * a_{k,j}
      END DO
    END DO
  END DO
END Decompose
```

Observe que esse algoritmo é "ingênuo" no sentido que o pivotamento não está incluído. Esse recurso será acrescentado mais tarde, quando será desenvolvido um algoritmo completo para a decomposição *LU*.

Depois que a matriz é decomposta, pode-se produzir uma solução para um valor particular do vetor à direita $\{B\}$, o que é feito em duas etapas. Primeiro, uma substituição progressiva é executada, resolvendo-se a Equação (10.8) para $\{D\}$. É importante reconhecer que isso significa simplesmente efetuar as manipulações da eliminação em $\{B\}$. Assim, ao final desse passo, o lado direito estará no mesmo estado em que estaria se tivesse sido efetuada a eliminação progressiva em $[A]$ e $\{B\}$ simultaneamente.

O passo de substituição progressiva pode ser representado de modo conciso por

$$d_i = b_i - \sum_{j=1}^{i-1} a_{ij} d_j \qquad \text{para } i = 2, 3, \ldots, n \tag{10.12}$$

O segundo passo, então, resume-se simplesmente em implementar a substituição regressiva, como na Equação (10.4). Novamente, é importante perceber que isso é idêntico à fase de substituição regressiva da eliminação de Gauss convencional. Assim, de maneira similar às Equações (9.16) e (9.17), o passo de substituição regressiva pode ser representado concisamente por

$$x_n = d_n / a_{nn} \tag{10.13}$$

$$x_i = \frac{d_i - \sum_{j=i+1}^{n} a_{ij} x_j}{a_{ii}} \qquad \text{para } i = n-1, n-2, \ldots, 1 \tag{10.14}$$

EXEMPLO 10.2 Os Passos de Substituição

Enunciado do Problema. Complete o problema iniciado no Exemplo 10.1, determinando a solução final por substituição progressiva e regressiva.

Solução. Como foi dito, o objetivo da substituição progressiva é impor as manipulações da eliminação, que foram aplicadas anteriormente a $[A]$, no vetor do lado direito, $\{B\}$. Lembre-se que o sistema resolvido no Exemplo 9.5 é

$$\begin{bmatrix} 3 & -0{,}1 & -0{,}2 \\ 0{,}1 & 7 & -0{,}3 \\ 0{,}3 & -0{,}2 & 10 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 7{,}85 \\ -19{,}3 \\ 71{,}4 \end{Bmatrix}$$

e que a fase de eliminação progressiva da eliminação de Gauss convencional resultou em

$$\begin{bmatrix} 3 & -0{,}1 & -0{,}2 \\ 0 & 7{,}00333 & -0{,}293333 \\ 0 & 0 & 10{,}0120 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 7{,}85 \\ -19{,}5617 \\ 70{,}0843 \end{Bmatrix} \tag{E10.2.1}$$

A fase de substituição progressiva é implementada aplicando-se a Equação (10.7) ao problema,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0{,}0333333 & 1 & 0 \\ 0{,}100000 & -0{,}0271300 & 1 \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} = \begin{Bmatrix} 7{,}85 \\ -19{,}3 \\ 71{,}4 \end{Bmatrix}$$

ou multiplicando o lado esquerdo,

$$\begin{aligned} d_1 &= 7{,}85 \\ 0{,}0333333 d_1 + d_2 &= -19{,}3 \\ 0{,}1 d_1 - 0{,}02713 d_2 + d_3 &= 71{,}4 \end{aligned}$$

Pode-se determinar d_1 a partir da primeira equação,

$$d_1 = 7{,}85$$

o que pode ser substituído na segunda equação para determinar

$$d_2 = -19{,}3 - 0{,}0333333(7.85) = -19{,}5617$$

Ambos, d_1 e d_2, podem então ser substituídos na terceira equação para dar

$$d_3 = 71{,}4 - 0{,}1(7{,}85) + 0{,}02713(-19{,}5617) = 70{,}0843$$

Assim,

$$\{D\} = \begin{Bmatrix} 7{,}85 \\ -19{,}5617 \\ 70{,}0843 \end{Bmatrix}$$

que é idêntico ao lado direito da Equação (E10.2.1).

Esse resultado pode ser substituído na Equação (10.4), $[U]\{X\} = \{D\}$, para dar

$$\begin{bmatrix} 3 & -0{,}1 & -0{,}2 \\ 0 & 7{,}00333 & -0{,}293333 \\ 0 & 0 & 10{,}0120 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 7{,}85 \\ -19{,}5617 \\ 70{,}0843 \end{Bmatrix}$$

que pode ser resolvida por substituição regressiva (ver Exemplo 9.5 para detalhes) para determinar a solução final,

$$\{X\} = \begin{Bmatrix} 3 \\ -2{,}5 \\ 7{,}00003 \end{Bmatrix}$$

A seguir tem-se um pseudocódigo para uma sub-rotina que implementa as duas fases de substituição:

```
SUB Substitute (a, n, b, x)
  'substituição progressiva
  DOFOR i = 2, n
    soma = b_i
    DOFOR j = 1, i − 1
      soma = soma − a_i,j * b_j
    END DO
    b_i = soma
  END DO
  'substituição regressiva
  x_n = b_n / a_n,n
  DOFOR i = n − 1, 1, −1
    soma = 0
    DOFOR j = i + 1, n
      soma = soma + a_i,j * x_j
    END DO
    x_i = (b_i − soma) / a_i,i
  END DO
END Substitute
```

O algoritmo de decomposição *LU* exige o mesmo número de flops de multiplicação/divisão que a eliminação de Gauss. A única diferença é que um pouco menos de esforço é gasto na fase de decomposição, já que as operações não são aplicadas ao lado direito. Assim, o número de flops de multiplicação/divisão envolvidos na fase de decomposição pode ser calculado por

$$\frac{n^3}{3} - \frac{n}{3} \xrightarrow{\text{quando } n \text{ cresce}} \frac{n^3}{3} + O(n) \tag{10.15}$$

Reciprocamente, a fase de substituição gasta um pouco mais de esforço. Assim, o número de flops para substituição progressiva e regressiva é n^2. O esforço total é, portanto, idêntico ao da eliminação de Gauss

$$\frac{n^3}{3} - \frac{n}{3} + n^2 \xrightarrow{\text{quando } n \text{ cresce}} \frac{n^3}{3} + O(n^2) \tag{10.16}$$

10.1.3 Algoritmo de Decomposição *LU*

Um algoritmo para implementar a decomposição *LU* com base na eliminação de Gauss é descrito na Figura 10.2. Quatro características desse algoritmo merecem menção:

- Os fatores gerados durante a fase de eliminação são armazenados na parte de baixo da matriz. Isso pode ser feito porque eles são transformados em zeros de qualquer forma e não são necessários para a solução final. O armazenamento economiza espaço.
- Esse algoritmo mantém as informações do pivotamento usando um vetor de ordem *o*. Isso acelera bastante o algoritmo porque apenas o vetor de ordem (ao contrário de toda a linha) é pivotado.
- As equações não sofrem mudança de escala, mas valores dos elementos em uma nova escala são usados para determinar quando o pivotamento deve ser implementado.
- O termo diagonal é monitorado durante a fase de pivotamento para se detectar ocorrências próximas de zero para sinalizar sistemas singulares. Se ele devolver um valor $er = -1$, uma matriz singular foi detectada e os cálculos devem parar. O usuário escolhe um pequeno valor para um parâmetro *tol* com a finalidade de detectar ocorrências próximas de zero.

10.1.4 Decomposição de Crout

Observe que, para a implementação da decomposição *LU* na eliminação de Gauss, a matriz $[L]$ tem 1's na diagonal. Isso é formalmente chamado de *decomposição de Doolittle* ou fatoração. Uma abordagem alternativa envolve a matriz $[U]$ com 1's na diagonal, o que é chamado de *decomposição de Crout*. Embora haja algumas diferenças entre essas abordagens (Atkinson, 1978; Ralston e Rabinowitz, 1978), seus desempenhos são comparáveis.

A abordagem da decomposição de Crout gera $[U]$ e $[L]$ varrendo a matriz pelas colunas e pelas linhas, como descrito na Figura 10.3. Ela pode ser implementada pela seguinte série de fórmulas concisas:

$$l_{i,1} = a_{i,1} \qquad \text{para } i = 1, 2, \ldots, n \tag{10.17}$$

$$u_{1j} = \frac{a_{1j}}{l_{11}} \qquad \text{para } j = 2, 3, \ldots, n \tag{10.18}$$

Para $j = 2, 3, \ldots, n-1$

$$l_{ij} = a_{ij} - \sum_{k=1}^{j-1} l_{ik} u_{kj} \qquad \text{para } i = j, j+1, \ldots, n \tag{10.19}$$

$$u_{jk} = \frac{a_{jk} - \sum_{i=1}^{j-1} l_{ji} u_{ik}}{l_{jj}} \qquad \text{para } k = j+1, j+2, \ldots, n \tag{10.20}$$

e

$$l_{nn} = a_{nn} - \sum_{k=1}^{n-1} l_{nk} u_{kn} \tag{10.21}$$

```
SUB Ludecomp (a, b, n, tol, x, er)
  DIM o_n, s_n
  er = 0
  CALL Decompose(a, n, tol, o, s, er)
  IF er <> −1 THEN
     CALL Substitute(a, o, n, b, x)
  END IF
END Ludecomp

SUB Decompose (a, n, tol, o, s, er)
  DOFOR i = 1, n
    o_i = i
    s_i = ABS(a_i,1)
    DOFOR j = 2, n
       IF ABS(a_i,j)>s_i THEN s_i = ABS(a_i,j)
    END DO
  END DO
  DOFOR k = 1, n − 1
    CALL Pivot(a, o, s, n, k)
    IF ABS(a_o(k),k/s_o(k)) < tol THEN
       er = −1
       PRINT a_o(k),k/s_o(k)
       EXIT DO
    END IF
    DOFOR i = k + 1, n
       fator = a_o(i),k/a_o(k),k
       a_o(i),k = fator
       DOFOR j = k + 1, n
         a_o(i),j = a_o(i),j − factor * a_o(k),j
       END DO
    END DO
  END DO
  IF ABS(a_o(k),k/s_o(k)) < tol THEN
     er = −1
     PRINT a_o(k),k/s_o(k)
  END IF
END Decompose

SUB Pivot (a, o, s, n, k)
  p = k
  maior = ABS(a_o(k),k/s_o(k))
  DOFOR ii = k + 1, n
    dummy = ABS(a_o(ii),k/s_o(ii))
    IF dummy > maior THEN
       maior = dummy
       p = ii
    END IF
  END DO
  dummy = o_p
  o_p = o_k
  o_k = dummy
END Pivot

SUB Substitute (a, o, n, b, x)
  DOFOR i = 2, n
    soma = b_o(i)
    DOFOR j = 1, i − 1
      soma = soma − a_o(i),j * b_o(j)
    END DO
    b_o(i) = soma
  END DO
  x_n = b_o(n)/a_o(n),n
  DOFOR i = n − 1, 1, −1
    soma = 0
    DOFOR j = i + 1, n
      soma = soma + a_o(i),j * x_j
    END DO
    x_i = (b_o(i) − soma)/a_o(i),i
  END DO
END Substitute
```

FIGURA 10.2
Pseudocódigo para um algoritmo da decomposição *LU*.

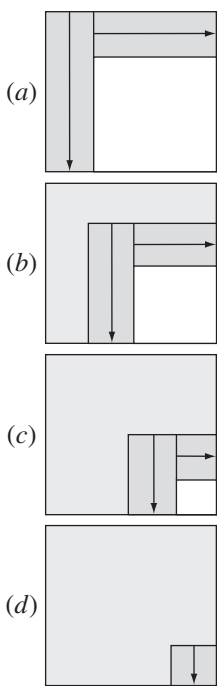


FIGURA 10.3
Um esquema descrevendo os cálculos envolvidos na decomposição *LU* de Crout.

Além do fato de consistir em uns poucos laços concisos, a abordagem precedente tem o benefício de que se pode economizar espaço de armazenamento. Não há necessidade de armazenar os 1's na diagonal de $[U]$ ou os zeros na diagonal de $[L]$ ou $[U]$ pois eles são dados pelo método. Conseqüentemente, os valores de $[U]$ podem ser armazenados nos espaços de zeros de $[L]$. Além disso, um exame mais profundo do processo precedente torna claro que, depois que cada elemento de $[A]$ for usado uma só vez, ele nunca mais será utilizado. Assim, à medida que cada elemento de $[L]$ e $[U]$ for calculado, ele pode substituir o elemento correspondente (de acordo com os índices) de $[A]$.

Um pseudocódigo para executar isso é apresentado na Figura 10.4. Observe que a Equação (10.17) não é incluída no pseudocódigo porque a primeira coluna de $[L]$ já está armazenada em $[A]$. Em vez disso, o algoritmo decorre diretamente das Equações (10.18) a (10.21).

```
DOFOR j = 2, n
  a1,j = a1,j/a1,1
END DO
DOFOR j = 2, n − 1
  DOFOR i = j, n
    soma = 0
    DOFOR k = 1, j − 1
      soma = soma + ai,k · ak,j
    END DO
    ai,j = ai,j − soma
  END DO
  DOFOR k = j + 1, n
    soma = 0
    DOFOR i = 1, j − 1
      soma = soma + aj,i · ai,k
    END DO
    aj,k = (aj,k − soma
  END DO
END DO
soma = 0
DOFOR k = 1, n − 1
  soma = soma + an,k · ak,n
END DO
an,n = an,n − soma
```

FIGURA 10.4
Pseudocódigo para algoritmo de decomposição *LU* de Crout.

10.2 A MATRIZ INVERSA

Na discussão sobre operações com matrizes (Seção PT3.2.2), foi introduzida a noção de que, se a matriz $[A]$ for quadrada, existe outra matriz, $[A]^{-1}$, chamada inversa de $[A]$, para a qual [Equação (PT3.3)]

$$[A][A]^{-1} = [A]^{-1}[A] = [I]$$

Agora, o foco será em como calcular essa inversa numericamente. A seguir, será explorado como isso pode ser usado para análise de engenharia.

10.2.1 Calculando a Inversa

A inversa pode ser calculada coluna a coluna, gerando soluções para vetores unitários como constantes do lado direito. Por exemplo, se o vetor do lado direito tem 1 na primeira posição e zeros no resto,

$$\{b\} = \begin{Bmatrix} 1 \\ 0 \\ 0 \end{Bmatrix}$$

a solução resultante será a primeira coluna da matriz inversa. Analogamente, se um vetor unitário com um 1 na segunda linha for usado

$$\{b\} = \begin{Bmatrix} 0 \\ 1 \\ 0 \end{Bmatrix}$$

o resultado será a segunda coluna da matriz inversa.

A melhor forma de implementar tais cálculos é com o algoritmo de decomposição *LU* descrito no começo do capítulo. Lembre-se que uma das grandes vantagens da decomposição *LU* é que ela fornece uma forma muito eficiente de calcular diversos vetores do lado direito. Assim, ela é ideal para manipular os vários vetores unitários necessários para calcular a inversa.

EXEMPLO 10.3 Inversão de Matrizes

Enunciado do Problema. Utilize a decomposição LU para determinar a inversa da matriz do sistema do Exemplo 10.2.

$$[A] = \begin{bmatrix} 3 & -0,1 & -0,2 \\ 0,1 & 7 & -0,3 \\ 0,3 & -0,2 & 10 \end{bmatrix}$$

Lembre-se de que a decomposição resultou nas seguintes matrizes triangulares superior e inferior:

$$[U] = \begin{bmatrix} 3 & -0,1 & -0,2 \\ 0 & 7,00333 & -0,293333 \\ 0 & 0 & 10,0120 \end{bmatrix} \quad [L] = \begin{bmatrix} 1 & 0 & 0 \\ 0,0333333 & 1 & 0 \\ 0,100000 & -0,0271300 & 1 \end{bmatrix}$$

Solução. A primeira coluna da matriz inversa pode ser determinada efetuando-se o procedimento de solução da substituição progressiva com um vetor unitário (com o 1 na primeira linha) como vetor do lado direito. Assim, a Equação (10.8), o sistema triangular inferior, pode ser escrita como

$$\begin{bmatrix} 1 & 0 & 0 \\ 0,0333333 & 1 & 0 \\ 0,100000 & -0,0271300 & 1 \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 0 \\ 0 \end{Bmatrix}$$

e resolvido com substituição progressiva para determinar $\{D\}^T = \lfloor 1 \ \ -0,03333 \ \ -0,1009 \rfloor$. Esse vetor pode ser usado como lado direito na Equação (10.3),

$$\begin{bmatrix} 3 & -0,1 & -0,2 \\ 0 & 7,00333 & -0,293333 \\ 0 & 0 & 10,0120 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ -0,03333 \\ -0,1009 \end{Bmatrix}$$

a qual pode ser resolvida por substituição regressiva para determinar $\{X\}^T = \lfloor 0,33249 \ \ -0,00518 \ \ -0,01008 \rfloor$, que é a primeira coluna da matriz,

$$[A]^{-1} = \begin{bmatrix} 0,33249 & 0 & 0 \\ -0,00518 & 0 & 0 \\ -0,01008 & 0 & 0 \end{bmatrix}$$

Para determinar a segunda coluna, a Equação (10.8) é formulada como

$$\begin{bmatrix} 1 & 0 & 0 \\ 0,0333333 & 1 & 0 \\ 0,100000 & -0,0271300 & 1 \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 1 \\ 0 \end{Bmatrix}$$

Isso pode ser usado para determinar $\{D\}$, e os resultados podem ser utilizados na Equação (10.3) para determinar $\{X\}^T = \lfloor 0,004944 \ \ 0,142903 \ \ 0,00271 \rfloor$, que é a segunda coluna da matriz,

$$[A]^{-1} = \begin{bmatrix} 0,33249 & 0,004944 & 0 \\ -0,00518 & 0,142903 & 0 \\ -0,01008 & 0,00271 & 0 \end{bmatrix}$$

Finalmente, o procedimento de substituição progressiva e regressiva pode ser implementado com $\{B\}^T = \lfloor 0 \ \ 0 \ \ 1 \rfloor$ para determinar $\{X\}^T = \lfloor 0,006798 \ \ 0,004183 \ \ 0,09988 \rfloor$, que é a última coluna da matriz,

$$[A]^{-1} = \begin{bmatrix} 0,33249 & 0,004944 & 0,006798 \\ -0,00518 & 0,142903 & 0,004183 \\ -0,01008 & 0,00271 & 0,09988 \end{bmatrix}$$

A validade desse resultado é verificada por $[A][A]^{-1} = [I]$.

```
CALL Decompose (a, n, tol, o, s, er)
IF er = 0 THEN
   DOFOR i = 1, n
     DOFOR j = 1, n
       IF i = j THEN
          b(j) = 1
       ELSE
          b(j) = 0
       END IF
     END DO
     CALL Substitute (a, o, n, b, x)
     DOFOR j = 1, n
       ai(j, i) = x(j)
     END DO
   END DO
   Output ai, se desejado
ELSE
   PRINT "sistema mal condicionado"
END IF
```

FIGURA 10.5
Programa principal que usa alguns dos subprogramas da Figura 10.2 para gerar a matriz inversa.

Um pseudocódigo para determinar a inversa de uma matriz é mostrado na Figura 10.5. Observe como a sub-rotina de decomposição da Figura 10.2 é chamada para executar a decomposição e então determinar a inversa, chamando repetidamente o algoritmo de substituição com vetores unitários.

O esforço exigido por esse algoritmo é calculado simplesmente por

$$\begin{array}{ccccc} \dfrac{n^3}{3} - \dfrac{n}{3} & + & n(n^2) & = \dfrac{4n^3}{3} - \dfrac{n}{3} \\ \text{decomposição} & + & n \times \text{substituições} \end{array} \tag{10.22}$$

em que, da Seção 10.1.2, a decomposição é definida pela Equação (10.15) e o esforço envolvido na manipulação de cada lado à direita envolve n^2 flops de multiplicação/divisão.

10.2.2 Cálculos de Resposta a Estímulos

Como foi discutido na Seção PT3.1.2, muitos dos sistemas de equações lineares confrontados na prática da engenharia são originários das leis de conservação. A expressão matemática dessas leis é alguma forma de equação de balanço para garantir que uma propriedade particular — massa, força, calor, momento, ou outra — seja conservada. Para um balanço de forças em uma estrutura, as propriedades poderiam ser as componentes horizontais ou verticais das forças agindo sobre cada nó da estrutura (ver Seção 12.2). Para um balanço de massas, as propriedades poderiam ser a massa em cada reator do processo químico (ver Seção 12.1). Outras áreas da engenharia vão dar origem a exemplos semelhantes.

Uma única equação de balanço pode ser escrita para cada parte do sistema, resultando em um conjunto de equações que define o comportamento da propriedade para o sistema inteiro. Essas equações são inter-relacionadas, ou acopladas, no sentido de que cada equação pode incluir uma ou mais das variáveis das outras equações. Em muitos casos, esses sistemas são lineares e, portanto, da forma exata tratada neste capítulo:

$$[A]\{X\} = \{B\} \tag{10.23}$$

Agora, para equações de balanço, os termos da Equação (10.23) têm uma interpretação física definida. Por exemplo, os elementos de $\{X\}$ são os níveis da propriedade

sendo balanceada para cada parte do sistema. Em um balanço de forças em uma estrutura, representam as forças horizontais e verticais em cada membro. Para o balanço de massas, são as massas de cada composto químico em cada reator. Em qualquer caso, representam o estado do sistema, ou a resposta que se tenta determinar.

O vetor $\{B\}$ do lado direito contém aqueles elementos do balanço que são independentes do comportamento do sistema — isto é, que são constantes. Assim, muitas vezes eles representam as forças externas ou os estímulos que conduzem o sistema.

Finalmente, a matriz $[A]$ dos coeficientes normalmente contém os parâmetros que expressam como as partes do sistema interagem ou são acopladas. Conseqüentemente, a Equação (10.23) pode ser expressa como

$$[\text{Interações}]\{\text{respostas}\} = \{\text{estímulos}\}$$

Logo, a Equação (10.23) pode ser vista como uma expressão do modelo matemático fundamental formulado previamente como uma única equação no Capítulo 1 [lembre-se da Equação (1.1)]. É possível agora ver que a Equação (10.23) representa uma versão destinada a sistemas acoplados envolvendo várias variáveis dependentes $\{X\}$.

Como se sabe deste e dos capítulos anteriores, há diversas formas de resolver a Equação (10.23). No entanto, usar a matriz inversa leva a um resultado particularmente interessante. A solução formal pode ser escrita como

$$\{X\} = [A]^{-1}\{B\}$$

ou (lembrando a definição da multiplicação de matrizes do Quadro PT3.2)

$$x_1 = a_{11}^{-1}b_1 + a_{12}^{-1}b_2 + a_{13}^{-1}b_3$$
$$x_2 = a_{21}^{-1}b_1 + a_{22}^{-1}b_2 + a_{23}^{-1}b_3$$
$$x_3 = a_{31}^{-1}b_1 + a_{32}^{-1}b_2 + a_{33}^{-1}b_3$$

Portanto, vê-se que a própria matriz inversa, além de fornecer uma solução, tem propriedades extremamente úteis. Isto é, cada um dos seus elementos representa a resposta de uma parte do sistema a um estímulo unitário de qualquer outra parte do sistema.

Observe que essas formulações são lineares e, portanto, valem a superposição e a proporcionalidade. A *superposição* significa que, se um sistema está sujeito a diversos estímulos diferentes (os b's), as respostas podem ser calculadas individualmente e os resultados podem ser somados para se obter uma resposta total. A *proporcionalidade* significa que a multiplicação do estímulo por uma quantidade faz com que a resposta a esse estímulo fique multiplicada pela mesma quantidade. Logo, o coeficiente a_{11}^{-1} é uma constante de proporcionalidade que dá o valor de x_1 que é causado pelo nível unitário de b_1. Esse resultado é independente dos efeitos de b_2 e b_3 em x_1, os quais estão refletidos nos coeficientes a_{12}^{-1} e a_{13}^{-1}, respectivamente. Portanto, chega-se à conclusão geral de que o elemento a_{ij}^{-1} da matriz inversa representa o valor de x_i causado por uma quantidade unitária de b_j. Usando o exemplo da estrutura, o elemento a_{ij}^{-1} da matriz inversa representaria a força no membro i decorrente de uma força externa unitária no nó j. Mesmo para sistemas pequenos, tal comportamento de interações resposta-estímulo individuais não seria intuitivamente óbvio. Assim, a matriz inversa fornece uma técnica poderosa para entender as inter-relações de partes componentes de sistemas complicados. Esse poder será ilustrado nas Seções 12.1 e 12.2.

10.3 ANÁLISE DE ERRO E CONDICIONAMENTO DO SISTEMA

Além das aplicações na engenharia, a inversa também fornece formas de discernir se um sistema é mal condicionado. Três métodos estão disponíveis para esse propósito:

1. Faça uma mudança de escala na matriz de coeficientes $[A]$ de modo que o maior elemento em cada linha seja 1. Inverta a matriz na nova escala e, se existirem elementos de $[A]^{-1}$ que sejam várias ordens de grandeza maiores que 1, é provável que o sistema seja mal condicionado (ver Quadro 10.1).

Quadro 10.1 Interpretando os Elementos da Matriz Inversa como uma Medida de Mal Condicionamento

Um método de determinar o condicionamento do sistema é mudar a escala de $[A]$, de modo que o maior elemento em cada linha seja 1, e então calcular $[A]^{-1}$. Se existirem elementos de $[A]^{-1}$ que sejam várias ordens de grandeza maiores que 1, é provável que o sistema seja mal condicionado.

Pode-se ter uma visão melhor dessa abordagem lembrando que uma forma de verificar se uma solução aproximada $\{X\}$ é aceitável é substituí-la na equação original e ver se o resultado é o vetor constante do lado direito. Isto é equivalente a

$$\{R\} = \{B\} - [A]\{\tilde{X}\} \tag{Q10.1.1}$$

onde $\{R\}$ é o resíduo entre as constantes do lado direito e os valores calculados com a solução $\{\tilde{X}\}$. Se $\{R\}$ for pequeno, conclui-se que os valores de $\{\tilde{X}\}$ são adequados. No entanto, suponha que $\{X\}$ seja a solução exata, que leva a um resíduo 0, como em

$$\{0\} = \{B\} - [A]\{X\} \tag{Q10.1.2}$$

Subtraindo a Equação (Q10.1.2) da Equação (Q10.1.1), tem-se

$$\{R\} = [A]\left\{\{X\} - \{\tilde{X}\}\right\}$$

Multiplicando ambos os lados dessa equação por $[A]^{-1}$, obtém-se

$$\{X\} - \{\tilde{X}\} = [A]^{-1}\{R\}$$

Esse resultado indica por que verificar uma solução por substituição pode ser enganoso. Para os casos em que elementos de $[A]^{-1}$ são grandes, uma pequena discrepância nos resíduos $\{R\}$ do lado direito poderia corresponder a um grande erro $\{X\} - \{\tilde{X}\}$ nos valores calculados das variáveis. Em outras palavras, um resíduo pequeno não garante uma solução acurada. No entanto, pode-se concluir que, se o maior elemento de $[A]^{-1}$ for da mesma ordem de grandeza da unidade, o sistema pode ser considerado bem condicionado. Reciprocamente, se $[A]^{-1}$ incluir elementos muito maiores do que a unidade, conclui-se que o sistema é mal condicionado.

2. Multiplique a inversa pela matriz de coeficientes original e verifique se o resultado está próximo da matriz identidade. Se não, isso indica mal condicionamento.
3. Inverta a matriz invertida e verifique se o resultado está suficientemente próximo da matriz de coeficientes original. Caso contrário, novamente isso pode indicar que o sistema é mal condicionado.

Apesar de esses métodos poderem indicar mal condicionamento, seria preferível obter um único número (como o número de condicionamento da Seção 4.2.3) que servisse como um indicador do problema. As tentativas de formular tal número de condicionamento da matriz são baseadas no conceito matemático de norma.

10.3.1 Normas de Vetores e de Matrizes

Uma *norma* é uma função a valores reais que fornece uma medida do tamanho ou "comprimento" de entidades matemáticas com muitas componentes, como vetores e matrizes (ver Quadro 10.2).

Um exemplo simples é um vetor no espaço euclidiano tridimensional (Figura 10.6), que pode ser representado como

$$\lfloor F \rfloor = \lfloor a \quad b \quad c \rfloor$$

onde a, b e c são as distâncias ao longo dos eixos x, y e z, respectivamente. O comprimento desse vetor — isto é, a distância das coordenadas (0, 0, 0) até (a, b, c) — pode ser simplesmente calculado como

$$\|F\|_e = \sqrt{a^2 + b^2 + c^2}$$

onde a nomenclatura $\|F\|_e$ indica que esse comprimento se chama norma euclidiana de $[F]$.

De forma semelhante, para um vetor de dimensão n, $\lfloor X \rfloor = \lfloor x_1 \;\; x_2 \;\; \cdots \;\; x_n \rfloor$, uma norma euclidiana pode ser calculada como

$$\|X\|_e = \sqrt{\sum_{i=1}^{n} x_i^2}$$

O conceito pode ser estendido para uma matriz $[A]$, como em

$$\|A\|_e = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}^2} \tag{10.24}$$

Quadro 10.2 Normas de Matrizes

Como explicado nesta Seção, as normas euclidianas podem ser empregadas para quantificar o tamanho de um vetor,

$$\|X\|_e = \sqrt{\sum_{i=1}^{n} x_i^2}$$

ou de uma matriz,

$$\|A\|_e = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} a_{i,j}^2}$$

Para vetores, há alternativas chamadas de p normas que podem ser representadas genericamente por

$$\|X\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

Também se percebe que a norma euclidiana e a 2 norma, $\|X\|_2$, são idênticas para os vetores.

Outros exemplos importantes são

$$\|X\|_1 = \sum_{i=1}^{n} |x_i|$$

que representa a norma como a soma dos valores absolutos dos elementos. Outra é o *módulo máximo* ou *norma uniforme do vetor*,

$$\|X\|_\infty = \max_{1\le i\le n} |x_i| \text{ t}$$

que define a norma como o elemento com o maior valor absoluto.

Usando uma abordagem semelhante, pode-se desenvolver normas para as matrizes. Por exemplo,

$$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}|$$

Ou seja, a soma dos valores absolutos dos coeficientes é calculada para cada coluna, e a maior dessas somas é tomada como a norma. Isso é chamado de *norma da soma das colunas*.

Uma construção semelhante pode ser feita para as linhas, resultando em uma *norma da soma das linhas* ou *norma uniforme da matriz*,

$$\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|$$

Deve-se observar que, em contraste com os vetores, a 2 norma e a norma euclidiana de uma matriz não são iguais. Enquanto a norma euclidiana $\|A\|_e$ pode ser facilmente determinada pela Equação (10.24), a 2 norma de uma matriz $\|A\|_2$é calculada como

$$\|A\|_2 = (\mu_{\text{máx}})^{1/2}$$

onde $\mu_{\text{máx}}$ é o maior autovalor de $[A]^T[A]$. No Capítulo 27, iremos aprender mais sobre autovalores. No momento, o ponto importante é que $\|A\|_2$ ou *norma espectral* é a norma mínima e, portanto, fornece a medida mais justa do tamanho (Ortega, 1972).

o que recebe um nome especial — a *norma de Frobenius*. Entretanto, como as outras normas de vetores, ela fornece um único valor para quantificar o "tamanho" de $[A]$.

Deve-se observar que há alternativas às normas euclidianas e de Frobenius (ver Quadro 10.2). Por exemplo, a *norma uniforme do vetor* é definida por

$$\|X\|_\infty = \max_{1\le i\le n} |x_i|$$

Isto é, o elemento com maior valor absoluto é tomado como a medida do tamanho do vetor. De maneira similar, uma *norma uniforme da matriz* ou *norma da soma das linhas* é definida como

$$\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}| \tag{10.25}$$

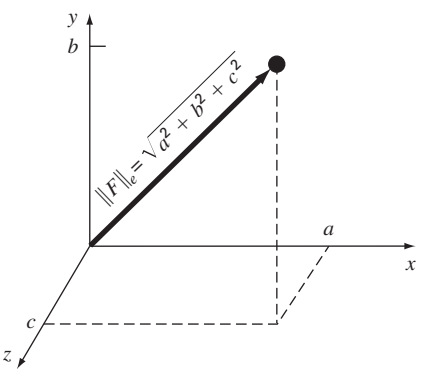


FIGURA 10.6
Representação gráfica de um vetor $[F] = [a \quad b \quad c]$ no espaço euclidiano.

Nesse caso, a soma dos valores absolutos dos elementos é calculada para cada linha, e a maior destas é tomada como norma.

Apesar de existirem vantagens teóricas no uso de cada uma dessas normas, a escolha às vezes é influenciada por considerações práticas. Por exemplo, a norma uniforme de linha é muito usada pela facilidade com que pode ser calculada, e pelo fato de que, em geral, fornece uma medida adequada do tamanho da matriz.

10.3.2 Número de Condicionamento da Matriz

Agora que já foi apresentado, o conceito de norma pode ser usado para definir

$$\text{Cond}\,[A] = \|A\| \cdot \|A^{-1}\| \tag{10.26}$$

onde Cond [A] é chamado *número de condicionamento da matriz*. Observe que, para uma matriz [A], esse número será superior ou igual a 1. Pode-se mostrar (Ralston e Rabinowitz, 1978; Gerald e Wheatley, 1989) que

$$\frac{\|\Delta X\|}{\|X\|} \leq \text{Cond}\,[A]\frac{\|\Delta A\|}{\|A\|}$$

Isto é, o erro relativo da norma de uma solução calculada pode ser tão grande quanto o erro relativo da norma dos coeficientes de [A] multiplicado pelo número de condicionamento. Por exemplo, se os coeficientes da matriz [A] são conhecidos com uma precisão de t dígitos (ou seja, erros de arredondamento são da ordem de 10^{-t}), e Cond $[A] = 10^c$, a solução [X] pode ser válida para apenas $t - c$ dígitos (erros de arredondamento $\sim 10^{c-t}$).

EXEMPLO 10.4 Determinação do Condicionamento de uma Matriz

Enunciado do Problema. A matriz de Hilbert, que é notoriamente mal condicionada, pode ser genericamente representada por

$$\begin{bmatrix} 1 & 1/2 & 1/3 & \cdots & 1/n \\ 1/2 & 1/3 & 1/4 & \cdots & 1/(n+1) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1/n & 1/(n+1) & 1/(n+2) & \cdots & 1/(2n-1) \end{bmatrix}$$

Use a norma da soma das linhas para fazer uma estimativa do número de condicionamento da matriz de Hilbert 3×3,

$$[A] = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$$

Solução. Primeiro, pode-se mudar a escala da matriz de modo que o máximo elemento em cada linha seja 1,

$$[A] = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1 & 2/3 & 1/2 \\ 1 & 3/4 & 3/5 \end{bmatrix}$$

Somando cada uma dessas linhas, temos 1,833, 2,1667 e 2,35. Assim, a terceira linha tem a maior soma, e a norma da soma das linhas é

$$\|A\|_\infty = 1 + \frac{3}{4} + \frac{3}{5} = 2{,}35$$

A inversa da matriz com mudança de escala pode ser calculada por

$$[A]^{-1} = \begin{bmatrix} 9 & -18 & 10 \\ -36 & 96 & -60 \\ 30 & -90 & 60 \end{bmatrix}$$

Observe que os elementos dessa matriz são maiores do que os elementos da matriz original. Isso também se reflete na sua norma da soma das linhas, que é calculada como

$$\left\|A^{-1}\right\|_\infty = |-36| + |96| + |-60| = 192$$

Logo, o número de condicionamento pode ser calculado por

$$\text{Cond}\,[A] = 2{,}35(192) = 451{,}2$$

O fato de o número de condicionamento ser consideravelmente maior do que a unidade sugere que o sistema é mal condicionado. A extensão do mal condicionamento pode ser quantificada calculando-se $c = \log 451{,}2 = 2{,}65$. Os computadores usando representação em ponto flutuante IEEE têm aproximadamente $t = \log 2^{-24} = 7{,}2$ algarismos significativos na base 10 (lembre-se da Seção 3.4.1). Assim, a solução pode exibir erros de arredondamento de até $10^{(2{,}65\text{-}7{,}2)} = 3 \times 10^{-5}$. Observe que tais estimativas quase sempre superestimam o erro real. No entanto, elas são úteis ao alertarem para a possibilidade de que os erros de arredondamento possam ser significativos.

Do ponto de vista prático, o problema na implementação da Equação (10.26) é o preço computacional exigido para o cálculo de $\|A^{-1}\|$. Rice (1983) esboça algumas estratégias possíveis para aliviar o problema. Mais do que isso, ele sugere uma forma alternativa de avaliar o condicionamento do sistema: rodar a mesma solução em dois compiladores diferentes. Como os códigos resultantes irão provavelmente implementar a aritmética de modo diferente, o efeito do mal condicionamento deve ser evidente a partir de tal experiência. Finalmente, deve-se mencionar que pacotes de software e bibliotecas como MATLAB e Mathcad têm recursos para calcular convenientemente o condicionamento de uma matriz. Tais recursos serão abordados de novo quando se revisar esses pacotes no fim do Capítulo 11.

10.3.3 Refinamento Iterativo

Em alguns casos, os erros de arredondamento podem ser reduzidos pelo procedimento a seguir. Suponha que alguém esteja resolvendo o seguinte conjunto de equações:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned} \tag{10.27}$$

Por concisão, a discussão será limitada a esse pequeno sistema (3×3). No entanto, a abordagem é aplicável, em geral, para conjuntos maiores de equações lineares.

Suponha que um vetor solução aproximado seja dado por $\{\tilde{X}\}^T = \lfloor \tilde{x}_1 \;\; \tilde{x}_2 \;\; \tilde{x}_3 \rfloor$. Essa solução pode ser substituída na Equação (10.27) para dar

$$\begin{aligned} a_{11}\tilde{x}_1 + a_{12}\tilde{x}_2 + a_{13}\tilde{x}_3 &= \tilde{b}_1 \\ a_{21}\tilde{x}_1 + a_{22}\tilde{x}_2 + a_{23}\tilde{x}_3 &= \tilde{b}_2 \\ a_{31}\tilde{x}_1 + a_{32}\tilde{x}_2 + a_{33}\tilde{x}_3 &= \tilde{b}_3 \end{aligned} \tag{10.28}$$

Agora, suponha que a solução exata $\{X\}$ se expresse como uma função da solução aproximada e de um vetor de fatores de correção $\{\Delta X\}$, onde

$$x_1 = \tilde{x}_1 + \Delta x_1$$

$$x_2 = \tilde{x}_2 + \Delta x_2 \qquad (10.29)$$
$$x_3 = \tilde{x}_3 + \Delta x_3$$

Se esses resultados forem substituídos na Equação (10.27), será obtido o seguinte sistema:

$$\begin{aligned} a_{11}(\tilde{x}_1 + \Delta x_1) + a_{12}(\tilde{x}_2 + \Delta x_2) + a_{13}(\tilde{x}_3 + \Delta x_3) &= b_1 \\ a_{21}(\tilde{x}_1 + \Delta x_1) + a_{22}(\tilde{x}_2 + \Delta x_2) + a_{23}(\tilde{x}_3 + \Delta x_3) &= b_2 \\ a_{31}(\tilde{x}_1 + \Delta x_1) + a_{32}(\tilde{x}_2 + \Delta x_2) + a_{33}(\tilde{x}_3 + \Delta x_3) &= b_3 \end{aligned} \qquad (10.30)$$

Agora, a Equação (10.28) pode ser subtraída da Equação (10.30) para fornecer

$$\begin{aligned} a_{11}\Delta x_1 + a_{12}\Delta x_2 + a_{13}\Delta x_3 &= b_1 - \tilde{b}_1 = E_1 \\ a_{21}\Delta x_1 + a_{22}\Delta x_2 + a_{23}\Delta x_3 &= b_2 - \tilde{b}_2 = E_2 \\ a_{31}\Delta x_1 + a_{32}\Delta x_2 + a_{33}\Delta x_3 &= b_3 - \tilde{b}_3 = E_3 \end{aligned} \qquad (10.31)$$

Esse sistema é ele próprio um conjunto de equações lineares simultâneas que pode ser resolvido para se obter os fatores de correção. Esses fatores podem ser usados para melhorar a solução, como especificado na Equação (10.29).

É relativamente fácil integrar um procedimento de refinamento iterativo em programas de computador para métodos de eliminação. O procedimento é especialmente efetivo nas abordagens de decomposição *LU* descritas anteriormente, que são planejadas para calcular vetores diferentes do lado direito de forma eficiente. Observe que, para serem efetivos na correção de sistemas mal condicionados, os *E*'s na Equação (10.31) devem ser expressos em dupla precisão.

PROBLEMAS

10.1 Use as regras de multiplicação de matrizes para mostrar que as Equações (10.7) e (10.8) são conseqüências da Equação (10.6).

10.2 **(a)** Use a eliminação de Gauss ingênua para decompor o seguinte sistema, de acordo com a descrição dada na Seção 10.2.

$$\begin{aligned} 10x_1 + 2x_2 - x_3 &= 27 \\ -3x_1 - 6x_2 + 2x_3 &= -61{,}5 \\ x_1 + x_2 + 5x_3 &= -21{,}5 \end{aligned}$$

A seguir, multiplique as matrizes resultantes [*L*] e [*U*] para verificar que se obtém [*A*]. (**b**) Use a decomposição *LU* para resolver o sistema. Mostre cada passo dos cálculos. (**c**) Resolva o sistema também para um outro vetor do lado direito: $\{B\}^T = \lfloor 12\ 18\ -6 \rfloor$.

10.3

(a) Resolva o seguinte sistema de equações por decomposição *LU* sem pivotamento

$$\begin{aligned} 8x_1 + 4x_2 - x_3 &= 11 \\ -2x_1 + 5x_2 + x_3 &= 4 \\ 2x_1 - x_2 + 6x_3 &= 7 \end{aligned}$$

(b) Determine a matriz inversa. Verifique seus resultados comprovando que $[A][A]^{-1} = [I]$.

10.4 Resolva o seguinte sistema de equações por decomposição *LU* com pivotamento parcial:

$$\begin{aligned} 2x_1 - 6x_2 - x_3 &= -38 \\ -3x_1 - x_2 + 7x_3 &= -34 \\ -8x_1 + x_2 - 2x_3 &= -20 \end{aligned}$$

10.5 Determine o número total de flops como função do número de equações *n* para as fases de (**a**) decomposição, (**b**) substituição progressiva e (**c**) substituição regressiva da versão de decomposição *LU* da eliminação de Gauss.

10.6 Use a decomposição *LU* para determinar a matriz inversa para o seguinte sistema. Não use pivotamento e verifique seus resultados comprovando que $[A][A]^{-1} = [I]$.

$$\begin{aligned} 10x_1 + 2x_2 - x_3 &= 27 \\ -3x_1 - 6x_2 + 2x_3 &= -61{,}5 \\ x_1 + x_2 + 5x_3 &= -21{,}5 \end{aligned}$$

10.7 Efetue a decomposição de Crout em

$$\begin{aligned} 2x_1 - 6x_2 + x_3 &= 12 \\ -x_1 + 7x_2 - x_3 &= -8 \\ x_1 - 3x_2 + 2x_3 &= 16 \end{aligned}$$

Então, multiplique as matrizes resultantes [*L*] e [*U*] para verificar que se obtém [*A*].

10.8 O seguinte sistema de equações é projetado para determinar as concentrações (os *c*'s em g/m^3) em uma série de reatores acoplados como função da quantidade de entrada de massa em cada reator (o lado direito está em g/dia),

$$\begin{aligned} 15c_1 - 3c_2 - c_3 &= 3800 \\ -3c_1 + 18c_2 - 6c_3 &= 1200 \\ -4c_1 - c_2 + 12c_3 &= 2350 \end{aligned}$$

(a) Determine a matriz inversa.
(b) Use a inversa para determinar a solução.
(c) Determine de quanto o fluxo de entrada de massa no reator 3 deve ser aumentado para induzir um aumento de 10g/m^3 na concentração do reator 1.
(d) De quanto a concentração no reator 3 será reduzida se o fluxo de entrada de massa nos reatores 1 e 2 for reduzido para 500 e 250 g/dia, respectivamente?

10.9 Determine $\|A\|_e$, $\|A\|_1$ e $\|A\|_\infty$ para

$$[A] = \begin{bmatrix} 8 & 2 & -10 \\ -9 & 1 & 3 \\ 15 & -1 & 6 \end{bmatrix}$$

Mude a escala da matriz, fazendo o maior elemento de cada linha igual a um.

10.10 Determine as normas euclidiana e da soma das linhas para os sistemas dos Problemas 10.3 e 10.4. Mude a escala das matrizes fazendo o maior elemento de cada linha igual a um.

10.11 Uma matriz [*A*] é definida por

$$[A] = \begin{bmatrix} 0{,}125 & 0{,}25 & 0{,}5 & 1 \\ 0{,}015625 & 0{,}625 & 0{,}25 & 1 \\ 0{,}00463 & 0{,}02777 & 0{,}16667 & 1 \\ 0{,}001953 & 0{,}015625 & 0{,}125 & 1 \end{bmatrix}$$

Usando a norma da soma das linhas, calcule o número de condicionamento e quantos algarismos suspeitos serão gerados por essa matriz.

10.12 **(a)** Determine o número de condicionamento do seguinte sistema, usando a norma da soma das linhas. Não mude a escala do sistema.

$$\begin{bmatrix} 1 & 4 & 9 & 16 & 25 \\ 4 & 9 & 16 & 25 & 36 \\ 9 & 16 & 25 & 36 & 49 \\ 16 & 25 & 36 & 49 & 64 \\ 25 & 36 & 49 & 64 & 81 \end{bmatrix}$$

Quantos algarismos de precisão serão perdidos por causa do mal condicionamento? (**b**) Repita (**a**), mas mudando a escala da matriz fazendo o maior elemento de cada linha igual a um.

10.13 Determine o número de condicionamento da matriz com base na norma da soma das linhas para a matriz de Hilbert 5×5, com mudança de escala. Quantos algarismos de precisão serão perdidos por causa do mal condicionamento?

10.14 Além da matriz de Hilbert, há outras matrizes que são inerentemente mal condicionadas. Uma delas é a *matriz de Vandermonde*, que tem a forma:

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{bmatrix}$$

(a) Determine o número de condicionamento da matriz com base na norma da soma das linhas para o caso em que $x_1 = 4$, $x_2 = 2$ e $x_3 = 7$.

(b) Use o software MATLAB para calcular os números de condicionamento espectral e de Frobenius.

10.15 Desenvolva um programa amigável ao usuário para a decomposição *LU* com base no pseudocódigo da Figura 10.2.

10.16 Desenvolva um programa amigável ao usuário para a decomposição *LU*, incluindo o recurso de calcular a matriz inversa. Use como base as Figuras 10.2. e 10.5.

10.17 Use técnicas de refinamento iterativo para melhorar $x_1 = 2$, $x_2 = -3$ e $x_3 = 8$, que são soluções aproximadas de

$$\begin{aligned} 2x_1 + 5x_2 + x_3 &= -5 \\ 6x_1 + 2x_2 + x_3 &= 12 \\ x_1 + 2x_2 + x_3 &= 3 \end{aligned}$$

10.18 Considere os vetores

$$\begin{aligned} \vec{A} &= 2\vec{i} - 3\vec{j} + a\vec{k} \\ \vec{B} &= b\vec{i} + \vec{j} - 4\vec{k} \\ \vec{C} &= 3\vec{i} + c\vec{j} + 2\vec{k} \end{aligned}$$

O vetor $\vec{A}$ é perpendicular a $\vec{B}$ e a $\vec{C}$. Sabe-se também que $\vec{B} \cdot \vec{C} = 2$. Use qualquer método estudado neste capítulo para determinar as incógnitas *a*, *b* e *c*.

10.19 Considere os seguintes vetores:

$$\begin{aligned} \vec{A} &= a\vec{i} + b\vec{j} + c\vec{k} \\ \vec{B} &= -2\vec{i} + \vec{j} - 4\vec{k} \\ \vec{C} &= \vec{i} + 3\vec{j} + 2\vec{k} \end{aligned}$$

onde $\vec{A}$ é um vetor desconhecido. Se

$$(\vec{A} \times \vec{B}) + (\vec{A} \times \vec{C}) = (5a + b)\vec{i} + (3b - 2)c\vec{j} + (-4c + 1)\vec{k}$$

use qualquer método estudado neste capítulo para determinar as incógnitas *a*, *b*, e *c*.

10.20 Considere a função definida no intervalo [0,2] por:

$$f(x) = \begin{cases} ax + b, & 0 \le x \le 1 \\ cx + d, & 1 \le x \le 2 \end{cases}$$

Determine as constantes *a*, *b* *c* e *d* de modo que a função *f* satisfaça:

- $f(0) = f(2) = 1$.
- *f* é contínua em todo o intervalo.
- $a + b = 4$.

Deduza e resolva um sistema de equações algébricas lineares com uma forma matricial idêntica à da Equação (10.1).

10.21

(a) Crie uma matriz de Hilbert 3×3. Ela será sua matriz [*A*]. Multiplique a matriz pelo vetor coluna $\{x\} = [1, 1, 1]^T$. A solução de $[A]\{x\}$ será outro vetor coluna $\{b\}$. Usando qualquer pacote numérico e a eliminação de Gauss, determine a solução de $[A]\{x\} = \{b\}$ usando a matriz de Hilbert e o $\{b\}$ que você calculou. Compare o resultado com o seu vetor conhecido $\{x\}$. Use precisão suficiente na exibição de seus resultados para detectar imprecisões.

(b) Repita a parte (**a**) usando uma matriz de Hilbert 7×7.

(c) Repita a parte (**a**) usando uma matriz de Hilbert 10×10.

CAPÍTULO 11

Matrizes Especiais e Gauss-Seidel

Certas matrizes têm uma estrutura particular que pode ser explorada para desenvolver esquemas eficientes de solução. A primeira parte deste capítulo é dedicada a dois desses tipos de sistemas: as matrizes *de banda* e as *simétricas*. Serão descritos métodos de eliminação eficientes para ambas.

A segunda parte do capítulo se volta para uma alternativa aos métodos de eliminação, isto é, os métodos iterativos de aproximação. O foco será no *método de Gauss-Seidel*, que emprega aproximações iniciais e itera o processo para obter estimativas refinadas da solução. O método de Gauss-Seidel é particularmente bem adaptado para um número grande de equações. Nesses casos, os métodos de eliminação podem estar sujeitos a erros de arredondamento. Como o erro no método de Gauss-Seidel é controlado pelo número de iterações, os erros de arredondamento não são um motivo de preocupação nesse método. Entretanto, há certos casos para os quais a técnica de Gauss-Seidel não convergirá para a resposta correta. Esses e outros prós e contras entre a eliminação e os métodos iterativos serão discutidos nas próximas páginas.

11.1 MATRIZES ESPECIAIS

Como mencionado no Quadro PT3.1, uma *matriz de banda* é uma matriz quadrada que tem todos os elementos iguais a zero, com exceção de uma faixa centrada na diagonal principal. Os sistemas de banda são encontrados freqüentemente na prática da engenharia e na prática científica. Por exemplo, eles ocorrem tipicamente na solução de equações diferenciais. Além disso, outros métodos numéricos, como os *splines* cúbicos (Seção 18.5), envolvem a solução de sistemas de banda.

As dimensões de um sistema de banda podem ser quantificadas por dois parâmetros: a largura da banda LB e metade da largura da banda MLB (Figura 11.1) Esses dois valores são relacionados por $\text{LB} = 2\text{MLB} + 1$. Assim, em geral, um sistema de banda é um sistema para o qual $a_{ij} = 0$ se $|i - j| > \text{MLB}$.

Embora a eliminação de Gauss ou a decomposição *LU* convencional possam ser empregadas para resolver sistemas de banda, elas são ineficientes, pois, se o pivotamento não for necessário, nenhum dos elementos fora da banda mudará de seu valor original nulo. Então, espaço e tempo desnecessários seriam desperdiçados no armazenamento e na manipulação desses zeros inúteis. Se for previamente conhecido que o pivotamento não é necessário, podem ser desenvolvidos algoritmos muito eficientes que não envolvem os elementos nulos fora da banda. Já que muitos problemas envolvendo sistemas de banda não requerem pivotamento, tais algoritmos alternativos, como os descritos a seguir, são as melhores escolhas.

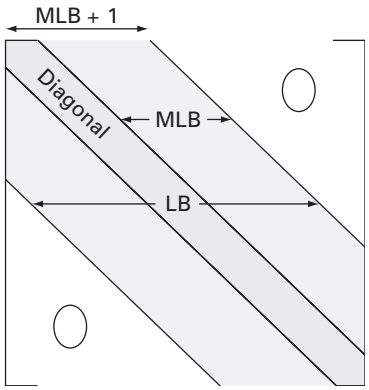


FIGURA 11.1
Parâmetros usados para quantificar as dimensões de um sistema de banda. LB e MLB designam a largura da faixa e a metade dessa largura, respectivamente.

11.1.1 Sistemas Tridiagonais

Um sistema tridiagonal — isto é, um que tenha a largura de faixa 3 — pode ser expresso no caso geral como

$$\begin{bmatrix} f_1 & g_1 & & & & & & \\ e_2 & f_2 & g_2 & & & & & \\ & e_3 & f_3 & g_3 & & & & \\ & & \cdot & \cdot & \cdot & & & \\ & & & \cdot & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \cdot & \\ & & & & & e_{n-1} & f_{n-1} & g_{n-1} \\ & & & & & & e_n & f_n \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{Bmatrix} = \begin{Bmatrix} r_1 \\ r_2 \\ r_3 \\ \cdot \\ \cdot \\ \cdot \\ r_{n-1} \\ r_n \end{Bmatrix} \tag{11.1}$$

Observe que trocamos nossa anotação dos coeficientes dos *a*'s e *b*'s para *e*'s, *f*'s, *g*'s e *r*'s. Isso foi feito para evitar o armazenamento de grandes quantidades de zeros inúteis na matriz quadrada dos *a's*. Essa modificação, que tem o objetivo de economizar espaço, é vantajosa, pois o algoritmo resultante requer menos memória de computador.

A Figura 11.2 mostra um pseudocódigo para um método eficiente, chamado *algoritmo de Thomas*, para resolver a Equação (11.1). Da mesma forma como na decomposição *LU* convencional, o algoritmo consiste em três passos: decomposição e substituições progressiva e regressiva. Assim, todas as vantagens da decomposição *LU*, como o cálculo conveniente para diversos vetores do lado direito e o cálculo da matriz inversa, podem ser obtidas pela aplicação adequada desse algoritmo.

(*a*) Decomposição

```
DOFOR k = 2, n
  e_k = e_k/f_k–1
  f_k = f_k – e_k · g_k–1
END DO
```

(*b*) Substituição progressiva

```
DOFOR k = 2, n
  r_k = r_k – e_k · r_k–1
END DO
```

(*c*) Substituição regressiva

```
x_n = r_n/f_n
DOFOR k = n – 1, 1, –1
  x_k = (r_k – g_k · x_k+1)/f_k
END DO
```

FIGURA 11.2
Pseudocódigo para implementar o algoritmo de Thomas, um método de decomposição *LU* para sistemas tridiagonais.

EXEMPLO 11.1

Solução Tridiagonal com o Algoritmo de Thomas

Enunciado do Problema. Resolva o seguinte sistema tridiagonal com o algoritmo de Thomas.

$$\begin{bmatrix} 2{,}04 & -1 & & \\ -1 & 2{,}04 & -1 & \\ & -1 & 2{,}04 & -1 \\ & & -1 & 2{,}04 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} 40{,}8 \\ 0{,}8 \\ 0{,}8 \\ 200{,}8 \end{Bmatrix}$$

Solução. Primeiro, a decomposição é implementada por

$$e_2 = -1/2{,}04 = -0{,}49$$
$$f_2 = 2{,}04 - (-0{,}49)(-1) = 1{,}550$$
$$e_3 = -1/1{,}550 = -0{,}645$$
$$f_3 = 2{,}04 - (-0{,}645)(-1) = 1{,}395$$
$$e_4 = -1/1{,}395 = -0{,}717$$
$$f_4 = 2{,}04 - (-0{,}717)(-1) = 1{,}323$$

Assim, a matriz foi transformada em

$$\begin{bmatrix} 2{,}04 & -1 & & \\ -0{,}49 & 1{,}550 & -1 & \\ & -0{,}645 & 1{,}395 & -1 \\ & & -0{,}717 & 1{,}323 \end{bmatrix}$$

e a decomposição LU é

$$[A] = [L][U] = \begin{bmatrix} 1 & & & \\ -0{,}49 & 1 & & \\ & -0{,}645 & 1 & \\ & & -0{,}717 & 1 \end{bmatrix} \begin{bmatrix} 2{,}04 & -1 & & \\ & 1{,}550 & -1 & \\ & & 1{,}395 & -1 \\ & & & 1{,}323 \end{bmatrix}$$

Pode-se verificar que isso está correto multiplicando $[L][U]$ para obter $[A]$.

A substituição progressiva é implementada por

$$r_2 = 0{,}8 - (-0{,}49)40{,}8 = 20{,}8$$
$$r_3 = 0{,}8 - (-0{,}645)20{,}8 = 14{,}221$$
$$r_4 = 200{,}8 - (-0{,}717)14{,}221 = 210{,}996$$

Portanto, o vetor do lado direito foi modificado para

$$\begin{Bmatrix} 40{,}8 \\ 20{,}8 \\ 14{,}221 \\ 210{,}996 \end{Bmatrix}$$

o qual pode então ser usado em combinação com a matriz $[U]$ para se efetuar a substituição regressiva e obter a solução

$$T_4 = 210{,}996/1{,}323 = 159{,}480$$
$$T_3 = [14{,}221 - (-1)159{,}48]/1{,}395 = 124.538$$
$$T_2 = [20{,}800 - (-1)124{,}538]/1{,}550 = 93{,}778$$
$$T_1 = [40{,}800 - (-1)93{,}778]/2{,}040 = 65{,}970$$

11.1.2 Decomposição de Cholesky

Do quadro PT3.1, tem-se que uma matriz simétrica é aquela na qual $a_{ij} = a_{ji}$ para todo i e j. Em outras palavras, $[A] = [A]^T$. Tais sistemas ocorrem usualmente no contexto tanto de problemas matemáticos quanto de engenharia. Eles oferecem vantagens computacionais, pois apenas metade do armazenamento é necessária e, na maioria dos casos, é preciso apenas a metade do tempo de computação para sua solução.

Uma das abordagens mais populares envolve a *decomposição de Cholesky*, algoritmo baseado no fato de que uma matriz simétrica pode ser decomposta em

$$[A] = [L][L]^T \tag{11.2}$$

Isto é, os fatores triangulares resultantes são transpostos um do outro.

Os termos da Equação (11.2) podem ser multiplicados e igualados. O resultado é expresso de forma simples por relações de recorrência. Para a k-ésima linha,

$$l_{ki} = \frac{a_{ki} - \sum_{j=1}^{i-1} l_{ij} l_{kj}}{l_{ii}} \qquad \text{para } i = 1, 2, \ldots, k-1 \tag{11.3}$$

e

$$l_{kk} = \sqrt{a_{kk} - \sum_{j=1}^{k-1} l_{kj}^2} \tag{11.4}$$

EXEMPLO 11.2 Decomposição de Cholesky

Enunciado do Problema. Aplique a decomposição de Cholesky à matriz simétrica

$$[A] = \begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix}$$

Solução. Para a primeira linha ($k = 1$), a Equação (11.3) não se aplica e a Equação (11.4) é usada para calcular

$$l_{11} = \sqrt{a_{11}} = \sqrt{6} = 2{,}4495$$

Para a segunda linha ($k = 2$), a Equação (11.3) dá

$$l_{21} = \frac{a_{21}}{l_{11}} = \frac{15}{2{,}4495} = 6{,}1237$$

e a Equação (11.4) fornece

$$l_{22} = \sqrt{a_{22} - l_{21}^2} = \sqrt{55 - (6{,}1237)^2} = 4{,}1833$$

Para a terceira linha ($k = 3$), a Equação (11.3) dá ($i = 1$)

$$l_{31} = \frac{a_{31}}{l_{11}} = \frac{55}{2{,}4495} = 22{,}454$$

e ($i = 2$)

$$l_{32} = \frac{a_{32} - l_{21} l_{31}}{l_{22}} = \frac{225 - 6{,}1237(22{,}454)}{4{,}1833} = 20{,}916$$

e a Equação (11.4) fornece

$$l_{33} = \sqrt{a_{33} - l_{31}^2 - l_{32}^2} = \sqrt{979 - (22{,}454)^2 - (20{,}916)^2} = 6{,}1106$$

Assim, a decomposição de Cholesky fornece

$$[L] = \begin{bmatrix} 2{,}4495 & & \\ 6{,}1237 & 4{,}1833 & \\ 22{,}454 & 20{,}916 & 6{,}1106 \end{bmatrix}$$

A validade dessa decomposição pode ser verificada substituindo-se a matriz L e sua transposta na Equação (11.2) para ver se seu produto fornece a matriz original $[A]$. Isso é deixado como exercício.

```
DOFOR k = 1, n
  DOFOR i = 1, k − 1
    soma = 0.
    DOFOR j = 1, i − 1
      soma = soma + a_ij · a_kj
    END DO
    a_ki = (a_ki − soma)/a_ii
  END DO
  soma = 0.
  DOFOR j = 1, k − 1
    soma = soma + a²_kj
  END DO
  a_kk = √(a_kk − soma)
END DO
```

FIGURA 11.3
Pseudocódigo para o algoritmo de decomposição *LU* de Cholesky.

A Figura 11.3 apresenta um pseudocódigo para implementar o algoritmo de decomposição de Cholesky. Deve-se observar que o algoritmo na Figura 11.3 pode resultar em um erro de execução se o cálculo de a_{kk} envolver a extração da raiz quadrada de um número negativo. Porém, nos casos em que a matriz é *positiva definida*,[1] isso nunca vai ocorrer. Como muitas das matrizes simétricas que surgem na engenharia são, de fato, positivas definidas, o algoritmo de Cholesky tem vasta aplicação. Outra vantagem de lidar com matrizes simétricas positivas definidas é que pivotamento não é necessário para evitar a divisão por zero. Assim, é possível implementar o algoritmo da Figura 11.3 sem a complicação do pivotamento.

11.2 GAUSS-SEIDEL

Os métodos iterativos ou de aproximação fornecem uma alternativa aos métodos de eliminação descritos até agora. Tais abordagens são semelhantes às técnicas desenvolvidas para obter as raízes de uma única equação do Capítulo 6. Aquelas abordagens consistiam em escolher um valor e então usar um método sistemático para obter uma estimativa refinada da raiz. Como esta parte do livro trata de um problema similar — obter valores que satisfaçam simultaneamente um conjunto de equações — pode-se suspeitar que tais métodos de aproximação são úteis neste contexto.

O *método de Gauss-Seidel* é o método iterativo mais comumente usado. Suponha que tenhamos um conjunto de n equações:

$$[A]\{X\} = \{B\}$$

Suponha que, para sermos concisos, nos limitemos a um conjunto 3×3 de equações. Se os elementos da diagonal forem todos não-nulos, é possível isolar x_1 na primeira equação, x_2 na segunda e x_3 na terceira para obter

$$x_1 = \frac{b_1 - a_{12}x_2 - a_{13}x_3}{a_{11}} \tag{11.5a}$$

$$x_2 = \frac{b_2 - a_{21}x_1 - a_{23}x_3}{a_{22}} \tag{11.5b}$$

$$x_3 = \frac{b_3 - a_{31}x_1 - a_{32}x_2}{a_{33}} \tag{11.5c}$$

Agora, começa o processo de solução escolhendo-se aproximações para os x's. Uma forma simples de obter aproximações iniciais é supor que elas são todas nulas. Esses zeros podem ser substituídos na Equação (11.5a), que pode ser usada para calcular um novo valor para $x_1 = b_1/a_{11}$. Então, substitui-se esse novo valor de x_1 junto com a aproximação anterior nula para x_3 na Equação (11.5b) para calcular um novo valor para x_2. O processo é repetido para a Equação (11.5c) para se calcular uma nova estimativa para x_3. Então, volta-se para a primeira equação e o procedimento inteiro é repetido até que a solução convirja para valores suficientemente próximos dos valores verdadeiros. A convergência pode ser verificada usando-se o critério [lembre-se da Equação (3.5)]

$$|\varepsilon_{a,i}| = \left| \frac{x_i^j - x_i^{j-1}}{x_i^j} \right| 100\% < \varepsilon_s \tag{11.6}$$

para todo i, onde j e $j - 1$ representam a iteração atual e a anterior.

[1] Uma *matriz positiva definida* é aquela para a qual o produto $\{X\}^T[A]\{X\}$ é maior do que zero para todo vetor não nulo $\{X\}$.

EXEMPLO 11.3 Método de Gauss-Seidel

Enunciado do Problema. Use o método de Gauss-Seidel para obter a solução do mesmo sistema usado no Exemplo 11.1:

$$\begin{aligned} 3x_1 - 0{,}1x_2 - 0{,}2x_3 &= 7{,}85 \\ 0{,}1x_1 + 7x_2 - 0{,}3x_3 &= -19{,}3 \\ 0{,}3x_1 - 0{,}2x_2 + 10x_3 &= 71{,}4 \end{aligned}$$

Lembre-se de que a solução verdadeira é $x_1 = 3$, $x_2 = -2{,}5$ e $x_3 = 7$.

Solução. Primeiro, em cada uma das equações, isole a variável na diagonal.

$$x_1 = \frac{7{,}85 + 0{,}1x_2 + 0{,}2x_3}{3} \tag{E11.3.1}$$

$$x_2 = \frac{-19{,}3 - 0{,}1x_1 + 0{,}3x_3}{7} \tag{E11.3.2}$$

$$x_3 = \frac{71{,}4 - 0{,}3x_1 + 0{,}2x_2}{10} \tag{E11.3.3}$$

Supondo-se que x_2 e x_3 são iguais a zero, a Equação (E11.3.1) pode ser usada para calcular

$$x_1 = \frac{7{,}85 + 0 + 0}{3} = 2{,}616667$$

Esse valor, junto com o valor suposto de $x_3 = 0$, pode ser substituído na Equação (E11.3.2) para calcular

$$x_2 = \frac{-19{,}3 - 0{,}1(2{,}616667) + 0}{7} = -2{,}794524$$

A primeira iteração é completada substituindo-se os valores calculados para x_1 e x_2 na Equação (E11.3.3) para obter

$$x_3 = \frac{71{,}4 - 0{,}3(2{,}616667) + 0{,}2(-2{,}794524)}{10} = 7{,}005610$$

Para a segunda iteração, o mesmo processo é repetido para calcular

$$x_1 = \frac{7{,}85 + 0{,}1(-2{,}794524) + 0{,}2(7{,}005610)}{3} = 2{,}990557 \qquad |\varepsilon_t| = 0{,}31\%$$

$$x_2 = \frac{-19{,}3 - 0{,}1(2{,}990557) + 0{,}3(7{,}005610)}{7} = -2{,}499625 \qquad |\varepsilon_t| = 0{,}015\%$$

$$x_3 = \frac{71{,}4 - 0{,}3(2{,}990557) + 0{,}2(-2{,}499625)}{10} = 7{,}000291 \qquad |\varepsilon_t| = 0{,}0042\%$$

O método está, portanto, convergindo para a verdadeira solução. Iterações adicionais podem ser aplicadas para melhorar as respostas. Porém, em um problema real, a resposta verdadeira não seria conhecida *a priori*. Conseqüentemente, a Equação (11.6) fornece um meio de estimar o erro. Por exemplo, para x_1,

$$|\varepsilon_{a,1}| = \left| \frac{2{,}990557 - 2{,}616667}{2{,}990557} \right| 100\% = 12{,}5\%$$

Para x_2 e x_3, as estimativas de erro são $|\varepsilon_{a,2}| = 11{,}8\%$ e $|\varepsilon_{a,3}| = 0{,}076\%$. Observe que, como no caso em que determinamos raízes de uma única equação, as formulações como a Equação (11.6) em geral fornecem uma estimativa conservadora de convergência. Assim, quando elas são satisfeitas, garantem que o resultado é conhecido pelo menos dentro da tolerância especificada por ε_s.

À medida que cada novo valor de x é calculado pelo método de Gauss-Seidel, ele é imediatamente usado na próxima equação para se determinar outro valor de x. Assim, se a solução estiver convergindo, a melhor estimativa disponível será empregada. Uma abordagem alternativa, chamada *iteração de Jacobi*, utiliza uma tática um pouco diferente. Em vez de usar os últimos x's disponíveis, essa técnica usa a Equação (11.5) para calcular um conjunto de novos x's com base no conjunto de antigos x's. Logo, conforme novos valores são gerados, não são imediatamente usados, mas guardados para a próxima iteração.

A diferença entre o método de Gauss-Seidel e a iteração de Jacobi é descrita na Figura 11.4. Embora haja certos casos para os quais o método de Jacobi é útil, a utilização das melhores estimativas disponíveis no método de Gauss-Seidel em geral o torna o preferido.

11.2.1 Critérios de Convergência para o Método de Gauss-Seidel

Observe que o método de Gauss-Seidel é similar à técnica de iteração de ponto fixo simples, que foi usada na Seção 6.1 para determinar raízes de uma única equação. Lembre-se de que a iteração de ponto fixo simples tem dois problemas fundamentais: (1) às vezes, não é convergente, e (2) quando converge, muitas vezes o faz muito lentamente. O método de Gauss-Seidel também pode exibir essas limitações.

Os critérios de convergência podem ser deduzidos lembrando, da Seção 6.5.1, que condições suficientes para a convergência de duas equações não-lineares, $u(x, y)$ e $v(x, y)$, são

$$\left|\frac{\partial u}{\partial x}\right| + \left|\frac{\partial u}{\partial y}\right| < 1 \tag{11.7a}$$

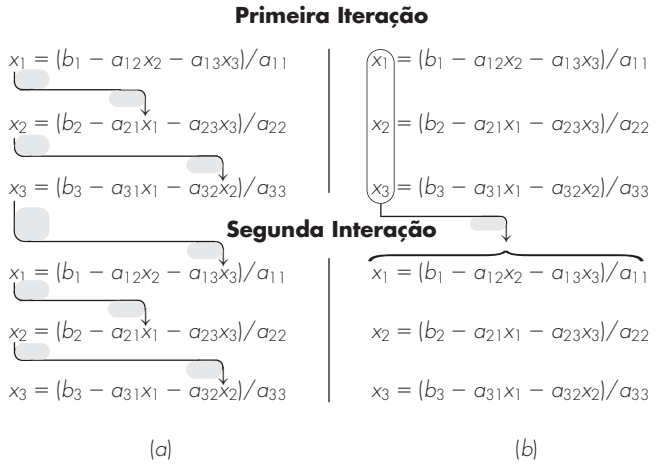
e

$$\left|\frac{\partial v}{\partial x}\right| + \left|\frac{\partial v}{\partial y}\right| < 1 \tag{11.7b}$$

Esses critérios também se aplicam a equações lineares do tipo que estamos resolvendo com o método de Gauss-Seidel. Por exemplo, no caso de duas equações simultâneas, o algoritmo de Gauss-Seidel [Equação (11.5)] pode ser expresso como

$$u(x_1, x_2) = \frac{b_1}{a_{11}} - \frac{a_{12}}{a_{11}} x_2 \tag{11.8a}$$

FIGURA 11.4
Descrição gráfica da diferença entre (*a*) o método de Gauss-Seidel e (*b*) a iteração de Jacobi para resolver equações algébricas lineares simultâneas.



e

$$v(x_1, x_2) = \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1 \tag{11.8b}$$

As derivadas parciais dessas equações podem ser calculadas com respeito a cada uma das variáveis como

$$\frac{\partial u}{\partial x_1} = 0 \qquad \frac{\partial u}{\partial x_2} = -\frac{a_{12}}{a_{11}}$$

e

$$\frac{\partial v}{\partial x_1} = -\frac{a_{21}}{a_{22}} \qquad \frac{\partial v}{\partial x_2} = 0$$

que podem ser substituídos na Equação (11.7) para fornecer

$$\left|\frac{a_{12}}{a_{11}}\right| < 1 \tag{11.9a}$$

e

$$\left|\frac{a_{21}}{a_{22}}\right| < 1 \tag{11.9b}$$

Em outras palavras, os valores absolutos das inclinações da Equação (11.8) devem ser menores do que 1 para garantir a convergência. Isso é mostrado graficamente na Figura 11.5. A Equação (11.9) pode também ser reformulada como

$$|a_{11}| > |a_{12}|$$

e

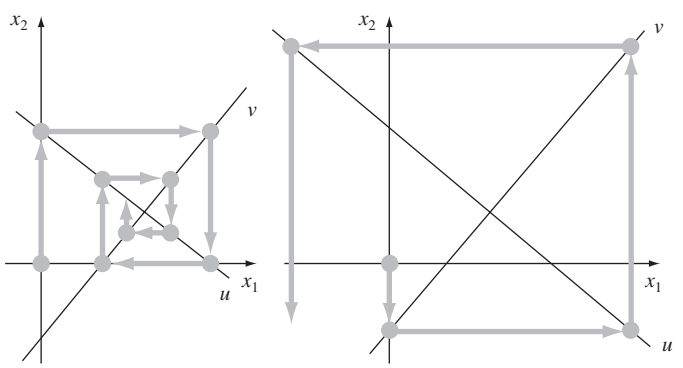
$$|a_{22}| > |a_{21}|$$

Isto é, o elemento da diagonal deve ser maior do que o elemento fora da diagonal em cada linha.

A extensão do que foi visto para n equações é imediata e pode ser expressa por

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{i,j}| \tag{11.10}$$

FIGURA 11.5
Ilustração da (*a*) convergência e (*b*) divergência do método de Gauss-Seidel. Observe que as mesmas funções são traçadas em ambos os casos (u: $11x_1 + 13x_2 = 286$; v: $11x_1 - 9x_2 = 99$). Assim, a ordem na qual as equações são implementadas (como descrito pela direção da primeira flecha saindo da origem) determina se os cálculos convergem.



(*a*) (*b*)

Isto é, o coeficiente da diagonal em cada equação deve ser maior que a soma dos valores absolutos dos outros coeficientes na equação. Esse critério é suficiente, mas não necessário para a convergência. Ou seja, embora o método possa às vezes funcionar se a Equação (11.10) não for satisfeita, a convergência é garantida se essa condição for satisfeita. Os sistemas para os quais a Equação (11.10) é satisfeita são chamados *diagonal dominantes*. Felizmente, muitos problemas de importância prática na engenharia satisfazem tal exigência.

11.2.2 Aprimoramento da Convergência Usando Relaxamento

O *relaxamento* representa uma pequena modificação no método de Gauss-Seidel e foi criado para melhorar a convergência. Após cada novo valor de x ser calculado usando-se a Equação (11.5), esse valor é modificado por uma média ponderada dos resultados da iteração anterior e da atual:

$$x_i^{\text{novo}} = \lambda x_i^{\text{novo}} + (1 - \lambda) x_i^{\text{velho}} \tag{11.11}$$

onde λ é um fator de peso para o qual deve ser escolhido um valor entre 0 e 2.

Se $\lambda = 1$, $(1 - \lambda)$ é igual a 0, e o resultado não é modificado. No entanto, se a λ for atribuído um valor entre 0 e 1, o resultado é uma média ponderada do resultado atual e do anterior. Esse tipo de modificação é chamado *sub-relaxamento* e é tipicamente usado para fazer com que um sistema não-convergente passe a convergir, ou para apressar a convergência amortecendo as oscilações.

Para valores de λ entre 1 e 2, peso extra é colocado no valor atual. Nesse caso, existe uma suposição implícita de que o novo valor está movendo-se na direção correta da solução verdadeira, mas a uma taxa muito lenta. Assim, o peso extra λ se presta a melhorar a estimativa aproximando-a do valor verdadeiro. Portanto, esse tipo de modificação, que é chamado *sobre-relaxamento*, é implementado para acelerar a convergência de um sistema já convergente. Essa abordagem é também chamada de *sobre-relaxamento simultâneo* ou *sucessivo* ou *SOR*.

A escolha de um valor adequado para λ depende criticamente do problema tratado e é, muitas vezes, determinada de maneira empírica. Para uma única solução de um conjunto de equações, tal técnica freqüentemente é desnecessária. Contudo, se o sistema que está sendo estudado deve ser resolvido repetidamente, a eficiência introduzida por uma boa escolha de λ pode ser extremamente importante. Bons exemplos são os sistemas muito grandes de equações diferenciais parciais que aparecem bastante quando se modelam variações contínuas das variáveis (lembre-se dos sistemas distribuídos esboçados na Figura PT3.1*b*). Esse tópico será retomado na Parte Oito.

11.2.3 Algoritmo para Gauss-Seidel

Um algoritmo para o método de Gauss-Seidel, com relaxamento, é descrito na Figura 11.6. Observe que não é garantido que esse algoritmo convirja se as equações não estiverem em uma forma diagonal dominante.

O pseudocódigo tem duas características que merecem menção. Primeiro, há um conjunto inicial de laços encaixados para dividir cada equação por seu elemento diagonal. Isso reduz o número total de operações no algoritmo. Segundo, observe que a verificação de erro é designada por uma variável denominada *sentinela*. Se alguma das equações tiver um erro aproximado maior que o critério de parada (e_s), então as iterações continuam. O uso da sentinela permite evitar cálculos desnecessários de estimativa de erro quando uma das equações excede o critério.

11.2.4 Contexto de Problemas para o Método de Gauss-Seidel

Além de evitar os problemas de arredondamento, a técnica de Gauss-Seidel tem várias outras vantagens que a torna particularmente atrativa no contexto de certos problemas de engenharia. Por exemplo, quando a matriz em questão é muito grande e muito esparsa (isto é, a maior parte dos elementos é zero), os métodos de eliminação desperdiçam enorme quantidade de memória de computador para armazenar zeros.

```
SUBROUTINE Gseid (a,b,n,x,imax,es,lambda)
  DOFOR i = 1,n
    dummy = a_{i,i}
    DOFOR j = 1,n
      a_{i,j} = a_{i,j}/dummy
    END DO
    b_i = b_i/dummy
  END DO
  DOFOR i = 1, n
    soma  = b_i
    DOFOR j = 1, n
      IF i≠j THEN soma  = soma − a_{i,j}*x_j
    END DO
    x_i= soma
  END DO
  iter=1
  DO
    sentinela = 1
    DOFOR i = 1,n
      old = x_i
      soma = b_i
      DOFOR j = 1,n
        IF i≠j THEN soma = soma − a_{i,j}*x_j
      END DO
      x_i = lambda* soma +(1.−lambda)*old
      IF sentinela = 1 AND x_i ≠ 0. THEN
         ea = ABS((x_i − old)/x_i)*100.
         IF ea > es THEN sentinela = 0
      END IF
    END DO
    iter = iter + 1
    IF sentinela = 1 OR (iter ≥ imax) EXIT
  END DO
END Gseid
```

FIGURA 11.6
Pseudocódigo para o método de Gauss-Seidel com relaxamento.

No começo deste capítulo, viu-se como essas limitações poderiam ser evitadas se a matriz de coeficientes fosse de banda. Para sistemas que não são de banda, não há usualmente uma maneira simples de evitar a utilização de muita memória quando se utilizam métodos de eliminação. Como todos os computadores têm uma quantidade finita de memória, essa ineficiência pode colocar uma limitação no tamanho dos sistemas para os quais os métodos de eliminação são práticos.

Embora um algoritmo geral como o da Figura 11.6 seja suscetível às mesmas limitações, a estrutura das equações de Gauss-Seidel [Equação (11.5)] permite que programas concisos sejam desenvolvidos para sistemas específicos. Como apenas os coeficientes não-nulos precisam ser incluídos na Equação (11.5), é possível grande economia de memória de computador. Mesmo que isso signifique um maior investimento inicial no desenvolvimento de software, as vantagens a longo prazo são substanciais quando se lida com sistemas grandes para os quais muitas simulações devem ser efetuadas. Tanto sistemas de variáveis agregadas quanto aqueles de distribuídas podem resultar em matrizes grandes e esparsas, para as quais o método de Gauss-Seidel é útil.

11.3 EQUAÇÕES ALGÉBRICAS LINEARES COM BIBLIOTECAS E PACOTES

As bibliotecas e os pacotes de software têm excelentes recursos para resolver sistemas de equações algébricas lineares. Antes de descrever tais ferramentas, deve-se mencionar que as abordagens descritas no Capítulo 7 para resolver sistemas não-lineares podem ser aplicadas a sistemas lineares. No entanto, nesta seção o foco será em abordagens que foram criadas especialmente para sistemas lineares.

11.3.1 Excel

Há duas formas de resolver equações algébricas lineares com Excel: (1) usando a ferramenta Solver ou (2) usando funções de inversão e multiplicação de matrizes.

Lembre-se de que uma forma de determinar a solução de equações algébricas lineares é

$$\{X\} = [A]^{-1}\{B\} \tag{11.12}$$

O Excel tem funções embutidas tanto para a inversão quanto para a multiplicação de matrizes que podem ser usadas para implementar essa fórmula.

EXEMPLO 11.4 Usando o Excel para Resolver Sistemas Lineares

Enunciado do Problema. Lembre-se de que no Capítulo 10 foi apresentada a matriz de Hilbert, na qual o sistema a seguir é baseado. Observe que ela já está em uma escala adequada, como foi feito previamente no Exemplo 10.3, de modo que o maior coeficiente em cada linha é a unidade.

$$\begin{bmatrix} 1 & 1/2 & 1/3 \\ 1 & 2/3 & 1/2 \\ 1 & 3/4 & 3/5 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 1{,}833333 \\ 2{,}166667 \\ 2{,}35 \end{Bmatrix}$$

A solução desse sistema é $\{X\}^T = \lfloor 1\ 1\ 1 \rfloor$. Use o Excel para obtê-la.

Solução. A planilha eletrônica para resolver esse problema é mostrada na Figura 11.7. Primeiro, a matriz $[A]$ e as constantes $\{B\}$ do lado direito são digitadas nas células da planilha. Então, um conjunto de células de dimensões apropriadas (no exemplo, 3×3) é selecionado ou clicando e arrastando o mouse ou utilizando as setas do teclado enquanto se mantém pressionada a tecla *shift*. Como na Figura 11.7, selecionamos o campo B5. .D7.

	A	B	C	D	E	F
1		1	0.5	0.33333333		1.83333333333333
2	[A] =	1	0.66666667	0.5	{B} =	2.16666666666667
3		1	0.75	0.6		2.35000000000000
4						
5		**9**	**-18**	**10**		**0.99999999999992**
6	[A]-1 =	**-36**	**96**	**-60**	{X} =	**1.00000000000043**
7		**30**	**-90**	**60**		**0.99999999999960**

=MINVERSE(B1:D3) **=MMULT(B5:D7,F1:F3)**

FIGURA 11.7

A seguir, a fórmula chamando a função de inversão de matriz é digitada,

```
=minverse(B1..D3)
```

Observe que o argumento é o campo que contém os elementos de [*A*]. As teclas *Ctrl* e *Shift* são mantidas pressionadas enquanto a tecla *Enter* é apertada. A inversa de [*A*] resultante será calculada pelo Excel e mostrada no campo B5. .D7, como ilustrado na Figura 11.7.

Uma abordagem parecida é usada para multiplicar a inversa pelo vetor do lado direito. Para isso, o intervalo F5. .F7 é selecionado e a seguinte fórmula é digitada:

```
=mmult(B5..D7,F1..F3)
```

onde o primeiro campo é a primeira matriz a ser multiplicada, $[A]^{-1}$, e o segundo campo é a segunda matriz a ser multiplicada, $\{B\}$. Novamente pressionando *Ctrl-Shift-Enter*, a solução, $\{X\}$, será calculada pelo Excel e mostrada no campo F5. .F7, como ilustrado na Figura 11.7. Como pode ser visto, o resultado é a resposta correta.

Observe que deliberadamente os resultados no Exemplo 11.4 foram reformatados para mostrar 15 dígitos. Isso foi feito porque o Excel usa dupla precisão para armazenar valores numéricos. Assim, vê-se que os erros de arredondamento ocorrem nos últimos dois dígitos. Isso implica um número de condicionamento da ordem de 100, o que está de acordo com o resultado de 451,2 originalmente calculado no Exemplo 10.3. O Excel não tem o recurso de calcular um número de condicionamento. Na maioria dos casos, principalmente porque ele usa números em dupla precisão, isso não representa um problema. No entanto, para casos em que se suspeita que o sistema é mal condicionado, a determinação do número de condicionamento é útil. O software MATLAB e o IMSL são capazes de calcular esse número.

11.3.2 MATLAB

Como o nome sugere, MATLAB (abreviação para MATriz e LABoratório) foi criado para facilitar manipulações de matrizes. Assim, como seria de esperar, seus recursos nessa área são excelentes. Algumas das funções principais do MATLAB, relacionadas a operações de matrizes, estão listadas na Tabela 11.1. O seguinte exemplo ilustra um pouco desses recursos.

TABELA 11.1 Funções do MATLAB para implementar análise de matrizes e álgebra linear numérica.

Análise de Matrizes		Equações Lineares	
Função	**Descrição**	**Função**	**Descrição**
cond	Número de condicionamento da matriz	\ and /	Solução da equação linear; usa "help slash"
norm	Norma da matriz ou do vetor	chol	Fatorização de Cholesky
rcond	Estimador de condicionamento recíproco LINPACK	lu	Fatores da eliminação de Gauss
rank	Número de linhas ou colunas linearmente independentes	inv	Inversa da matriz
det	Determinante	qr	Decomposição ortogonal-triangular
trace	Soma dos elementos da diagonal	qrdelete	Apaga uma coluna da fatorização QR
null	Espaço nulo	qrinsert	Insere uma coluna na fatorização QR
orth	Ortogonalização	nnls	Mínimos quadrados não-negativos
rref	Forma reduzida escalonada por linha	pinv	Pseudo-inversa
		lscov	Mínimos quadrados na presença de covariância conhecida

EXEMPLO 11.5 Usando o MATLAB para Manipular Equações Algébricas Lineares

Enunciado do Problema. Explore como o MATLAB pode ser usado para resolver e analisar equações algébricas lineares. Use o mesmo sistema do Exemplo 11.4.

Solução. Primeiro, pode-se digitar a matriz $[A]$ e o vetor $\{B\}$,

```
>> A = [ 1  1/2  1/3 ; 1  2/3  2/4 ; 1  3/4  3/5 ]

A =
   1.0000    0.5000    0.3333
   1.0000    0.6667    0.5000
   1.0000    0.7500    0.6000

>> B=[1+1/2+1/3;1+2/3+2/4;1+3/4+3/5]

B =
   1.8333
   2.1667
   2.3500
```

Então, determina-se o número de condicionamento para $[A]$, como em

```
>> Cond(A)

ans =
  366.3503
```

Esse resultado é baseado em $\|A\|_2$, ou seja, a norma espectral discutida no Quadro 10.2. Observe que ele é da mesma ordem de grandeza que o número de condicionamento 451,2 baseado na norma da soma das linhas do Exemplo 10.3. Ambos os resultados implicam que podem ser perdidos entre 2 e 3 algarismos de precisão.

Agora, é possível resolver o sistema de equações de duas formas diferentes. O mais direto e eficiente é empregar a barra invertida, ou "divisão à esquerda":

```
>> X=A\B

X =
   1.0000
   1.0000
   1.0000
```

Para casos como o nosso, o MATLAB usa a eliminação de Gauss para resolver tais sistemas.

Como uma alternativa, pode-se implementar a Equação (PT3.6) diretamente, como em

```
>> X=inv(A)*B

X =
   1.0000
   1.0000
   1.0000
```

Essa abordagem de fato determina primeiro a matriz inversa e depois efetua a multiplicação matricial. Portanto, consome mais tempo do que usando a barra invertida.

11.3.3 IMSL

O IMSL tem diversas rotinas para sistemas lineares, inversão de matrizes e cálculo de determinantes. A Tabela 11.2 enumera as categorias cobertas.

Como se vê na Tabela 11.3, oito rotinas são dedicadas ao caso específico de matrizes reais arbitrárias. Nessa discussão, iremos concentrar-nos em duas rotinas: LFCRG e LFIRG.

A rotina LFCRG efetua uma decomposição *LU* da matriz [*A*] e calcula seu número de condicionamento. A LFCRG é implementada pelo seguinte comando CALL:

```
CALL LFCRG(N, A, LDA, FAC, LDFAC, IPVT, RCOND)
```

onde

N = Ordem da matriz. (Entrada)

A = Matriz N × N a ser decomposta. (Entrada)

LDA = Dimensão dominante de A como especificado na declaração de dimensão dada pelo programa. (Entrada)

FAC = Matriz N × N contendo a decomposição *LU* de A. (Saída)

LDFAC = Dimensão dominante de FAC como especificado na declaração de dimensão dada pelo programa. (Entrada)

IPVT = Vetor de comprimento N contendo a informação de pivotamento para a decomposição *LU*. (Saída)

RCOND = Escalar contendo o recíproco do número de condicionamento de A. (Saída)

TABELA 11.2 Categorias de rotinas de IMSL para a solução de sistemas lineares, inversão de matrizes e cálculo de determinantes.

- Matrizes reais arbitrárias
- Matrizes complexas arbitrárias
- Matrizes triangulares reais
- Matrizes triangulares complexas
- Matrizes reais positivas definidas
- Matrizes reais simétricas
- Matrizes complexas hermitianas positivas definidas
- Matrizes complexas hermitianas
- Matrizes de banda reais em armazenamento de banda
- Matrizes de banda reais simétricas positivas definidas em armazenamento de banda
- Matrizes de banda complexas em armazenamento de banda
- Matrizes de banda complexas simétricas positivas definidas em armazenamento de banda
- Solucionador de equações lineares reais esparsas
- Solucionador de equações lineares complexas esparsas
- Solucionador de equações lineares reais esparsas simétricas positivas definidas
- Solucionador de equações lineares complexas esparsas hermitianas positivas definidas
- Matrizes Toeplitz reais em armazenamento Toeplitz
- Matrizes Toeplitz complexas em armazenamento Toeplitz
- Matrizes complexas circulantes em armazenamento circulante
- Métodos iterativos
- Mínimos quadrados lineares e fatorização de matrizes
- Mínimos quadrados, decomposição QR e a inversa generalizada
- Fatorização de Cholesky
- Decomposição em valores singulares (SVD)
- Apoio matemático para estatística

TABELA 11.3 Rotinas IMSL para a solução de matrizes reais arbitrárias.

Rotina	Capacidade
LSARG	Resolve sistemas lineares com alta acurácia
LSLRG	Resolve um sistema linear
LFCRG	Fatora e calcula o número de condicionamento
LFTRG	Fatora
LFSRG	Resolve depois de fatorar
LFIRG	Resolve sistemas lineares com alta acurácia depois de fatorar
LFDRG	Calcula o determinante depois de fatorar
LINRG	Inverte

O LFIRG usa a decomposição *LU* e um vetor particular do lado direito para gerar uma solução de alta acurácia utilizando refinamento iterativo. O LFIRG é implementado pelo seguinte comando CALL:

```
CALL LFIRG(N, A, LDA, FAC, LDFAC, IPVT, B, IPATH, X, RES)
```

onde

N = Ordem da matriz. (Entrada)

A = Matriz N × N a ser decomposta. (Entrada)

LDA = Dimensão dominante de A como especificado na declaração de dimensão dada pelo programa. (Entrada)

FAC = Matriz N × N contendo a decomposição *LU* de A. (Entrada)

LDFAC = Dimensão dominante de FAC como especificado na declaração de dimensão dada pelo programa. (Entrada)

IPVT = Vetor de comprimento N contendo a informação de pivotamento para a decomposição *LU*. (Entrada)

B = Vetor de comprimento N contendo o lado direito do sistema linear.

IPATH = Indicador de caminho. (Entrada)

= 1 significa que o sistema AX = B deve ser resolvido

= 2 significa que o sistema $A^TX = B$ deve ser resolvido

X = Vetor de comprimento N contendo a solução do sistema linear. (Saída)

RES = Vetor de comprimento N contendo o vetor residual da solução refinada. (Saída)

Essas duas rotinas são usadas no seguinte exemplo. Primeiro, o LFCRG é chamado para decompor a matriz e devolver o número de condicionamento. Então, o LFIRG é chamado N vezes com o vetor B contendo cada coluna da matriz identidade para gerar as colunas da matriz inversa. Finalmente, LFIRG pode ser chamado uma vez adicional para obter a solução de um único vetor do lado direito.

EXEMPLO 11.6 Usando o IMSL para Analisar e Resolver uma Matriz de Hilbert

Enunciado do Problema. Use LFCRG e LFIRG para determinar o número de condicionamento, a matriz inversa e a solução para o sistema com a matriz de Hilbert,

$$\begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 1{,}833333 \\ 1{,}083333 \\ 0{,}783333 \end{Bmatrix}$$

Solução. Um exemplo de um programa principal em Fortran 90 usando LFCRG e LFIRG para resolver esse problema pode ser escrito como

```
PROGRAM Lineqs
USE msimsl
IMPLICIT NONE
INTEGER::ipath,lda,n,ldfac
PARAMETER(ipath=1,lda=3,ldfac=3,n=3)
INTEGER::ipvt(n),i,j,itmax=50
REAL::A(lda,lda),Ainv(lda,lda),factor(ldfac,ldfac),Rcond,Res(n)
REAL::Rj(n),B(n),x(n)

DATA A/1.0,0.5,0.3333333,0.5,0.3333333,0.25,0.3333333,0.25,0.2/
DATA B/1.833333,1.083333,0.783333/
!Fazer decomposição LU; determinar e mostrar o número de condicionamento

CALL LFCRG(n,A,lda,factor,ldfac,ipvt,Rcond)
PRINT *, "número de condicionamento = ", 1.0E0/Rcond
PRINT *

!Inicializar o vetor Rj com zero
DO i = 1,n
  Rj(i) = 0.
END DO

!Alimentar as colunas da matriz identidade através do LFIRG para
gerar a inversa e armazenar o resultado em Ainv. Mostra Ainv

DO j = 1, n
  Rj(j) = 1.0
  CALL LFIRG(n,A,lda,fator,ldfac,ipvt,Rj,ipath,ainv(1,j),Res)
  Rj(j) = 0.0
END DO
PRINT *, "Inversa da Matriz:"
DO i = 1,n
  PRINT *, (Ainv(i,j),j=1,n)
END DO
PRINT *

!Usar LFIRG para obter solução para B. Mostrar os resultados

PRINT *, "Solução:"
DO I = 1,n
  PRINT *, x(i)
END DO

END PROGRAM
```

A saída é:

```
Número de Condicionamento =    680.811600

Inversa da Matriz:
       9.000033      -36.000180       30.000160
     -36.000180      192.000900     -180.000800
      30.000160     -180.000800      180.000800

Solução:
   9.999986E-01
       1.000010
   9.999884E-01
```

Novamente, o número de condicionamento é da mesma ordem de grandeza que o número de condicionamento baseado na norma da soma das linhas no Exemplo 10.3 (451,2). Ambos os resultados implicam que entre 2 e 3 dígitos de precisão podem ser perdidos. Isso é confirmado pela solução, na qual se vê que os erros de arredondamento ocorrem nos últimos dois ou três algarismos.

PROBLEMAS

11.1 Efetue os mesmos cálculos do Exemplo 11.1, agora para o sistema triangular

$$\begin{bmatrix} 0{,}8 & -0{,}4 & \\ -0{,}4 & 0{,}8 & -0{,}4 \\ & -0{,}4 & 0{,}8 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 41 \\ 25 \\ 105 \end{Bmatrix}$$

11.2 Determine a matriz inversa para o Exemplo 11.1, baseando-se na decomposição *LU* e nos vetores unitários.

11.3 O seguinte sistema tridiagonal deve ser resolvido como parte de um algoritmo maior (Crank-Nicolson) para resolver equações diferenciais parciais:

$$\begin{bmatrix} 2{,}01475 & -0{,}020875 & & \\ -0{,}020875 & 2{,}01475 & -0{,}020875 & \\ & -0{,}020875 & 2{,}01475 & -0{,}020875 \\ & & -0{,}020875 & 2{,}01475 \end{bmatrix} \times \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} 4{,}175 \\ 0 \\ 0 \\ 2{,}0875 \end{Bmatrix}$$

Use o algoritmo de Thomas para obter uma solução.

11.4 Confirme a validade da decomposição de Cholesky do Exemplo 11.2 substituindo os resultados na Equação (11.2) para verificar que o produto de $[L]$ por $[L]^T$ fornece $[A]$.

11.5 Efetue a decomposição de Cholesky do seguinte sistema simétrico, à mão:

$$\begin{bmatrix} 8 & 20 & 15 \\ 20 & 80 & 50 \\ 15 & 50 & 60 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 50 \\ 250 \\ 100 \end{Bmatrix}$$

11.6 Efetue os mesmos cálculos que os do Exemplo 11.2 para o seguinte sistema simétrico

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152{,}6 \\ 585{,}6 \\ 2488{,}8 \end{Bmatrix}$$

Além de determinar a decomposição de Cholesky, empregue-a para determinar os *a's*.

11.7 **(a)** Use o método de Gauss-Seidel para resolver o sistema triangular do Problema 11.1 ($\varepsilon_s = 5\%$). **(b)** Repita **(a)**, mas use sobre-relaxamento com $\lambda = 1{,}2$.

11.8 Lembre-se, do Problema 10.8, que o seguinte sistema de equações é criado para determinar concentrações (com os *c*'s em g/m^3) em uma série de reatores acoplados como função da quantidade de entrada de massa em cada reator (o lado direito, em g/d)

$$\begin{aligned} 15c_1 - 3c_2 - c_3 &= 3800 \\ -3c_1 + 18c_2 - 6c_3 &= 1200 \\ -4c_1 - c_2 + 12c_3 &= 2350 \end{aligned}$$

Resolva esse problema pelo método de Gauss-Seidel até $\varepsilon_s = 5\%$.

11.9 Repita o Problema 11.8, mas use iteração de Jacobi.

11.10 Use o método de Gauss-Seidel para resolver o seguinte sistema até que o erro relativo porcentual caia abaixo de $\varepsilon_s = 5\%$,

$$\begin{aligned} 10x_1 + 2x_2 - x_3 &= 27 \\ -3x_1 - 6x_2 + 2x_3 &= -61{,}5 \\ x_1 + x_2 + 5x_3 &= -21{,}5 \end{aligned}$$

11.11 Use o método de Gauss-Seidel **(a)** sem relaxamento e **(b)** com relaxamento ($\lambda = 0{,}95$) para resolver o seguinte sistema com tolerância de $\varepsilon_s = 5\%$. Se necessário, reorganize as equações para garantir convergência.

$$\begin{aligned} -3x_1 + x_2 + 12x_3 &= 50 \\ 6x_1 - x_2 - x_3 &= 3 \\ 6x_1 + 9x_2 + x_3 &= 40 \end{aligned}$$

11.12 Use o método de Gauss-Seidel **(a)** sem relaxamento e **(b)** com relaxamento ($\lambda = 1{,}2$) para resolver o seguinte sistema com tolerância de $\varepsilon_s = 5\%$. Se necessário, reorganize as equações para garantir a convergência.

$$\begin{aligned} 2x_1 - 6x_2 - x_3 &= -38 \\ -3x_1 - x_2 + 7x_3 &= -34 \\ -8x_1 + x_2 - 2x_3 &= -20 \end{aligned}$$

11.13 Desenhe novamente a Figura 11.5 para o caso em que as inclinações das equações forem 1 e –1. Qual o resultado de aplicar Gauss-Seidel a tal sistema?

11.14 Dos seguintes três conjuntos de equações lineares, identifique qual(is) conjunto(s) não pode(m) ser resolvido(s) usando um método iterativo como o de Gauss-Seidel. Mostre, utilizando qualquer número de iterações, que necessariamente sua solução não converge. Enuncie claramente seu critério de convergência (como você sabe que a solução não está convergindo).

Conjunto Um	Conjunto Dois	Conjunto Três
$9x + 3y + z = 13$	$x + y + 6z = 8$	$-3x + 4y + 5z = 6$
$-6x + 8z = 2$	$x + 5y - z = 5$	$-2x + 2y - 3z = -3$
$2x + 5y - z = 6$	$4x + 2y - 2z = 4$	$2y - z = 1$

11.15 Use a biblioteca ou pacote de software de sua preferência para obter uma solução, calcular a inversa e determinar o número de condicionamento (sem mudança de escala) com base na norma da soma das linhas para

(a)

$$\begin{bmatrix} 1 & 4 & 9 \\ 4 & 9 & 16 \\ 9 & 16 & 25 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 14 \\ 29 \\ 50 \end{Bmatrix}$$

(b)

$$\begin{bmatrix} 1 & 4 & 9 & 16 \\ 4 & 9 & 16 & 25 \\ 9 & 16 & 25 & 36 \\ 16 & 25 & 36 & 49 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{Bmatrix} = \begin{Bmatrix} 30 \\ 54 \\ 86 \\ 126 \end{Bmatrix}$$

Em ambos os casos, as respostas para todos os *x*'s devem ser 1.

11.16 Dado o par de equações simultâneas não-lineares:

$$\begin{aligned} f(x, y) &= 4 - y - 2x^2 \\ g(x, y) &= 8 - y^2 - 4x \end{aligned}$$

(a) Use o Solver do Excel para determinar os dois pares de valores de x e y que satisfazem essas equações.
(b) Usando um intervalo de aproximações iniciais ($x = -6$ até 6 e $y = -6$ até 6), determine quais aproximações iniciais resultarão em cada uma das soluções.

11.17 Uma companhia de eletrônica produz transistores, resistores e chips de computador. Cada transistor usa quatro unidades de cobre, uma unidade de zinco e duas unidades de vidro. Cada resistor usa três, três e uma unidades de cada material, respectivamente, e cada chip de computador usa duas, uma e três unidades desses materiais, respectivamente. Colocando essa informação em uma tabela, tem-se:

Componente	Cobre	Zinco	Vidro
Transistor	4	1	2
Resistor	3	3	1
Chip de computador	2	1	3

O fornecimento desses materiais varia de semana para semana. Assim, a companhia precisa determinar uma meta de produção diferente para cada semana. Por exemplo, em uma semana a quantidade total de materiais disponíveis era 960 unidades de cobre, 510 unidades de zinco e 610 unidades de vidro. Determine o sistema de equações que modela essa meta de produção e use o Excel e a informação deste capítulo sobre a resolução de sistemas lineares com o Excel para determinar o número de transistores, resistores e chips de computador fabricados nessa semana.

11.18 Use o software MATLAB para determinar o número de condicionamento espectral para uma matriz de Hilbert de dimensão 10. Quantos dígitos de precisão podem ser perdidos por causa do mal condicionamento? Determine a solução desse sistema no caso em que cada elemento do vetor à direita $\{b\}$ consiste na soma dos coeficientes de sua linha. Em outras palavras, resolva no caso em que todas as variáveis devem ser exatamente um. Compare os erros resultantes com aqueles esperados com base no número de condicionamento.

11.19 Repita o Problema 11.18, mas para o caso de uma matriz de Vandermonde de dimensão 6 (ver Problema 10.14), onde $x_1 = 4$, $x_2 = 2$, $x_3 = 7$, $x_4 = 10$, $x_5 = 3$ e $x_6 = 5$.

11.20 Na Seção 9.2.1, determinamos o número de operações necessárias para eliminação de Gauss sem pivotamento parcial. Faça uma determinação semelhante para o algoritmo de Thomas (Figura 11.2). Faça um gráfico de operações *versus* n (de 2 a 20) para ambas as técnicas.

11.21 Desenvolva um programa amigável ao usuário em uma linguagem de macro ou de alto nível à sua escolha para obter uma solução para um sistema tridiagonal com o algoritmo de Thomas (Figura 11.2). Teste seu programa repetindo o Exemplo 11.1.

11.22 Desenvolva um programa amigável ao usuário em uma linguagem de macro ou de alto nível à sua escolha para a decomposição de Cholesky com base na Figura 11.3. Teste seu programa repetindo o Exemplo 11.2.

11.23 Desenvolva um programa amigável ao usuário em uma linguagem de macro ou de alto nível à sua escolha para o método de Gauss-Seidel com base na Figura 11.6. Teste seu programa repetindo o Exemplo 11.3.

CAPÍTULO 12

Estudo de Casos: Equações Algébricas Lineares

O propósito deste capítulo é usar os procedimentos numéricos discutidos nos Capítulos 9, 10 e 11 para resolver sistemas de equações algébricas lineares para alguns estudos de casos da engenharia. Essas técnicas numéricas sistemáticas têm importância prática, pois os engenheiros freqüentemente encontram problemas envolvendo sistemas de equações que são muito grandes para serem resolvidos à mão. Os algoritmos numéricos nessas aplicações são particularmente convenientes para serem implementados em computadores pessoais.

A *Seção 12.1* mostra como um balanço de massas pode ser empregado para modelar um sistema de reatores. A *Seção 12.2* enfatiza de modo especial o uso da inversão de matrizes para determinar as complexas interações de causa e efeito entre as forças nos membros de uma treliça. A *Seção 12.3* é um exemplo do uso das leis de Kirchhoff para calcular as correntes e voltagens em um circuito de resistores. Finalmente, a *Seção 12.4* é uma ilustração de como as equações lineares são empregadas na determinação da configuração estacionária em um sistema massa-mola.

12.1 ANÁLISE DO ESTADO ESTACIONÁRIO DE UM SISTEMA DE REATORES (ENGENHARIA QUÍMICA/BIOENGENHARIA)

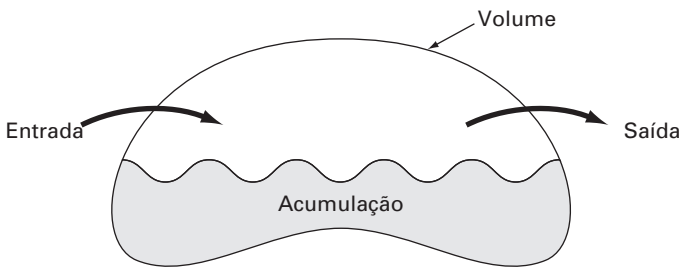
Fundamentos. Um dos princípios de organização mais importantes na engenharia química é a *conservação de massa* (lembre-se da Tabela 1.1). Em termos quantitativos, o princípio é expresso como um balanço de massas que contabiliza todas as fontes e todos os sorvedouros de um material que passa para dentro e para fora de um volume (Figura 12.1). Em um período finito de tempo, isso pode ser expresso como:

$$\text{Acumulação} = \text{entradas} - \text{saídas} \tag{12.1}$$

O balanço de massas representa um exercício de contabilidade para a substância específica sendo modelada. Para o período do cálculo, se as entradas forem maiores do que as saídas, a massa da substância dentro do volume cresce. Se as saídas forem maiores do que as entradas, a massa diminui. Se as entradas forem iguais às saídas, a acumulação é zero, e a massa permanece constante. Para essa condição estável, ou estacionária, a Equação (12.1) pode ser expressa como:

$$\text{Entradas} = \text{saídas} \tag{12.2}$$

FIGURA 12.1
Representação esquemática do balanço de massa.



Use a conservação da massa para determinar as concentrações estacionárias de um sistema de reatores acoplados.

Solução. O balanço de massas pode ser usado para resolver problemas de engenharia expressando as entradas e as saídas em termos de variáveis e parâmetros mensuráveis. Por exemplo, se estivermos fazendo o balanço de massas para uma substância conservativa (isto é, uma que não aumenta nem diminui por causa das transformações químicas) em um reator (Figura 12.2), devemos quantificar a taxa pela qual a massa escoa para dentro do reator através dos dois tubos de entrada e para fora do reator através do tubo de saída. Isso pode ser feito tomando-se o produto da vazão em volume Q (em m^3/min) pela concentração c (em mg/m^3) para cada tubo. Por exemplo, para o tubo 1 na Figura 12.2, $Q_1 = 2$ m^3/min e $c_1 = 25$ mg/m^3; assim, a taxa na qual a massa escoa para dentro do reator através do tubo 1 é $Q_1c_1 = (2\ m^3/min)(25\ mg/m^3) = 50$ mg/min. Então, 50 mg de produtos químicos entram no reator através desse tubo por minuto. Analogamente, para o tubo 2, a taxa de entrada de massa (vazão em massa) pode ser calculada como $Q_2c_2 = (1{,}5\ m^3/min)(10\ mg/m^3) = 15$ mg/min.

Observe que a concentração que sai do reator através do tubo 3 não é especificada pela Figura 12.2, porque já temos informação suficiente para calculá-la com base na conservação de massa. Como o reator está em um estado estacionário, a Equação (12.2) é válida e as entradas devem ser balanceadas pelas saídas, como em

$$Q_1c_1 + Q_2c_2 = Q_3c_3$$

Substituindo os valores dados nessa equação, temos:

$$50 + 15 = 3{,}5c_3$$

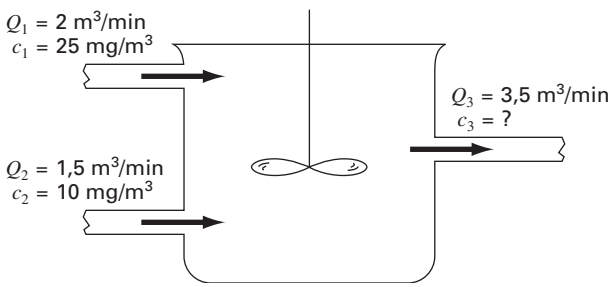


FIGURA 12.2
Um reator completamente misturado, em estado estacionário, com dois tubos de entrada e um tubo de saída. As vazões Q estão em metros cúbicos por minuto e a concentração c está em miligramas por metro cúbico.

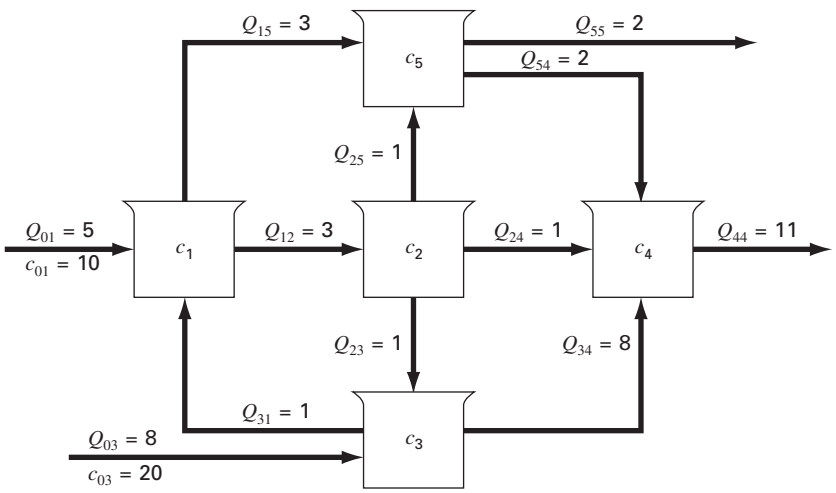


FIGURA 12.3
Cinco reatores interligados por tubos.

o que fornece $c_3 = 18{,}6$ mg/m^3. Assim, determinamos a concentração no terceiro tubo. No entanto, os cálculos levam a um bônus adicional. Como o reator é bem misturado (como representado pelo propulsor na Figura 12.2), a concentração será uniforme, ou homogênea, em todo o tanque. Desse modo, a concentração no tubo 3 deve ser idêntica à concentração no reator. Conseqüentemente, o balanço de massas permitiu calcular a concentração tanto no reator quanto no tubo de saída. Tal informação é de grande utilidade para engenheiros químicos e petroquímicos que devem projetar reatores que contêm misturas com uma concentração específica.

Como apenas álgebra simples foi usada para determinar a concentração para um único reator na Figura 12.2, talvez não fique óbvio onde computadores aparecem em cálculos de balanço de massa. A Figura 12.3 mostra um problema em que computadores não apenas são úteis como são uma necessidade prática. Pelo fato de haver cinco reatores conectados, ou acoplados, cinco equações simultâneas de balanço de massa são necessárias para caracterizar o sistema. Para o reator 1, a taxa de entrada de massa é

$$5(10) + Q_{31}c_3$$

e a taxa de saída de massa é

$$Q_{12}c_1 + Q_{15}c_1$$

Como o sistema está em um estado estacionário, as entradas e saídas devem ser iguais:

$$5(10) + Q_{31}c_3 = Q_{12}c_1 + Q_{15}c_1$$

ou, substituindo os valores da Figura 12.3,

$$6c_1 - c_3 = 50$$

Equações semelhantes podem ser geradas para os outros reatores:

$$-3c_1 + 3c_2 = 0$$

$$-c_2 + 9c_3 = 160$$

$$-c_2 - 8c_3 + 11c_4 - 2c_5 = 0$$

$$-3c_1 - c_2 + 4c_5 = 0$$

Um método numérico pode ser usado para resolver essas cinco equações e determinar as cinco concentrações desconhecidas:

$$\{C\}^T = \lfloor 11,51 \quad 11,51 \quad 19,06 \quad 17,00 \quad 11,51 \rfloor$$

Além disso, a matriz inversa pode ser calculada como sendo

$$[A]^{-1} = \begin{bmatrix} 0{,}16981 & 0{,}00629 & 0{,}01887 & 0 & 0 \\ 0{,}16981 & 0{,}33962 & 0{,}01887 & 0 & 0 \\ 0{,}01887 & 0{,}03774 & 0{,}11321 & 0 & 0 \\ 0{,}06003 & 0{,}07461 & 0{,}08748 & 0{,}09091 & 0{,}04545 \\ 0{,}16981 & 0{,}08962 & 0{,}01887 & 0 & 0{,}25000 \end{bmatrix}$$

Cada um dos elementos a_{ij} significa a mudança na concentração no reator i por causa de uma variação unitária na carga do reator j. Assim, os zeros na coluna 4 indicam que uma carga no reator 4 não vai ter efeito nos reatores 1, 2, 3 e 5. Isso é consistente com a configuração do sistema (Figura 12.3), que mostra que o escoamento para fora do reator 4 não retroalimenta nenhum dos outros reatores. Em contraste, cargas em qualquer dos três primeiros reatores irão afetar o sistema inteiro, como indicado pela ausência de zeros nas três primeiras colunas. Tal informação é de grande utilidade para os engenheiros que projetam e administram esses sistemas.

12.2 ANÁLISE DE UMA TRELIÇA ESTATICAMENTE DETERMINADA (ENGENHARIA CIVIL/AMBIENTAL)

Fundamentos. Um problema importante em engenharia estrutural é o de determinar as forças e reações associadas a uma treliça estaticamente determinada. A Figura 12.4 mostra um exemplo de tal treliça.

As forças (F) representam a tensão ou a compressão dos membros da treliça. As reações externas (H_2, V_2 e V_3) são forças que caracterizam como a treliça interage com a superfície que a suporta. A dobradiça no nó 2 pode transmitir tanto forças horizontais quanto verticais à superfície, enquanto o rolete no nó 3 transmite apenas forças verticais. Observe que o efeito da carga externa de 1.000 lb é distribuído ao longo dos vários membros da treliça.

Solução. Esse tipo de estrutura pode ser descrito por um sistema de equações lineares acopladas. Os diagramas de força de corpo livre são mostrados para cada nó na Figura 12.5. A soma das forças nas direções horizontal e vertical deve ser zero em cada nó, pois o sistema está em repouso. Assim, para o nó 1,

$$\Sigma F_H = 0 = -F_1 \cos 30^\circ + F_3 \cos 60^\circ + F_{1,h} \tag{12.3}$$

$$\Sigma F_V = 0 = -F_1 \text{ sen } 30^\circ - \mathrm{F}_3 \text{ sen } 60^\circ + \mathrm{F}_{1,\mathrm{v}} \tag{12.4}$$

Para o nó 2,

$$\Sigma F_H = 0 = F_2 + F_1 \cos 30^\circ + F_{2,h} + H_2 \tag{12.5}$$

$$\Sigma F_V = 0 = F_1 \text{ sen } 30^\circ + \mathrm{F}_{2,\mathrm{v}} + \mathrm{V}_2 \tag{12.6}$$

FIGURA 12.4
Forças em uma treliça estaticamente determinada.

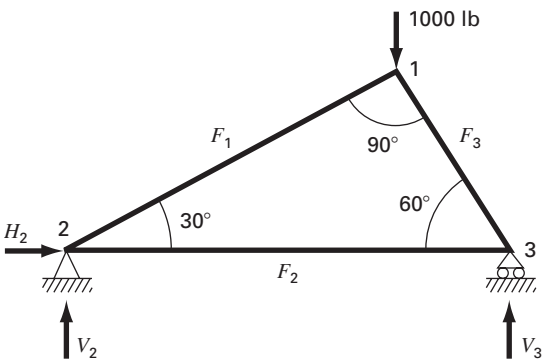
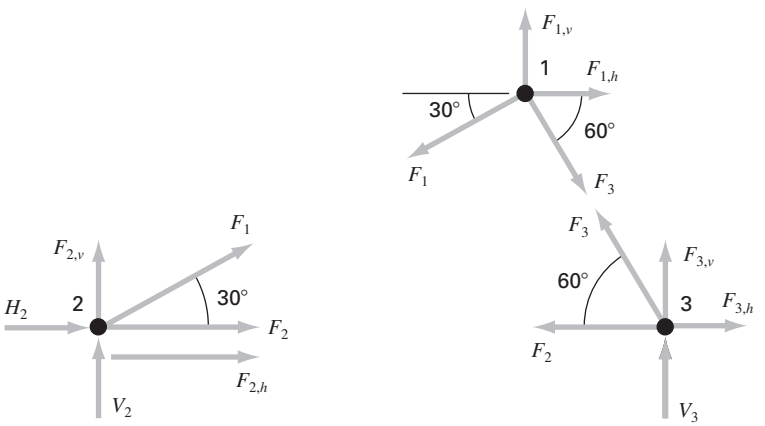


FIGURA 12.5
Diagrama de forças de corpo livre para os nós de uma treliça estaticamente determinada.



Para o nó 3,

$$\Sigma F_H = 0 = -F_2 - F_3 \cos 60^\circ + F_{3,h} \tag{12.7}$$

$$\Sigma F_V = 0 = F_3 \text{ sen } 60^\circ + \mathrm{F_{3,v}} + \mathrm{V_3} \tag{12.8}$$

onde $F_{i,h}$ é a força externa horizontal aplicada ao nó i (onde uma força positiva age da esquerda para a direita) e $F_{1,v}$ é a força externa vertical aplicada ao nó i (onde uma força positiva age para cima). Assim, nesse problema, a força de 1.000 lb para baixo no nó 1 corresponde a $F_{1,v} = -1.000$. Para esse caso, todas as outras $F_{i,v}$'s e $F_{i,h}$'s são nulas. Observe que os sentidos das forças internas e das forças de reações são desconhecidos. A aplicação apropriada das leis de Newton exige apenas hipóteses consistentes sobre o sentido. Se os sentidos forem escolhidos de forma incorreta, as soluções serão negativas. Observe também que, nesse problema, assumiu-se que as forças em todos os membros são de tensão e agem de modo a aproximar os nós adjacentes. Então, uma solução negativa corresponde a compressão. Esse problema pode ser escrito como o seguinte sistema de seis equações e seis incógnitas:

$$\begin{bmatrix} 0{,}866 & 0 & -0{,}5 & 0 & 0 & 0 \\ 0{,}5 & 0 & 0{,}866 & 0 & 0 & 0 \\ -0{,}866 & -1 & 0 & -1 & 0 & 0 \\ -0{,}5 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0{,}5 & 0 & 0 & 0 \\ 0 & 0 & -0{,}866 & 0 & 0 & -1 \end{bmatrix} \begin{Bmatrix} F_1 \\ F_2 \\ F_3 \\ H_2 \\ V_2 \\ V_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ -1000 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix} \tag{12.9}$$

Observe que, como formulado na Equação (12.9), o pivotamento parcial é necessário para evitar divisão por elementos diagonais nulos. Empregando-se uma estratégia de pivotamento, o sistema pode ser resolvido usando-se qualquer das técnicas de eliminação discutidas nos Capítulos 9 ou 10. No entanto, como esse problema é um estudo de caso ideal para demonstrar a utilidade da matriz inversa, a decomposição *LU* pode ser usada para calcular

$$F_1 = -500 \qquad F_2 = 433 \qquad F_3 = -866$$
$$H_2 = 0 \qquad V_2 = 250 \qquad V_3 = 750$$

e a matriz inversa é

$$[A]^{-1} = \begin{bmatrix} 0{,}866 & 0{,}5 & 0 & 0 & 0 & 0 \\ 0{,}25 & -0{,}433 & 0 & 0 & 1 & 0 \\ -0{,}5 & 0{,}866 & 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & -1 & 0 \\ -0{,}433 & -0{,}25 & 0 & -1 & 0 & 0 \\ 0{,}433 & -0{,}75 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Agora, observe que o vetor do lado direito representa as forças horizontais e verticais externas aplicadas em cada nó, assim

$$\{F\}^T = \lfloor F_{1,h} \quad F_{1,v} \quad F_{2,h} \quad F_{2,v} \quad F_{3,h} \quad F_{3,v} \rfloor \tag{12.10}$$

Como as forças externas não têm nenhum efeito na decomposição *LU*, o método não precisa ser implementado várias vezes para se estudar o efeito de diferentes forças externas na treliça. Em vez disso, tudo o que se tem de fazer é efetuar os passos de substituição progressiva e regressiva para cada vetor do lado direito para obter eficientemente várias soluções. Por exemplo, pode-se estar interessado em estudar o efeito de forças horizontais induzidas por um vento soprando da esquerda para a direita. Se a força do vento puder ser idealizada como duas forças pontuais de 1.000 lb nos nós 1 e 2 (Figura 12.6*a*), o vetor à direita é

$$\{F\}^T = \lfloor -1000 \quad 0 \quad 1000 \quad 0 \quad 0 \quad 0 \rfloor$$

que pode ser usado para calcular

$$F_1 = 866 \qquad F_2 = 250 \qquad F_3 = -500$$
$$H_2 = -2000 \qquad V_2 = -433 \qquad V_3 = 433$$

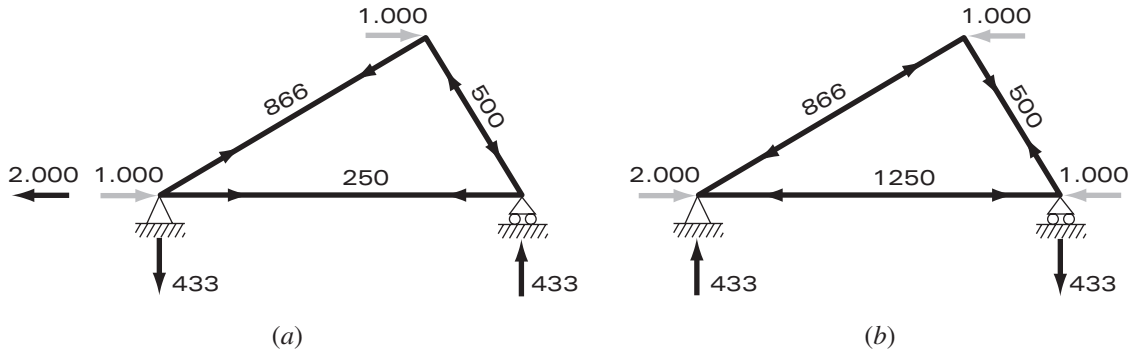


FIGURA 12.6
Dois casos de teste mostrando (*a*) ventos a partir da esquerda e (*b*) ventos a partir da direita.

Para um vento a partir da direita (Figura 12.6*b*), $F_{1,h} = -1.000$, $F_{3,h} = -1.000$, e todas as outras forças externas são nulas, resultando em

$$F_1 = -866 \qquad F_2 = -1.250 \qquad F_3 = 500$$
$$H_2 = 2.000 \qquad V_2 = 433 \qquad V_3 = -433$$

Os resultados indicam que os ventos têm efeitos marcadamente diferentes na estrutura. Ambos os casos estão descritos na Figura 12.6.

Os elementos individuais da matriz invertida também têm utilidade direta para elucidar interações resposta-estímulo para a estrutura. Cada elemento representa a mudança de uma das variáveis indeterminadas correspondentes a uma variação unitária de um dos estímulos externos. Por exemplo, o elemento a_{32}^{-1} indica que a terceira incógnita (F_3) vai mudar 0,866 por causa de uma variação unitária do segundo estímulo externo ($F_{1,v}$). Assim, se a carga vertical no primeiro nó crescer de uma unidade, F_3 irá crescer 0,866. O fato de que alguns elementos são zero indica que certas variáveis não são afetadas por alguns dos estímulos externos. Por exemplo, $a_{13}^{-1} = 0$ significa que F_1 não é afetada por mudanças em $F_{2,h}$. Essa habilidade de isolar interações tem um grande número de aplicações na engenharia, incluindo a identificação daqueles componentes que são mais sensíveis a estímulos externos e, como conseqüência, mais suscetíveis a falhas. Além disso, é usada para identificar componentes que podem ser desnecessários (ver Problema 12.18).

A abordagem anterior se torna particularmente útil quando aplicada a estruturas grandes e complexas. Na prática de engenharia, pode ser necessário resolver treliças com centenas ou mesmo milhares de membros estruturais. As equações lineares fornecem uma abordagem poderosa para se ganhar uma visão geral do comportamento dessas estruturas.

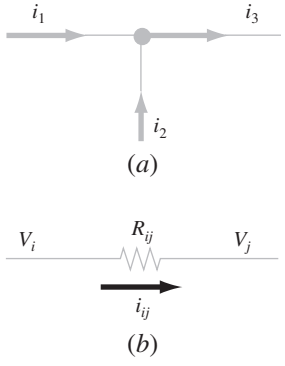
12.3 CORRENTES E VOLTAGENS EM CIRCUITOS DE RESISTORES (ENGENHARIA ELÉTRICA)

Fundamentos. Um problema comum na engenharia elétrica envolve a determinação das correntes e voltagens em várias posições em circuitos de resistores. Esses problemas são resolvidos usando-se as regras de Kirchhoff para correntes e voltagens. A regra da corrente (ou regra do ponto) estabelece que a soma algébrica de todas as correntes entrando em um nó deve ser zero (ver Figura 12.7*a*), ou

$$\Sigma i = 0 \tag{12.11}$$

onde todas as correntes entrando no nó são consideradas de sinal positivo. A regra da corrente é uma aplicação do princípio de conservação de carga (lembre-se da Tabela 1.1).

FIGURA 12.7
Representação esquemática (*a*) da regra de correntes de Kirchhoff e (*b*) da lei de Ohm.



A regra da voltagem (ou regra do laço) especifica que a soma algébrica das diferenças de potencial (isto é, variações de voltagem) em cada laço deve ser igual a zero. Para um circuito de resistores, isso é expresso por

$$\Sigma\xi - \Sigma i R = 0 \tag{12.12}$$

onde ξ é a fem (força eletromotriz) das fontes de voltagem e R é a resistência de cada um dos resistores do laço. Observe que o segundo termo vem da lei de Ohm (Figura 12.7*b*), que estabelece que a queda de voltagem através de um resistor ideal é igual ao produto da corrente pela resistência. A regra de voltagem de Kirchhoff é uma expressão da lei de conservação de energia.

Solução. A aplicação dessas regras resulta em um sistema de equações algébricas lineares simultâneas, pois os vários laços dentro de um circuito são acoplados. Por exemplo, considere o circuito mostrado na Figura 12.8. As correntes associadas a ele são desconhecidas tanto em módulo quanto em sentido. Isso não representa grande dificuldade, pois é possível simplesmente escolher um sentido para cada corrente. Se a solução resultante das leis de Kirchhoff for negativa, então o sentido foi escolhido incorretamente. A Figura 12.9 mostra algumas escolhas de correntes.

Dadas essas hipóteses, a regra de corrente de Kirchhoff é aplicada em cada nó para fornecer

$$i_{12} + i_{52} + i_{32} = 0$$

$$i_{65} - i_{52} - i_{54} = 0$$

$$i_{43} - i_{32} = 0$$

$$i_{54} - i_{43} = 0$$

A aplicação da regra de voltagem em cada um dos dois laços dá

$$-i_{54}R_{54} - i_{43}R_{43} - i_{32}R_{32} + i_{52}R_{52} = 0$$

$$-i_{65}R_{65} - i_{52}R_{52} + i_{12}R_{12} - 200 = 0$$

FIGURA 12.8
Um circuito de resistores a ser resolvido usando-se equações algébricas lineares simultâneas.

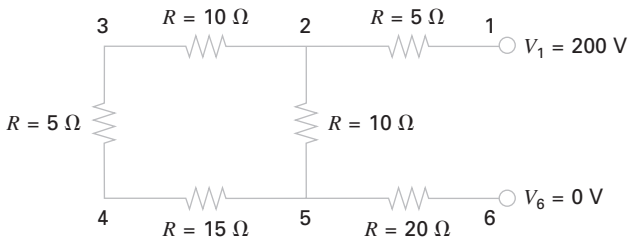
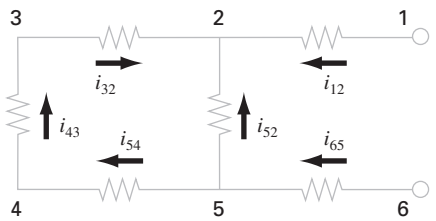


FIGURA 12.9
Escolha de correntes.



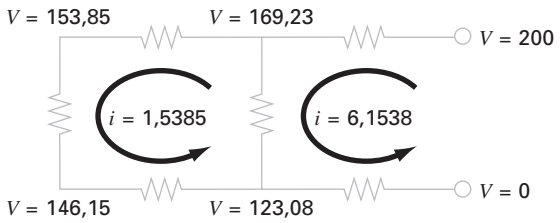


FIGURA 12.10
A solução para as correntes e voltagens obtidas usando-se um método de eliminação.

ou, substituindo as resistências da Figura 12.8 e trazendo as constantes para o lado direito,

$$-15i_{54} - 5i_{43} - 10i_{32} + 10i_{52} = 0$$

$$-20i_{65} - 10i_{52} + 5i_{12} = 200$$

Assim, o problema se transforma em resolver o seguinte conjunto de seis equações por seis correntes desconhecidas:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 10 & -10 & 0 & -15 & -5 \\ 5 & -10 & 0 & -20 & 0 & 0 \end{bmatrix} \begin{Bmatrix} i_{12} \\ i_{52} \\ i_{32} \\ i_{65} \\ i_{54} \\ i_{43} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 200 \end{Bmatrix}$$

Embora não seja prático resolvê-lo à mão, esse sistema é tratado facilmente usando-se um método de eliminação. Procedendo dessa forma, a solução é

$$i_{12} = 6,1538 \qquad i_{52} = -4,6154 \qquad i_{32} = -1,5385$$
$$i_{65} = -6,1538 \qquad i_{54} = -1,5385 \qquad i_{43} = -1,5385$$

Portanto, com a interpretação apropriada dos sinais do resultado, as correntes e voltagens do circuito são mostradas na Figura 12.10. As vantagens de usar algoritmos numéricos e computadores para problemas desse tipo devem ser evidentes.

12.4 SISTEMAS MASSA-MOLA (ENGENHARIA MECÂNICA/AEROESPACIAL)

Fundamentos. Sistemas idealizados de massa-mola desempenham um papel importante na mecânica e em outros problemas da engenharia. A Figura 12.11 mostra tal sistema. Assim que são soltas, as massas são puxadas para baixo pela força da gravidade. Observe que o deslocamento de cada mola na Figura 12.11*b* é medido em coordenadas locais relativas à sua posição inicial na Figura 12.11*a*.

Como introduzido no Capítulo 1, a segunda lei de Newton pode ser usada em conjunto com o balanço de forças para se desenvolver um modelo matemático do sistema. Para cada massa, a segunda lei pode ser expressa como

$$m\frac{d^2x}{dt^2} = F_D - F_U \tag{12.13}$$

Para simplificar as análises, suponha que todas as molas são idênticas e obedecem à lei de Hooke. Um diagrama de corpo livre para a primeira massa está esquematizado na Figura 12.12*a*. A força para cima é simplesmente a expressão direta da lei de Hooke:

$$F_U = kx_1 \tag{12.14}$$

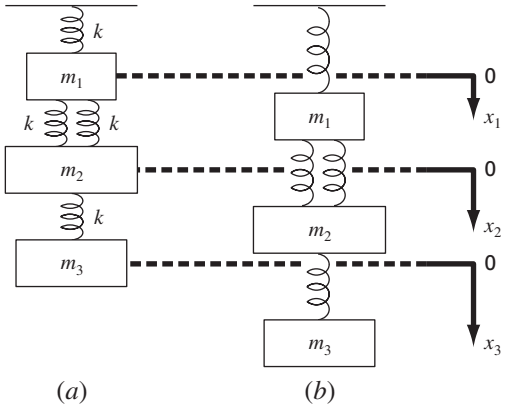


FIGURA 12.11
Um sistema composto por três massas suspensas verticalmente por uma série de molas. (*a*) O sistema antes de ser solto, isto é, antes da extensão ou compressão das molas. (*b*) O sistema depois de ser solto. Observe que as posições das massas são dadas em coordenadas locais com origens em suas posições antes de serem soltas.

A componente para baixo consiste na força das duas molas juntamente com a ação da gravidade na massa,

$$F_D = k(x_2 - x_1) + k(x_2 - x_1) = m_1 g \tag{12.15}$$

Observe como a componente de força das duas molas é proporcional ao deslocamento da segunda massa, x_2, corrigido pelo deslocamento da primeira massa, x_1.

As Equações (12.14) e (12.15) podem ser substituídas na Equação (12.13) para dar

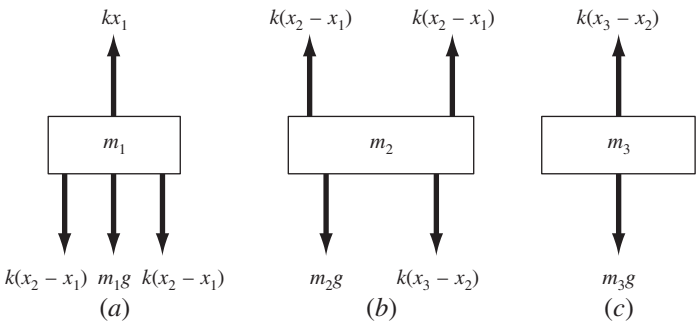
$$m_1 \frac{d^2 x_1}{dt^2} = 2k(x_2 - x_1) + m_1 g - kx_1 \tag{12.16}$$

Desse modo, deduz-se uma equação diferencial ordinária de segunda ordem para descrever o deslocamento da primeira massa com relação ao tempo. No entanto, observe que a solução não pode ser obtida, pois o modelo inclui uma segunda variável dependente, x_2. Conseqüentemente, diagramas de corpo livre devem ser desenvolvidos para a segunda e a terceira massas (Figuras 12.12*b* e *c*) que podem ser usados para deduzir

$$m_2 \frac{d^2 x_2}{dt^2} = k(x_3 - x_2) + m_2 g - 2k(x_2 - x_1) \tag{12.17}$$

e

FIGURA 12.12
Diagramas de corpo livre para as três massas da Figura 12.11.



$$m_3\frac{d^2x_3}{dt^2} = m_3g - k(x_3 - x_2) \tag{12.18}$$

As Equações (12.16), (12.17) e (12.18) formam um sistema de três equações diferenciais com três incógnitas. Com as condições iniciais apropriadas, elas podem ser usadas para determinar o deslocamento das massas como função do tempo (isto é, suas oscilações). Métodos numéricos de obter tais soluções serão discutidos na Parte Sete. No momento, é possível obter os deslocamentos que ocorrem quando o sistema eventualmente estiver em repouso, isto é, no estado estacionário. Para fazer isso, as derivadas nas Equações (12.16), (12.17) e (12.18) são igualadas a zero e obtém-se

$$\begin{aligned} 3kx_1 &- 2kx_2 &&&= m_1g\\ -2kx_1 &+ 3kx_2 &- kx_3 &= m_2g\\ &- kx_2 &+ kx_3 &= m_3g \end{aligned}$$

ou, em forma matricial,

$$[K]\{X\} = \{W\}$$

onde $[K]$, chamada de *matriz de rigidez*, é

$$[K] = \begin{bmatrix} 3k & -2k & \\ -2k & 3k & -k \\ & -k & k \end{bmatrix}$$

e $\{X\}$ e $\{W\}$ são os vetores coluna das incógnitas X e dos pesos mg, respectivamente.

Solução. Nesse ponto, os métodos numéricos podem ser usados para se obter uma solução. Se $m_1 = 2$ kg, $m_2 = 3$ kg, $m_3 = 2{,}5$ kg e $k = 10$ kg/s^2, use a decomposição LU para determinar os deslocamentos e gerar a inversa de $[K]$.

Substituindo os parâmetros do modelo, tem-se

$$[K] = \begin{bmatrix} 30 & -20 & \\ -20 & 30 & -10 \\ & -10 & 10 \end{bmatrix} \qquad \{W\} = \begin{Bmatrix} 19{,}6 \\ 29{,}4 \\ 24{,}5 \end{Bmatrix}$$

A decomposição LU pode ser usada para determinar $x_1 = 7{,}35$, $x_2 = 10{,}045$ e $x_3 = 12{,}495$. Esses deslocamentos foram utilizados para construir a Figura 12.11*d*. A inversa dessa matriz de rigidez é calculada por

$$[K]^{-1} = \begin{bmatrix} 0{,}1 & 0{,}1 & 0{,}1 \\ 0{,}1 & 0{,}15 & 0{,}15 \\ 0{,}1 & 0{,}15 & 0{,}25 \end{bmatrix}$$

Cada elemento dessa matriz, k_{ji}^{-1}, mostra o deslocamento da massa i decorrente de uma força unitária aplicada sobre a massa j. Assim, os valores 0,1 na coluna 1 mostram que uma carga unitária aplicada para baixo na primeira massa vai deslocar todas as outras massas 0,1 m para baixo. Os outros elementos podem ser interpretados de maneira similar. Então, a inversa da matriz de rigidez fornece um resumo fundamental de como as componentes do sistema respondem a forças aplicadas externamente.

PROBLEMAS

Engenharia Química/Bioengenharia

12.1 Efetue os mesmos cálculos da Seção 12.1, mas troque c_{01} para 40 e c_{03} para 10. Mude também os seguintes fluxos: $Q_{01} = 6$, $Q_{12} = 4$, $Q_{24} = 2$ e $Q_{44} = 12$.

12.2 Se a entrada no reator 3 na Seção 12.1 for diminuída de 25%, use a inversa da matriz para calcular a mudança porcentual na concentração dos reatores 1 e 4.

12.3 Já que o sistema mostrado na Figura 12.3 está no estado estacionário, o que pode ser dito a respeito dos quatro fluxos: Q_{01}, Q_{03}, Q_{44} e Q_{55}?

12.4 Recalcule as concentrações para os cinco reatores mostrados na Figura 12.3, se os fluxos forem trocados para

$$Q_{01} = 5 \quad Q_{31} = 3 \quad Q_{25} = 2 \quad Q_{23} = 2$$
$$Q_{15} = 4 \quad Q_{55} = 3 \quad Q_{54} = 3 \quad Q_{34} = 7$$
$$Q_{12} = 4 \quad Q_{03} = 8 \quad Q_{24} = 0 \quad Q_{44} = 10$$

12.5 Resolva o mesmo sistema do Problema 12.4, mas faça $Q_{12} = Q_{54} = 0$ e $Q_{15} = Q_{34} = 3$. Suponha que os fluxos internos (Q_{01}, Q_{03}) e os externos (Q_{44}, Q_{55}) sejam os mesmos. Use a conservação de fluxo para recalcular os valores dos outros fluxos.

12.6 A Figura P12.6 mostra três reatores interligados por tubos. Como indicado, a taxa de transferência de produtos químicos através de cada tubo é igual a uma vazão (Q, com unidade de metros cúbicos por segundo) multiplicada pela concentração no reator do qual o fluxo se origina (c, com unidades em miligramas por metro cúbico). Se o sistema estiver em um estado estacionário, a transferência para dentro de cada reator vai balancear a transferência para fora. Deduza equações de balanço de massa para os reatores e resolva as três equações lineares algébricas simultâneas para suas concentrações.

12.7 Empregando a mesma abordagem básica da Seção 12.1, determine a concentração de cloreto em cada um dos Grandes Lagos usando a informação mostrada na Figura P12.7.

12.8 O Baixo Rio Colorado consiste em uma série de quatro reservatórios, como mostrado na Figura P12.8. Os balanços de massas podem ser escritos para cada reservatório e o seguinte conjunto de equações algébricas lineares simultâneas é obtido:

$$\begin{bmatrix} 13{,}422 & 0 & 0 & 0 \\ -13{,}422 & 12{,}252 & 0 & 0 \\ 0 & -12{,}252 & 12{,}377 & 0 \\ 0 & 0 & -12{,}377 & 11{,}797 \end{bmatrix} \begin{Bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{Bmatrix} = \begin{Bmatrix} 750{,}5 \\ 300 \\ 102 \\ 30 \end{Bmatrix}$$

onde o vetor à direita consiste nas cargas de cloreto em cada um dos quatro lagos e c_1, c_2, c_3 e c_4 são iguais à concentração resultante de cloreto nos lagos Powell, Mead, Mohave e Havasu, respectivamente.

(a) Use a inversão de matrizes para determinar a concentração em cada um dos quatro lagos.

(b) Em quanto deve ser reduzida a carga no lago Powell para que a concentração de cloreto no lago Havasu seja 75?

(c) Usando a norma da soma das colunas, calcule o número de condicionamento e quantos algarismos suspeitos serão gerados ao resolver esse sistema.

Figura P12.6
Três reatores interligados por tubos. A taxa de transferência de massa através de cada tubo é igual ao produto da vazão Q pela concentração c do reator do qual o fluxo se origina.

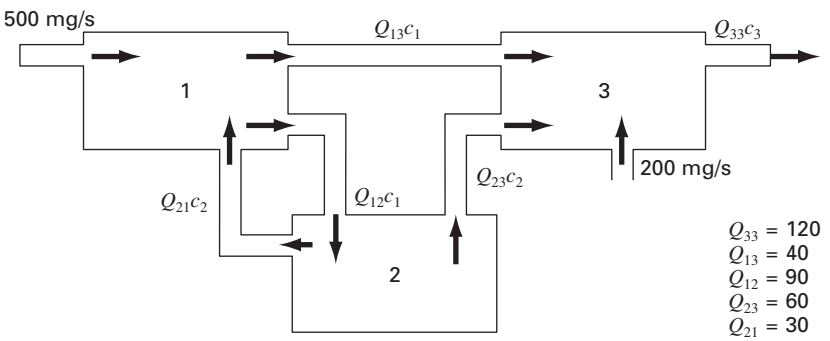
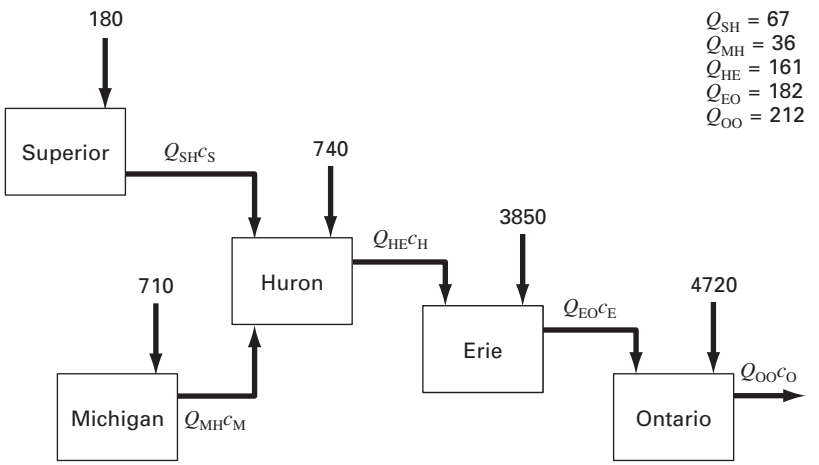


Figura P12.7
Um balanço de cloreto para os Grandes Lagos. As flechas numeradas são entradas diretas.



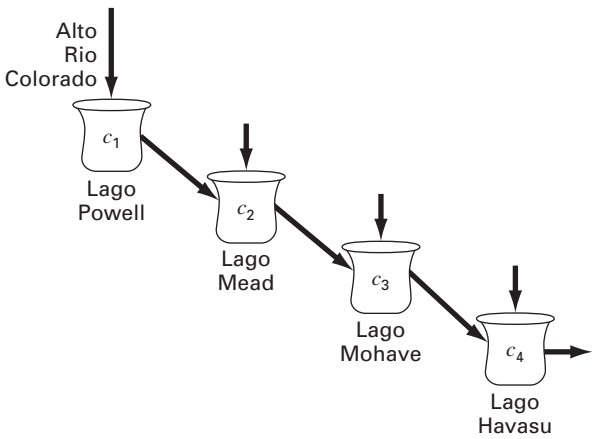


Figura P12.8
O Baixo Rio Colorado.

12.9 Um processo de extração em estágios é descrito na figura P12.9. Em tais sistemas, uma corrente contendo uma fração de peso Y_{ent} de um produto químico entra a partir da esquerda a uma vazão em massa F_1. Simultaneamente, um solvente carregando uma fração de peso X_{ent} do mesmo produto químico entra pela direita a uma vazão F_2. Assim, no estágio i, o balanço de massas pode ser representado por

$$F_1 Y_{i-1} + F_2 X_{i+1} = F_1 Y_i + F_2 X_i \qquad \text{(P12.9}a\text{)}$$

Em cada estágio, supõe-se que foi estabelecido um equilíbrio entre Y_i e X_i como em

$$K = \frac{X_i}{Y_i} \qquad \text{(P12.9}b\text{)}$$

onde K é chamado de coeficiente de distribuição. Pode-se isolar os X_i na Equação (P12.9b) e substituí-los na Equação (P12.9a) para obter:

$$Y_{i-1} - \left(1 + \frac{F_2}{F_1}K\right) Y_i + \left(\frac{F_2}{F_1}K\right) Y_{i+1} = 0 \qquad \text{(P12.9}c\text{)}$$

Se $F_1 = 500$ kg/h, $Y_{ent} = 0{,}1$, $F_2 = 1.000$ kg/h, $X_{ent} = 0$ e $K = 4$, determine os valores de $Y_{saída}$ e $X_{saída}$ se um reator de cinco estágios for usado. Observe que a Equação (P12.9c) deve ser modificada para levar em conta as frações de peso dos fluxos de entrada quando aplicada ao primeiro e ao último estágio.

12.10 Uma reação de primeira ordem, irreversível (ver Seção 28.1) acontece em quatro reatores bem misturados (Figura P12.10).

$$A \xrightarrow{k} B$$

Assim, a taxa na qual A é transformado em B pode ser representada como

$$R_{ab} = kVc$$

Os reatores têm diferentes volumes e, como são operados a diferentes temperaturas, cada um tem uma taxa de reação diferente:

Reator	V, L	k, h^{-1}
1	25	0,075
2	75	0,15
3	100	0,4
4	25	0,1

Determine a concentração de A e B em cada um dos reatores no estado estacionário.

12.11 Uma bomba peristáltica produz um escoamento unitário (Q_1) de um fluido altamente viscoso. A rede está descrita na Figura P12.11. Cada seção de tubo tem os mesmos comprimento e diâmetro. Os balanços de massa e energia mecânica podem ser simplificados para se obter os escoamentos em cada tubo. Resolva o sistema de equações a seguir para obter o escoamento de cada corrente.

$$Q_3 + 2Q_4 - 2Q_2 = 0$$
$$Q_5 + 2Q_6 - 2Q_4 = 0$$
$$3Q_7 - 2Q_6 = 0$$
$$Q_1 = Q_2 + Q_3$$
$$Q_3 = Q_4 + Q_5$$
$$Q_5 = Q_6 + Q_7$$

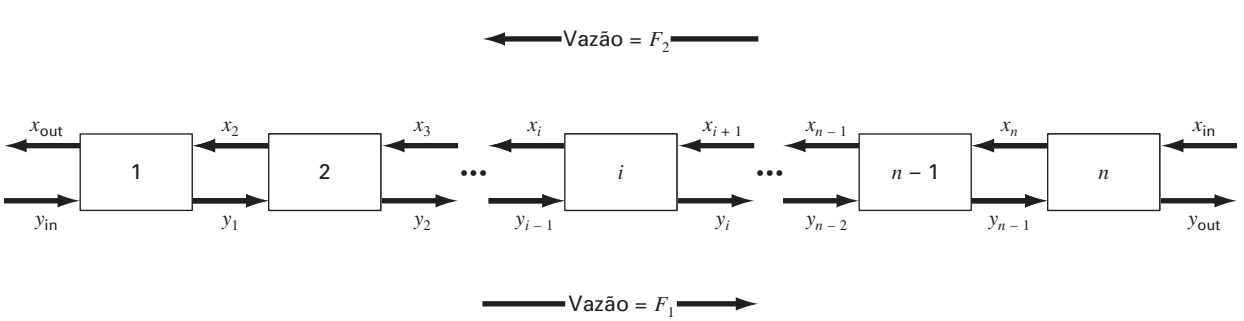
12.12 A Figura P12.12 mostra um processo de trocas químicas consistindo em uma série de reatores nos quais um gás escoando da esquerda para a direita passa sobre um líquido escoando da direita para a esquerda. A transferência de um produto químico do gás para o líquido se dá a uma taxa proporcional à diferença de concentrações entre o gás e o líquido em cada reator. No estado estacionário, um balanço de massas para o primeiro reator pode ser escrito para o gás como

$$Q_G c_{G0} - Q_G c_{G1} + D(c_{L1} - c_{G1}) = 0$$

e para o líquido como

$$Q_L c_{L2} - Q_L c_{L1} + D(c_{G1} - c_{L1}) = 0$$

Figura P12.9
Um processo de extração em estágios.



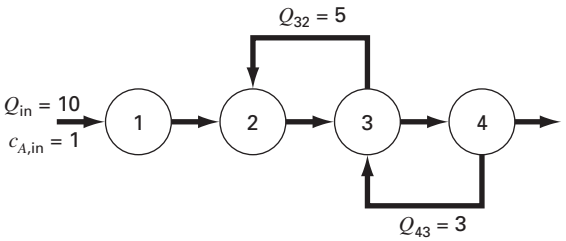


Figura P12.10

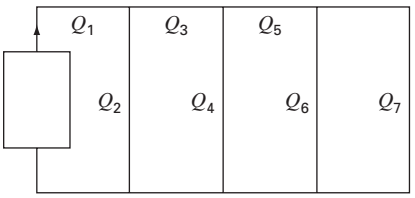


Figura P12.11

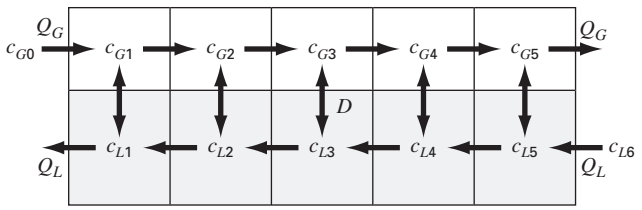


Figura P12.12

onde Q_G e Q_L são as vazões do gás e do líquido, respectivamente, e D é a taxa de troca entre gás e líquido. Balanços semelhantes podem ser escritos para os outros reatores. Determine as concentrações para os valores a seguir: $Q_G = 2$, $Q_L = 1$, $D = 0{,}8$, $c_{G0} = 100$, $c_{L6} = 10$.

Engenharia Civil/Ambiental

12.13 Um engenheiro civil envolvido em uma construção precisa de 4.800, 5.800 e 5.700 m^3 de areia, cascalho fino e cascalho grosso, respectivamente, para terminar a construção. Existem três minas de onde esses materiais podem ser obtidos. A composição dessas minas é:

	Areia %	Cascalho fino %	Cascalho grosso %
Mina 1	55	30	15
Mina 2	25	45	30
Mina 3	25	20	55

Quantos metros cúbicos devem ser minerados de cada mina para atender as necessidades do engenheiro?

12.14 Efetue os mesmos cálculos da Seção 12.2, mas para a treliça descrita na Figura P12.14.

12.15 Efetue os mesmos cálculos da Seção 12.2, mas para a treliça descrita na Figura P12.15.

12.16 Calcule as forças e as reações na treliça da Figura 12.4 se uma força para baixo de 2.500 kg e uma força para a direita de 2.000 kg forem aplicadas ao nó 1.

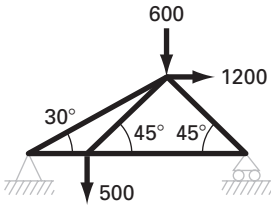


Figura P12.14

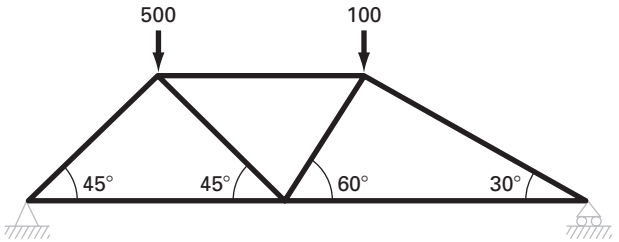


Figura P12.15

12.17 No exemplo da Figura 12.4, em que uma força de 1.000 lb foi aplicada para baixo ao nó 1, as reações externas V_2 e V_3 foram calculadas. Mas, se os comprimentos dos membros da treliça fossem dados, seria possível calcular V_2 e V_3 utilizando o fato de que $V_2 + V_3$ deve ser igual a 1.000 e somando os torques em torno do nó 2. No entanto, como V_2 e V_3 são conhecidas, pode-se trabalhar ao contrário para determinar os comprimentos dos membros da treliça. Observe que, como há três comprimentos desconhecidos e apenas duas equações, é possível determinar apenas a relação entre os comprimentos. Determine essa relação.

12.18 Empregando os mesmos métodos usados para analisar a Figura 12.4, determine as forças e as reações para a treliça mostrada na Figura P12.18.

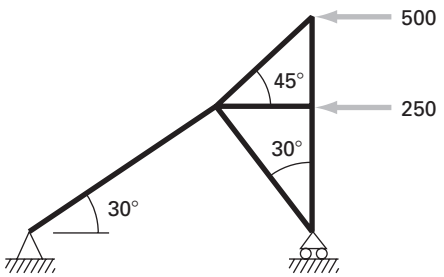
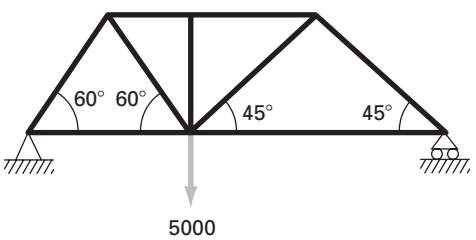


Figura P12.18

12.19 Determine as forças e as reações na treliça da Figura P12.19. Determine ainda a inversa da matriz para esse sistema. A força exercida pelo membro vertical no membro do meio parece razoável? Por quê?

Figura P12.19



12.20 Como o nome sugere, a poluição do ar de interiores refere-se à contaminação do ar em ambientes fechados, como casas, escritórios, áreas de serviço, etc. Suponha que você está projetando um sistema de ventilação para um restaurante como mostrado na Figura P12.20. As áreas de servir do restaurante consistem em duas salas quadradas e uma sala alongada. As salas 1 e 3 têm fontes de monóxido de carbono da área dos fumantes e de uma grelha defeituosa, respectivamente. O balanço de massa no estado estacionário pode ser escrito para cada área. Por exemplo, para a área dos fumantes (isto é, sala 1), o balanço pode ser escrito como

$$\begin{matrix} 0 = W_{\text{fumantes}} + & Q_a c_a & - & Q_a c_1 & + E_{13}(c_3 - c_1) \\ \text{(carga)} + & \text{(entrada)} & - & \text{(saída)} & + \quad \text{(mistura)} \end{matrix}$$

ou, substituindo os parâmetros,

$$225c_1 - 25c_3 = 1400$$

Balanços semelhantes podem ser escritos para as outras áreas.

(a) Determine a concentração no estado estacionário do monóxido de carbono em cada sala.

(b) Determine qual a percentagem de monóxido de carbono na área das crianças resultante (i) dos fumantes, (ii) da grelha e (iii) da entrada de ventilação.

(c) Se a carga dos fumantes e da grelha for aumentada para 2.000 e 5.000 mg/h, respectivamente, use a matriz inversa para determinar o crescimento da concentração na área das crianças.

(d) Como a concentração na área das crianças varia se uma tela for construída de modo que a mistura entre as áreas 2 e 4 decresça para 5 m^3/h?

12.21 Uma força para cima de 20 kN é aplicada no topo de um tripé, como descrito na Figura P12.21. Determine as forças nas pernas do tripé.

12.22 Uma treliça é submetida à carga como mostrado na Figura P12.22. Usando o seguinte conjunto de equações, determine as dez incógnitas: *AB*, *BC*, *AD*, *BD*, *CD*, *DE*, *CE*, A_x, A_y e E_y.

Figura P12.20
Vista superior dos ambientes de um restaurante. As flechas em uma única direção representam escoamento volumétrico de ar, enquanto as flechas em duplo sentido representam a mistura por difusão. As cargas dos fumantes e da grelha adicionam massa de monóxido de carbono ao sistema, mas escoamento de ar desprezível.

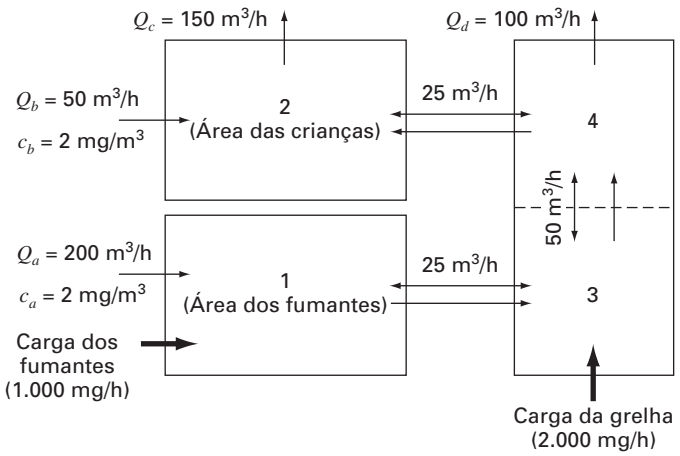
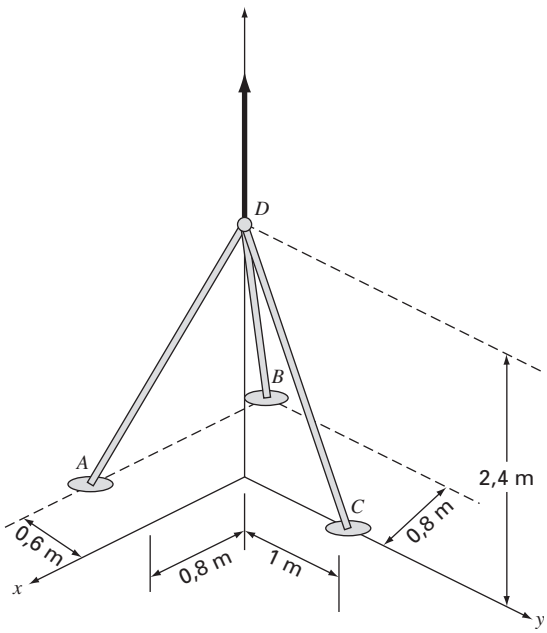


Figura P12.21



$$A_x + AD = 0 \qquad -24 - CD - (4/5)CE = 0$$
$$A_y + AB = 0 \qquad -AD + DE - (3/5)BD = 0$$
$$74 + BC + (3/5)BD = 0 \qquad CD + (4/5)BD = 0$$
$$-AB - (4/5)BD = 0 \qquad -DE - (3/5)CE = 0$$
$$-BC + (3/5)CE = 0 \qquad E_y + (4/5)CE = 0$$

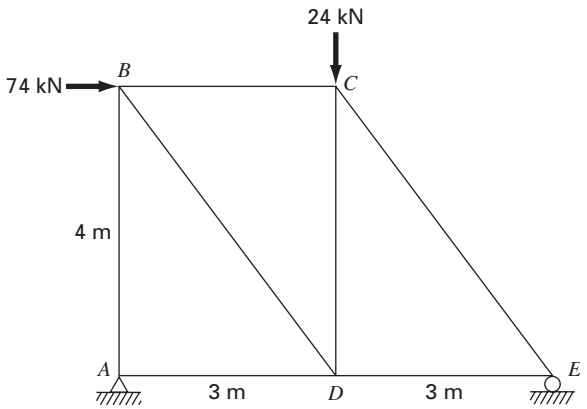


Figura P12.22

Engenharia Elétrica

12.23 Efetue os mesmos cálculos da Seção 12.3 para o circuito descrito na Figura P12.23.

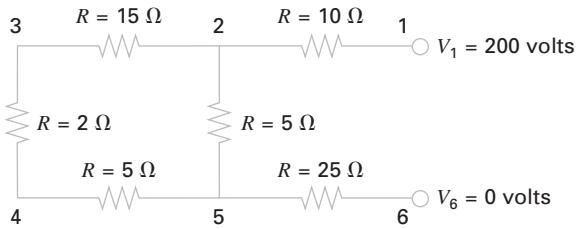


Figura P12.23

12.24 Efetue os mesmos cálculos da Seção 12.3 para o circuito descrito na Figura P12.24.

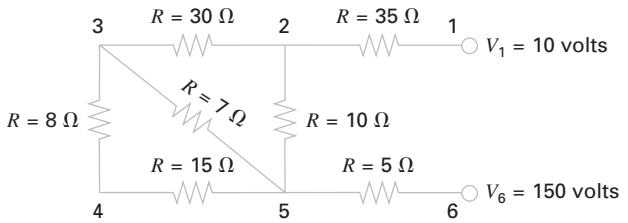


Figura P12.24

12.25 Resolva o circuito da Figura P12.25, determinando as correntes em cada fio. Use a eliminação de Gauss com pivotamento.

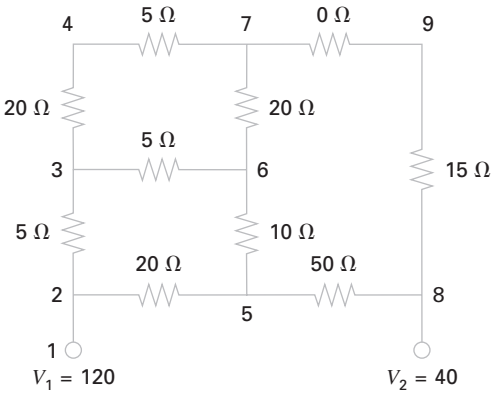


Figura P12.25

12.26 Um engenheiro elétrico supervisiona a produção de três tipos de componentes elétricos. Três tipos de material — metal, plástico e borracha — são necessários para a produção. As quantidades necessárias para a produção de cada componente são:

Componente	Metal, g/ por componente	Plástico, g/ por componente	Borracha, g/ por componente
1	15	0,30	1,0
2	17	0,40	1,2
3	19	0,55	1,5

Se um total de 3,89, 0,095 e 0,282 kg de metal, plástico e borracha, respectivamente, estiver disponível a cada dia, quantos componentes poderão ser produzidos por dia?

12.27 Determine as correntes para o circuito da Figura P12.27.

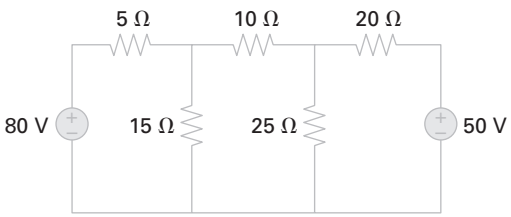
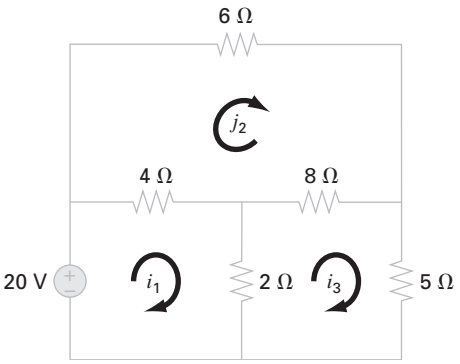


Figura P12.27

12.28 Determine as correntes para o circuito da Figura P12.28.

Figura P12.28



12.29 O sistema de equações a seguir foi gerado ao se aplicar a lei de correntes na malha ao circuito da Figura P12.29:

$$55I_1 - 25I_4 = -200$$
$$-37I_3 - 4I_4 = -250$$
$$-25I_1 - 4I_3 + 29I_4 = 100$$

Determine I_1, I_3 e I_4.

12.30 O sistema de equações a seguir foi gerado ao se aplicar a lei de correntes na malha ao circuito da Figura P12.30:

$$60I_1 - 40I_2 = 200$$
$$-40I_1 + 150I_2 - 100I_3 = 0$$
$$-100I_2 + 130I_3 = 230$$

Determine I_1, I_2 e I_3.

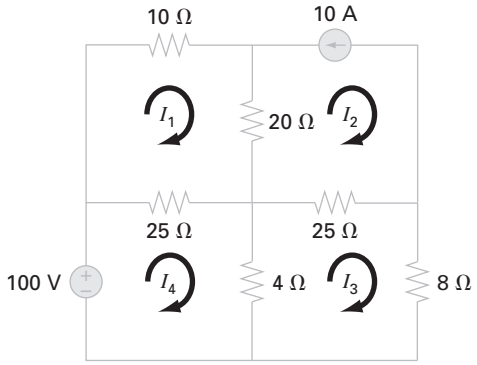


Figura P12.29

Engenharia Mecânica/Aeroespacial

12.31 Efetue os mesmos cálculos da Seção 12.4, mas adicione uma terceira mola entre as massas 1 e 2 e triplique k para todas as molas.

12.32 Efetue os mesmos cálculos da Seção 12.4, mas troque as massas de 2, 3 e 2,5 kg para 10, 3,5 e 2 kg, respectivamente.

12.33 Sistemas massa-mola idealizados têm numerosas aplicações em toda a engenharia. A Figura P12.33 mostra um arranjo de quatro molas em série sendo comprimidas por uma força de 1.500 kg. No equilíbrio, as equações de balanço das forças podem ser deduzidas escrevendo-se as interações entre as molas,

$$k_2(x_2 - x_1) = k_1 x_1$$
$$k_3(x_3 - x_2) = k_2(x_2 - x_1)$$
$$k_4(x_4 - x_3) = k_3(x_3 - x_2)$$
$$F = k_4(x_4 - x_3)$$

onde os k's são as constantes elásticas das molas. Se k_1 até k_4 forem 100, 50, 80 e 200 N/m, respectivamente, calcule os x's.

12.34 Três blocos são ligados por uma corda sem peso e repousam em um plano inclinado (Figura P12.34*a*). Empregar um procedimento similar àquele usado na análise dos pára-quedistas em queda-livre no Exemplo 9.11 fornece o seguinte conjunto de equações simultâneas (os diagramas de corpo livre são mostrados na Figura P12.34*b*):

$$100a + T = 519{,}72$$
$$50a - T + R = 216{,}55$$
$$25a - R = 108{,}27$$

Determine a aceleração a e as tensões T e R nas duas cordas.

12.35 Efetue cálculos semelhantes àqueles que foram usados no Problema 12.34 para o sistema mostrado na Figura P12.35.

Figura P12.30

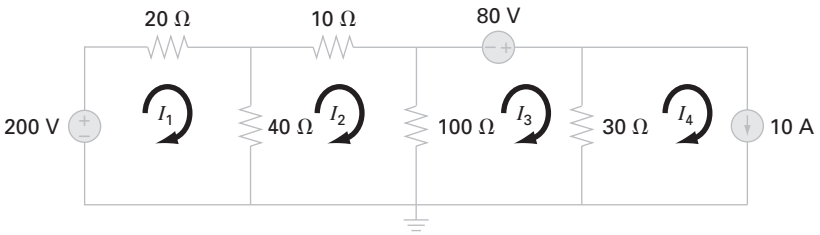
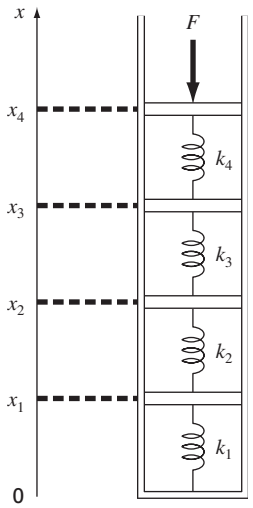


Figura P12.33



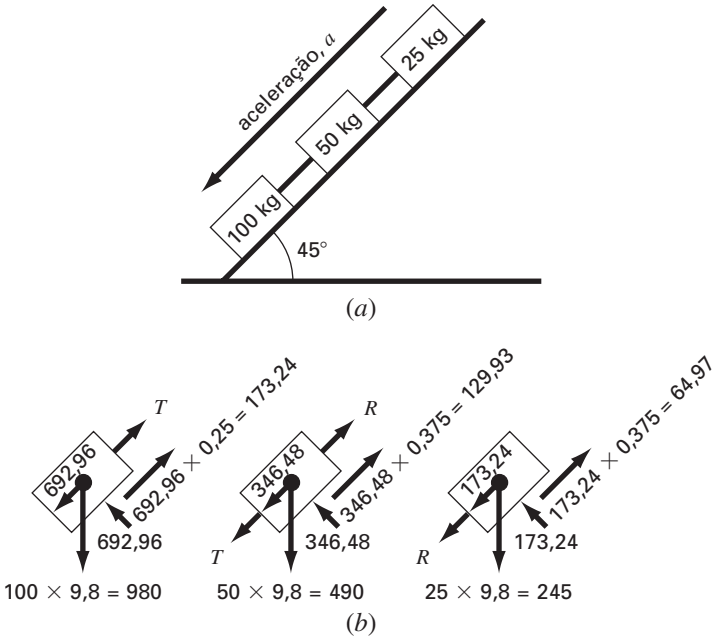


Figura P12.34

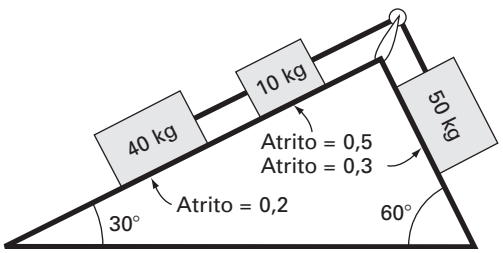


Figura P12.35

12.36 Efetue os mesmos cálculos que no Problema 12.34 para o sistema mostrado na Figura P12.36 (os ângulos são de 45º).

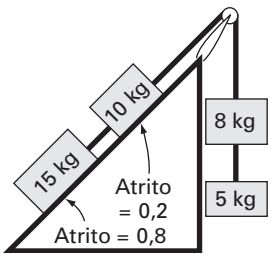


Figura P12.36

12.37 Considere o sistema de quatro massas e três molas da Figura P12.37. A determinação das equações de movimento de $\sum F_x = ma$, para cada massa, usando seu diagrama de corpo livre, resulta nas seguintes equações diferenciais:

$$\ddot{x}_1 + \left(\frac{k_1 + k_2}{m_1}\right) x_1 - \left(\frac{k_2}{m_1}\right) x_2 = 0$$

$$\ddot{x}_2 - \left(\frac{k_2}{m_2}\right) x_1 + \left(\frac{k_2 + k_3}{m_2}\right) x_2 - \left(\frac{k_3}{m_2}\right) x_3 = 0$$

$$\ddot{x}_3 - \left(\frac{k_3}{m_3}\right) x_2 + \left(\frac{k_3 + k_4}{m_3}\right) x_3 = 0$$

onde $k_1 = k_4 =$ 10 N/m, $k_2 = k_3 = 30$ N/m e $m_1 = m_2 = m_3 = m_4 = 2$ kg. Escreva as três equações na forma matricial:

0 = [Vetor aceleração] + [matriz k/m] [Vetor de deslocamento x]

Em um instante específico em que $x_1 = 0{,}05$ m, $x_2 = 0{,}04$ m e $x_3 = 0{,}03$ m, isso forma uma matriz tridiagonal. Determine a aceleração de cada massa.

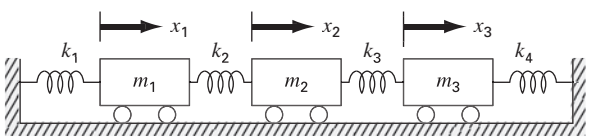


Figura P12.37

12.38 Equações algébricas lineares podem surgir na resolução de equações diferenciais. Por exemplo, a seguinte equação diferencial origina-se de um balanço de calor para um eixo longo e fino (Figura P12.38):

$$\frac{d^2T}{dx^2} + h'(T_a - T) = 0 \qquad \text{(P12.38)}$$

onde T é a temperatura (ºC), x é a distância ao longo do eixo (m), h' é o coeficiente de transferência de calor entre o eixo e o ambiente (m^{-2}) e T_a é a temperatura ambiente (ºC). Essa equação pode ser transformada em um conjunto de equações algébricas lineares usando-se uma aproximação por diferenças finitas divididas para a segunda derivada (lembre-se da Seção 4.1.3),

$$\frac{d^2T}{dx^2} = \frac{T_{i+1} - 2T_i + T_{i-1}}{\Delta x^2}$$

onde T_i designa a temperatura no nó i. Essa aproximação pode ser substituída na Equação (P12.38) para dar

$$-T_{i-1} + (2 + h'\Delta x^2)T_i - T_{i+1} = h'\Delta x^2 T_a$$

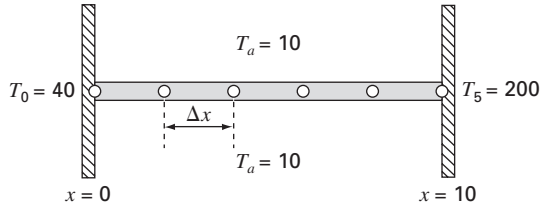


Figura P12.38
Um eixo uniforme não isolado posicionado entre duas paredes de temperaturas constantes, mas diferentes. A representação por diferenças finitas usa quatro nós interiores.

Essa equação pode ser escrita para cada um dos nós interiores do eixo, resultando em um sistema de equações tridiagonal. O primeiro e o último nós nas extremidades do eixo são fixados por condições de contorno.

(a) Determine uma solução analítica para a Equação (P12.38) para um eixo de 10 metros com $T_a = 20$, $T(x = 0) = 40$, $T(x = 10) = 200$ e $h' = 0{,}02$.

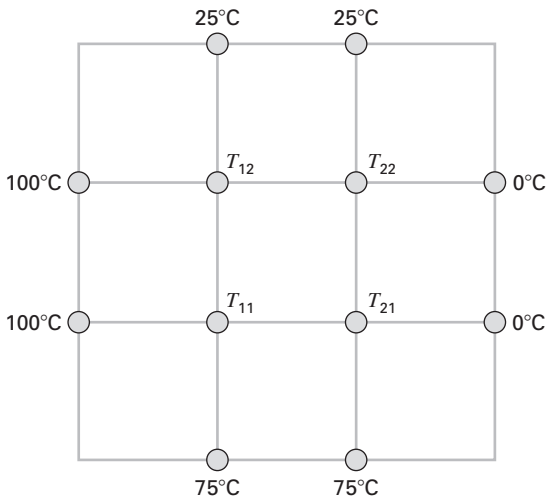
(b) Determine uma solução numérica para os mesmos valores dos parâmetros empregados em **(a)** usando uma solução por diferenças finitas com quatro nós interiores, como mostrado na Figura P12.38 ($\Delta x = 2$ m).

12.39 A distribuição estacionária de temperatura em uma placa aquecida pode ser modelada pela *equação de Laplace*.

$$0 = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}$$

Se a placa for representada por uma série de nós (Figura P12.39), pode-se substituir as segundas derivadas por diferenças divididas finitas centradas, o que resulta em um sistema de equações algébricas lineares. Use o método de Gauss-Seidel para determinar as temperaturas nos nós na Figura P12.39.

Figura P12.39

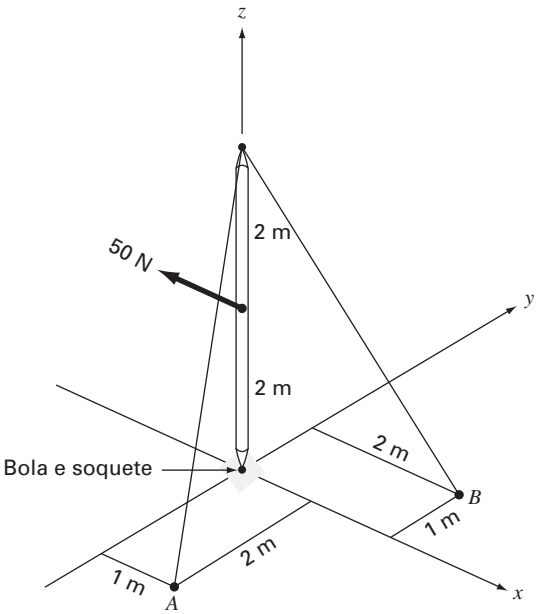


12.40 Um eixo sobre uma bola e um soquete está preso aos cabos *A* e *B* como na Figura 12.40.

(a) Se uma força de 50 N for exercida no eixo sem massa no ponto *G*, quais as forças de tensão nos cabos *A* e *B*?

(b) Determine as forças de reação na base do eixo. Chame esse ponto de *P*.

Figura P12.40





EPÍLOGO: PARTE TRÊS

PT3.4 PRÓS E CONTRAS

A Tabela PT3.2 fornece um resumo dos prós e contras envolvidos na resolução simultânea de equações algébricas lineares. Dois métodos — o método gráfico e a regra de Cramer — são limitados a um número pequeno (≤ 3) de equações e, portanto, têm pouca utilidade na resolução de problemas práticos. No entanto, essas técnicas são ferramentas didáticas úteis para entender o comportamento dos sistemas lineares em geral.

Os métodos numéricos propriamente ditos são divididos em duas categorias gerais: métodos exatos e métodos de aproximação. Como o nome indica, os primeiros são desenvolvidos para fornecer respostas exatas. No entanto, como são afetados por erros de arredondamento, esses métodos nos levam, às vezes, a resultados imprecisos. A ordem de grandeza dos erros de arredondamento varia de sistema para sistema e depende de vários fatores, os quais incluem a dimensão do sistema, seu condicionamento e se a matriz dos coeficientes é esparsa ou cheia. Além disso, a precisão computacional afetará os erros de arredondamento.

É recomendável que se use uma estratégia de pivotamento em qualquer programa de computador que implemente métodos exatos. A inclusão de tal estratégia minimiza erros de arredondamento e evita problemas como a divisão por zero. Se não houver outros fatores em jogo, os algoritmos baseados no método de decomposição *LU* são melhores por causa de sua eficiência e flexibilidade.

Apesar de os métodos de eliminação terem grande utilidade, seu uso na totalidade da matriz de coeficientes pode ser, um pouco, limitante quando se lida com sistemas muito grandes e esparsos. Isso se deve ao fato de que muita memória de computador será utilizada para armazenar zeros sem significado. Para sistemas de bandas, estão disponíveis técnicas para implementar métodos de eliminação, que não armazenam a matriz de coeficientes inteira.

TABELA PT3.2 Comparação das características de métodos alternativos para determinar soluções simultâneas de equações algébricas lineares.

Método	Estabilidade	Precisão	Aplicação	Esforço de Programação	Comentários
Gráfico	—	Ruim	Limitada	—	Pode levar mais tempo do que os métodos numéricos
Regra de Cramer	—	Afetada por erros de arredondamento	Limitada	—	É exigido esforço computacional excessivo para mais de três equações
Eliminação de Gauss (com pivotamento parcial)	—	Afetada por erros de arredondamento	Geral	Moderado	
Decomposição *LU*	—	Afetada por erros de arredondamento	Geral	Moderado	Método de eliminação preferencial; permite o cálculo da matriz inversa
Gauss-Seidel	Pode não convergir se não for diagonal dominante	Excelente	Apropriado apenas para sistemas diagonal dominantes	Fácil	

A técnica de aproximação descrita neste livro é chamada de método de Gauss-Seidel. Ela difere das técnicas exatas, pois emprega um esquema iterativo para obter estimativas progressivamente mais próximas da solução. Assim, o efeito de arredondamentos é um ponto discutível no método de Gauss-Seidel, já que podemos continuar as iterações enquanto for necessário até obter a precisão desejada. Além disso, podem ser desenvolvidas versões do método de Gauss-Seidel que permitam utilizar eficientemente as necessidades de armazenamento computacional dos sistemas esparsos. Conseqüentemente, a técnica de Gauss-Seidel é de utilidade para sistemas muito grandes de equações em que o armazenamento do sistema poderia tornar-se um problema significativo para as técnicas exatas.

A desvantagem do método de Gauss-Seidel é que ele nem sempre converge, ou, às vezes, converge lentamente para a solução verdadeira. Ele é estritamente confiável apenas para sistemas diagonal dominantes. No entanto, estão disponíveis métodos de relaxamento que podem, algumas vezes, contrabalançar essas desvantagens. Além disso, como muitos conjuntos de equações algébricas lineares oriundas de sistemas físicos exibem dominância diagonal, o método de Gauss-Seidel é de grande utilidade para resolver problemas de engenharia.

Em resumo, vários fatores podem influenciar sua escolha de uma técnica para a solução de um problema particular envolvendo equações algébricas lineares. No entanto, como delineado anteriormente, dois fatores particularmente importantes na determinação de sua escolha são o tamanho do sistema e quão esparso ele é.

PT3.5 RELAÇÕES E FÓRMULAS IMPORTANTES

Todas as partes deste livro incluem uma seção que resume as fórmulas importantes. Embora a Parte Três não lide com fórmulas únicas, a Tabela PT3.3 é utilizada para resumir os algoritmos estudados aqui. A tabela fornece uma visão geral que deveria ser útil para rever e esclarecer as principais diferenças entre os métodos.

PT3.6 MÉTODOS AVANÇADOS E REFERÊNCIAS ADICIONAIS

Referências gerais sobre a solução simultânea de equações lineares podem ser encontradas em Faddeev e Faddeeva (1963), Stewart (1973), Varga (1962) e Young (1971). Ralston e Rabinowitz (1978) fornecem um resumo geral.

Muitas técnicas avançadas estão disponíveis para aumentar a economia de tempo e/ou espaço durante a resolução de equações algébricas lineares. Muitas delas se concentram em explorar propriedades das equações, como simetria e matrizes de banda. Em particular, estão disponíveis algoritmos para se trabalhar com matrizes esparsas e transformá-las em um formato de banda mínimo. Jacobs (1977) e Tewarson (1973) incluem informações nessa área. Uma vez que estejam em um formato de banda mínimo, há uma variedade de estratégias de solução eficientes que são empregadas, como a abordagem de armazenamento de coluna ativa de Bathe e Wilson (1976).

Além de conjuntos de equações $n \times n$, existem outros sistemas nos quais o número de equações, m, e o número de incógnitas, n, não são iguais. Sistemas em que $m < n$ são chamados *subdeterminados*. Em tais casos, pode tanto não existir solução quanto existir mais de uma. Sistemas em que $m > n$ são chamados *sobredeterminados*. Para tais situações, em geral não existe uma solução exata. No entanto, muitas vezes é possível desenvolver uma solução de compromisso, que procura determinar as respostas que estão "mais próximas" de satisfazer todas as equações simultaneamente. Uma abordagem comum é resolver as equações no sentido dos "mínimos quadrados" (Lawson e Hanson, 1974; Wilkinson e Reinsch, 1971). De forma alternativa, os métodos de programação linear podem ser usados, nos casos em que as equações são resolvidas em um sentido "ótimo", minimizando-se alguma função objetivo (Dantzig, 1963; Luenberger, 1973; e Rabinowitz, 1968). Essa abordagem é descrita em detalhes no Capítulo 15.

TABELA PT3.3 Resumo das informações importantes apresentadas na Parte Três.

Método	Procedimento	Problemas em Potencial e Soluções
Eliminação de Gauss	$\left[\begin{array}{ccc\|c} a_{11} & a_{12} & a_{13} & c_1 \\ a_{21} & a_{22} & a_{23} & c_2 \\ a_{31} & a_{32} & a_{33} & c_3 \end{array}\right] \Rightarrow \left[\begin{array}{ccc\|c} a_{11} & a_{12} & a_{13} & c_1 \\ & a'_{22} & a'_{23} & c'_2 \\ & & a''_{33} & c''_3 \end{array}\right] \Rightarrow \begin{array}{l} x_3 = c''_3/a''_{33} \\ x_2 = (c'_2 - a'_{23}x_3)/a'_{22} \\ x_1 = (c_1 - a_{12}x_1 - a_{13}x_3)/a_{11} \end{array}$	Problemas: Mal condicionamento Arredondamento Divisão por zero Soluções: Maior precisão Pivotamento parcial
Decomposição *LU*	Decomposição; Substituição regressiva $\left[\begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array}\right] \Rightarrow \left[\begin{array}{ccc} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{array}\right] \left\{\begin{array}{c} d_1 \\ d_2 \\ d_3 \end{array}\right\} = \left\{\begin{array}{c} c_1 \\ c_2 \\ c_3 \end{array}\right\} \Rightarrow \left[\begin{array}{ccc} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{array}\right] \left\{\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right\} = \left\{\begin{array}{c} d_1 \\ d_2 \\ d_3 \end{array}\right\} \Rightarrow \left\{\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right\}$ Substituição progressiva	Problemas: Mal condicionamento Arredondamento Divisão por zero Soluções: Maior precisão Pivotamento parcial
Método de Gauss-Seidel	$\left.\begin{array}{l} x_1^j = (c_1 - a_{12}x_2^{j-1} - a_{13}x_3^{j-1})/a_{11} \\ x_2^j = (c_2 - a_{21}x_1^j - a_{23}x_3^{j-1})/a_{22} \\ x_3^j = (c_3 - a_{31}x_1^j - a_{32}x_2^j)/a_{33} \end{array}\right\}$ continue iterativamente até que $\left\|\frac{x_i^j - x_i^{j-1}}{x_i^j}\right\| 100\% < \epsilon_s$ para todos os x_is	Problemas: Divergência ou convergência lenta Soluções: Dominância diagonal Relaxamento

PARTE QUATRO



OTIMIZAÇÃO

PT4.1 MOTIVAÇÃO

A localização de raízes (Parte 2) e a otimização estão relacionadas no sentido que ambas envolvem a estimativa e a busca de um ponto na função. A diferença fundamental entre os dois tipos de problema é ilustrada na Figura PT4.1. A localização das raízes envolve a busca dos zeros de uma função ou funções. Em contraste, a otimização envolve a busca de um mínimo ou de um máximo.

O ponto ótimo é um ponto no qual a curva é achatada. Em termos matemáticos, isso corresponde ao valor de x onde a derivada $f'(x)$ é igual a zero. Adicionalmente, a segunda derivada, $f''(x)$, indica quando o ponto ótimo é um mínimo ou um máximo: se $f''(x) < 0$, o ponto é um máximo; se $f''(x) > 0$, o ponto é um mínimo.

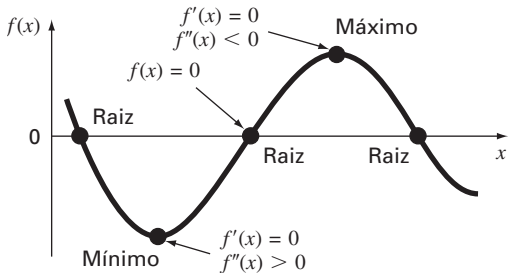
Agora, o entendimento da relação entre raízes e pontos ótimos sugere uma possível estratégia para achar os últimos. Isto é, você pode derivar a função e localizar a raiz (ou seja, o zero) da nova função. De fato, alguns métodos de otimização procuram o ponto ótimo resolvendo o problema de raiz: $f'(x) = 0$. Deve ser observado que tais buscas são geralmente complicadas porque $f'(x)$ não está disponível analiticamente. Portanto, algumas vezes são usadas aproximações por diferenças finitas para obter uma estimativa da derivada.

Além de se considerar a otimização como um problema de determinação de raízes, deve ser observado que a tarefa de localização de pontos ótimos é auxiliada por algumas estruturas matemáticas adicionais que não são parte da simples determinação de raízes. Isso tende a fazer que a otimização seja um problema mais abordável, particularmente para o caso multidimensional.

PT4.1.1 Métodos Não-computacionais e História

Como mencionado, os métodos do cálculo diferencial ainda são utilizados para determinar soluções ótimas. Todos os estudantes de engenharia e ciências lembram-se de como trabalhar nos problemas de máximos e mínimos determinando a primeira derivada das funções dos seus cursos de cálculo. Bernoulli, Euler, Lagrange e outros assentaram os fundamentos do cálculo das variações, que trata da minimização de funções. O método dos multiplicadores de Lagrange foi desenvolvido para otimizar problemas com restrições (ou vinculados), isto é, otimizar problemas nos quais as variáveis são limitadas de alguma forma.

FIGURA PT4.1
Uma função de uma única variável ilustrando a diferença entre raízes e pontos ótimos.



O primeiro grande avanço nas abordagens numéricas ocorreu apenas com o desenvolvimento dos computadores digitais depois da Segunda Guerra Mundial. Koopmans no Reino Unido e Kantorovich, na antiga União Soviética, trabalharam independentemente no problema geral da distribuição de custo mínimo de suprimentos e produtos. Em 1947, Dantzig, um aluno de Koopmans, inventou o *procedimento simplex* para resolver problemas de programação linear. Essa abordagem pavimentou o caminho para outros métodos para otimização restrita por diversos pesquisadores, especialmente Charnes e seus colaboradores. Abordagens para otimização irrestrita também se desenvolveram rapidamente, seguindo a ampla disponibilidade dos computadores.

PT4.1.2 Otimização e a Prática da Engenharia

Muitos dos modelos com os quais lidamos até aqui são modelos *descritivos*. Isto é, foram desenvolvidos para simular o comportamento de um artefato de engenharia ou sistema. Em contraste, a otimização tipicamente lida com a busca pelo "melhor resultado", ou com a solução ótima de um problema. Portanto, no contexto da modelagem, eles são denominados modelos *prescritivos*, uma vez que podem ser utilizados para prescrever uma linha de ação ou o melhor projeto.

Os engenheiros devem sempre projetar artefatos e produtos que desempenhem tarefas de forma eficiente. Ao fazer isso, eles estão restritos pelas limitações do mundo físico. Além disso, devem manter os custos baixos. Portanto, estão sempre confrontando problemas de otimização que equilibram eficiência e limitações. Alguns casos comuns estão listados na Tabela PT4.1. O exemplo a seguir foi desenvolvido para ajudar você a perceber a maneira como tais problemas podem ser formulados.

TABELA PT4.1 Alguns exemplos comuns de problemas de otimização em engenharia.

- Projeto de uma aeronave com peso mínimo e resistência máxima.
- Trajetória ótima de veículos espaciais.
- Projeto de estruturas de engenharia civil com custo mínimo.
- Projeto de reservatórios de água como represas para diminuir os danos de enchentes e ao mesmo tempo fornecendo a máxima potência hídrica.
- Previsão do comportamento de estruturas pela minimização da energia potencial.
- Estratégia de corte de material com custo mínimo.
- Projeto de bombas e equipamento de transferência de calor para eficiência máxima.
- Maximização da potência de saída de redes elétricas e de máquinas, ao mesmo tempo minimizando a geração de calor.
- Rota mais curta de um caixeiro-viajante visitando várias cidades em uma viagem de vendas.
- Planejamento e escalonamento ótimos.
- Análise estatística e modelos com erro mínimo.
- Redes de tubulação ótimas.
- Controle de estoque.
- Planejamento de manutenção para minimizar custos.
- Minimização de tempo de espera e de tempo ocioso.
- Projeto de sistemas de tratamento de dejetos para atingir padrões de qualidade de água com custo mínimo.

EXEMPLO PT4.1 Otimização do Custo do Pára-Quedas

Enunciado do Problema. Nos capítulos precedentes deste livro, foi usado o problema do pára-quedista em queda livre para ilustrar as áreas básicas de problemas dos métodos numéricos. Você deve ter observado que nenhum desses exemplos se concentra no que acontece depois que o pára-quedas se abre. Neste exemplo será examinado o caso em que o pára-quedas se abre e estamos interessados em prever a velocidade de impacto no solo.

Você é um engenheiro que trabalha para uma agência que planeja fornecer suprimentos, por via aérea, a refugiados em uma zona de guerra. Os suprimentos devem ser jogados de uma altitude baixa (500 m), de modo que a queda não seja detectada e que os supri-

mentos caiam tão perto quanto possível do campo de refugiados. O pára-quedas se abre imediatamente após ter saído do avião. Para reduzir os estragos, a velocidade vertical no impacto deve ser menor que um valor crítico $v_c = 20$ m/s.

O pára-quedas usado para a queda é mostrado na Figura PT4.2. A área da seção transversal do pára-quedas é o da semi-esfera,

$$A = 2\pi r^2 \tag{PT4.1}$$

O comprimento de cada uma das 16 cordas que conectam o pára-quedas à massa está relacionado com o raio do pára-quedas por

$$\ell = \sqrt{2}r \tag{PT4.2}$$

Você sabe que a força de arrasto do pára-quedas é uma função linear da área da seção transversal, descrita pela seguinte fórmula

$$c = k_c A \tag{PT4.3}$$

onde c é o coeficiente de arrasto (kg/s) e k_c é a constante de proporcionalidade que parametriza o efeito da área no arrasto [kg/(s · m^2)].

Além disso, você pode dividir a carga em tantos pacotes quantos quiser. Isto é, a massa de cada pacote individual pode ser calculada como

$$m = \frac{M_t}{n}$$

onde m é a massa de um pacote individual (kg), M_t é a carga total jogada (kg) e n é o número total de pacotes.

Finalmente, o custo de cada pára-quedas está relacionado ao tamanho do pára-quedas de forma não-linear,

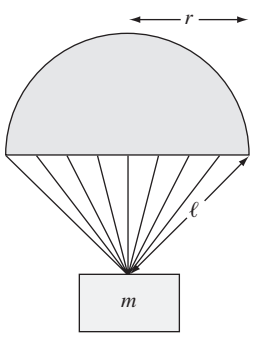
$$\text{Custo por pára-quedas} = c_0 + c_1\ell + c_2A^2 \tag{PT4.4}$$

onde c_0, c_1 e c_2 são os coeficientes de custo. O termo constante, c_0, é o preço-base dos pára-quedas. A relação não-linear entre custo e área existe porque pára-quedas grandes são muito mais difíceis de construir do que pára-quedas pequenos.

Determine o tamanho (r) e o número de pára-quedas (n) que resultam em um custo mínimo, ao mesmo tempo que satisfazem o requisito de ter uma velocidade de impacto suficientemente pequena.

Solução. O objetivo aqui é determinar o número e o tamanho dos pára-quedas para minimizar o custo do socorro aéreo. O problema tem restrições porque os pacotes devem ter uma velocidade de impacto menor que o valor crítico.

FIGURA PT4.2
Um pára-quedas aberto.



O custo pode ser calculado multiplicando-se o custo individual de cada pára-quedas [Eq. (PT4.4)] pelo número de pára-quedas (n). Portanto, a função que você quer minimizar, formalmente chamada de *função objetivo*, é escrita como

$$\text{Minimize } C = n(c_0 + c_1\ell + c_2A^2) \tag{PT4.5}$$

onde C é o custo (\$) e A e ℓ são calculados pelas Equações (PT4.1) e (PT4.2), respectivamente.

Em seguida, deve-se especificar as *restrições*. Para esse problema existem duas *restrições*. Primeiro, a velocidade de impacto deve ser menor ou igual à velocidade crítica,

$$v \leq v_c \tag{PT4.6}$$

Segundo, o número de pacotes deve ser um inteiro maior ou igual a 1,

$$n \geq 1 \tag{PT4.7}$$

onde n é um inteiro.

A essa altura, o problema de otimização está formulado. Como pode ser visto, é um problema não-linear com restrições.

Embora o problema tenha sido formulado de forma ampla, ainda há um aspecto a ser abordado: como determinaremos a velocidade de impacto v? Lembre-se, do Capítulo 1, de que a velocidade de um corpo em queda pode ser calculada como

$$v = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{1.10}$$

onde v é a velocidade (m/s), g é a aceleração da gravidade (m/s^2), m é a massa (kg) e t é o tempo (s).

Embora a Equação (1.10) forneça uma relação entre v e t, é preciso saber por quanto tempo a massa cai. Assim, é necessária uma relação entre a distância de queda z e o tempo de queda t. A distância de queda pode ser calculada a partir da velocidade na Equação (1.10) por integração,

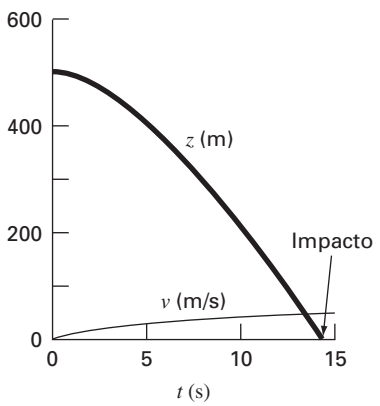
$$z = \int_0^t \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) dt \tag{PT4.8}$$

Essa integral pode ser calculada, fornecendo

$$z = z_0 - \frac{gm}{c}t + \frac{gm^2}{c^2}\left(1 - e^{-(c/m)t}\right) \tag{PT4.9}$$

onde z_0 é a altura inicial (m). Essa função, mostrada na Figura PT4.3, fornece uma maneira de prever z a partir do conhecimento de t.

FIGURA PT4.3
A altura z e a velocidade v de um pára-quedas aberto a medida que cai até a terra ($z = 0$).



Entretanto, não é preciso ter z como função de t para resolver esse problema. Em vez disso, é necessário calcular o tempo necessário para o pacote cair a distância z_0. Assim, devemos reformular a Equação (PT4.9) como um problema de determinação de raízes. Isto é, devemos determinar o instante no qual z é igual a zero.

$$f(t) = 0 = z_0 - \frac{gm}{c}t + \frac{gm^2}{c^2}\left(1 - e^{-(c/m)t}\right) \tag{PT4.10}$$

Uma vez que o tempo de impacto seja calculado, pode-se substituí-lo na Equação (1.10) e determinar a velocidade de impacto.

Portanto, a especificação final do problema deveria ser

$$\text{Minimize } C = n(c_0 + c_1\ell + c_2A^2) \tag{PT4.11}$$

sujeito a

$$v \leq v_c \tag{PT4.12}$$

$$n \geq 1 \tag{PT4.13}$$

onde

$$A = 2\pi r^2 \tag{PT4.14}$$

$$\ell = \sqrt{2}r \tag{PT4.15}$$

$$c = k_c A \tag{PT4.16}$$

$$m = \frac{M_t}{n} \tag{PT4.17}$$

$$t = \text{raiz}\left[z_0 - \frac{gm}{c}t + \frac{gm^2}{c^2}\left(1 - e^{-(c/m)t}\right)\right] \tag{PT4.18}$$

$$v = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{PT4.19}$$

Esse problema será resolvido no Exemplo 15.4 no final do Capítulo 15. Por enquanto, perceba que ele tem a maioria dos elementos fundamentais de outros problemas de otimização que você encontrará rotineiramente na prática da engenharia. São eles:

- O problema envolverá uma *função objetivo,* a qual representa sua meta.
- Existirão algumas *variáveis de projeto*. Tais variáveis podem ser números reais ou podem ser inteiros. No exemplo, elas são r (real) e n (inteiro).
- O problema incluirá *restrições* que refletem as limitações sob as quais você está trabalhando.

Deve-se salientar mais um ponto antes de prosseguir. Embora, superficialmente, a função objetivo e as restrições pareçam ser equações simples [por exemplo a Equação (PT4.12)], elas podem de fato ser o "topo do *iceberg*". Isto é, elas podem ser fundamentadas por dependências e modelos complexos. Por exemplo, como no nosso caso, elas podem envolver outros métodos numéricos [Equação (PT4.18)]. Isso significa que as relações funcionais que você estará usando podem, na realidade, representar cálculos longos e complicados. Portanto, técnicas capazes de encontrar uma solução ótima, ao mesmo tempo que minimizam o número de cálculos de função, são extremamente valiosas.

PT4.2 FUNDAMENTOS MATEMÁTICOS

Há diversos conceitos e operações matemáticos que fundamentam a otimização. Como eles serão mais relevantes para você em contexto, a discussão desses pré-requisitos matemáticos será adiada até que eles sejam de fato necessários. Por exemplo, os conceitos

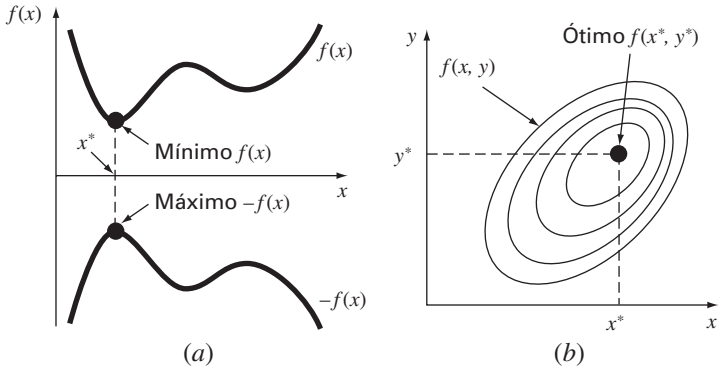


FIGURA PT4.4
(*a*) Otimização unidimensional. A figura também ilustra como a minimização de $f(x)$ é equivalente à maximização de $-f(x)$. (*b*) Otimização bidimensional. Observe que essa figura pode ser usada para representar ou a maximização (os contornos aumentam em elevação até o máximo, como em uma montanha) ou a minimização (os contornos decrescem em elevação até o mínimo, como em um vale).

importantes de gradiente e hessianas serão discutidos no início do Capítulo 14, na otimização irrestrita em muitas variáveis. Por enquanto, nos limitaremos aqui ao tópico mais geral de como os problemas de otimização são classificados.

Um problema de *otimização* ou *programação matemática* geralmente pode ser enunciado como:

Encontre $\boldsymbol{x}$, que minimiza ou maximiza $f(\boldsymbol{x})$
sujeito a

$$d_i(\boldsymbol{x}) \leq a_i \qquad i = 1, 2, \ldots, m \tag{PT4.20}$$

$$e_i(\boldsymbol{x}) = b_i \qquad i = 1, 2, \ldots, p \tag{PT4.21}$$

onde $\boldsymbol{x}$ é um *vetor de projeto n*-dimensional, $f(\boldsymbol{x})$ é a *função objetivo*, $d_i(\boldsymbol{x})$ são *restrições expressas por desigualdades*, $e_i(\boldsymbol{x})$ são *restrições expressas por igualdades* e a_i e b_i são constantes.

Problemas de otimização podem ser classificados com base na forma de $f(\boldsymbol{x})$:

- Se $f(\boldsymbol{x})$ e as restrições forem lineares, temos a *programação linear.*
- Se $f(\boldsymbol{x})$ for quadrática e as restrições forem lineares, temos a *programação quadrática.*
- Se $f(\boldsymbol{x})$ não for linear ou quadrática e/ou as restrições forem não-lineares, temos a *programação não-linear*.

Além disso, quando as Equações (PT4.20) e (PT4.21) são incluídas, temos problemas de *otimização com restrições*; caso contrário, temos problemas de *otimização sem restrições*.

Observe que, para os problemas com restrições, os graus de liberdade são dados por $n-p-m$. Geralmente, para obter uma solução, $p + m$ deve ser $\leq n$. Se $p + m > n$, o problema é dito *super-restrito*.

Outra forma de classificar os problemas de otimização é pela dimensão. Isso geralmente é feito dividindo-os em problemas unidimensionais e multidimensionais. Como o nome indica, os *problemas unidimensionais* envolvem funções que dependem de uma única variável dependente. Como na Figura PT4.4*a*, a busca, então, consiste em subir ou descer picos ou vales. Os *problemas multidimensionais* envolvem funções que dependem de duas ou mais variáveis dependentes. Seguindo o mesmo espírito, uma otimização bidimensional pode ser novamente visualizada como buscando picos ou vales (PT4.4*b*). Entretanto, do mesmo modo como em uma excursão real, não se está restrito a andar em uma única direção, em vez disso, a *topografia* deve ser examinada para se alcançar a meta de forma eficiente.

Finalmente, o processo de busca pelo máximo comparado com a busca pelo mínimo é basicamente idêntico, porque o mesmo valor $\boldsymbol{x}^*$ tanto minimiza $f(\boldsymbol{x})$ quanto maximiza $-f(\boldsymbol{x})$. Essa equivalência é ilustrada graficamente para uma função unidimensional na Figura PT4.4*a*.

PT4.3 ORIENTAÇÃO

Alguma orientação seria útil antes de prosseguir com os métodos numéricos para otimização. O que vem a seguir pretende dar uma visão geral de todo o material da Parte Quatro. Além disso, alguns objetivos foram incluídos para ajudar você a focalizar seus esforços quando estudar esse material.

PT4.3.1 Escopo e Visão Geral

A Figura PT4.5 é uma representação esquemática da organização da Parte Quatro. Examine-a cuidadosamente, começando pelo topo e trabalhando no sentido horário.

Depois dessa introdução, o **Capítulo 13** é destinado à *otimização unidimensional sem restrições*. Serão apresentados três métodos para encontrar o mínimo ou máximo de uma função de uma única variável: *busca da seção áurea*, *interpolação quadrática* e o *método de Newton*. Esses métodos também são relevantes para a otimização multidimensional.

O **Capítulo 14** cobre dois tipos de métodos gerais para resolver problemas de *otimização multidimensional sem restrições*. Os *métodos diretos*, como *busca aleatória*, *busca univariante* e *busca de padrões,* não requerem o cálculo da derivada da função. Por outro lado, os *métodos gradiente* usam a primeira e, algumas vezes, a segunda derivada para achar o ponto ótimo. Este capítulo introduz o *gradiente* e a *hessiana*, que são representações multidimensionais da primeira e da segunda derivadas. O método do aclive ou declive máximo será então coberto com algum detalhe. Em seguida, serão descritos alguns métodos avançados: *gradiente conjugado*, *método de Newton*, *método de Marquardt* e *métodos quasi-Newton*.

O **Capítulo 15** é dedicado à *otimização com restrições.* A *programação linear* é descrita em detalhes usando tanto a representação gráfica quanto o *método simplex*. Uma análise detalhada da *otimização não-linear com restrições* está além do escopo deste livro, mas será fornecida uma visão geral das principais abordagens. Adicionalmente, será ilustrado como tais problemas (juntamente com os problemas cobertos nos Capítulos 13 e 14) podem ser resolvidos com pacotes e bibliotecas de software como Excel, MATLAB e o IMSL.

O **Capítulo 16** estende os conceitos citados para problemas reais de engenharia. Aplicações de engenharia são usadas para ilustrar como problemas de otimização são formulados e para que se possa vislumbrar das aplicações técnicas de solução usadas na prática profissional.

Um epílogo foi incluído no final da Parte Quatro. Ele contém uma revisão dos métodos discutidos nos Capítulos 13, 14 e 15. Essa revisão inclui a descrição dos prós e contras associados ao uso de cada uma das técnicas. Essa seção também fornece referências de alguns métodos numéricos que estão além do escopo deste texto.

PT4.3.2 Metas e Objetivos

Objetivos de Estudo. Após completar a Parte Quatro, o estudante deve estar suficientemente informado para abordar com sucesso uma grande variedade de problemas em engenharia que envolvem otimização. Em geral, deve ter dominado as técnicas, aprendido a avaliar a confiabilidade delas e ser capaz de analisar métodos alternativos para um problema particular. Além dessas metas gerais, os conceitos específicos da Tabela PT4.2 devem ser assimilados para um entendimento completo do material da Parte Quatro.

Objetivos Computacionais. O aluno deve estar apto a escrever um subprograma para implementar uma busca unidimensional simples (como a busca da seção áurea ou interpolação quadrática) e uma busca multidimensional (como o método de busca aleatória). Além disso, bibliotecas de programa como o IMSL e pacotes de software como o Excel ou o MATLAB possuem recursos distintos para otimização. O estudante pode usar esta parte do livro para se familiarizar com esses recursos.

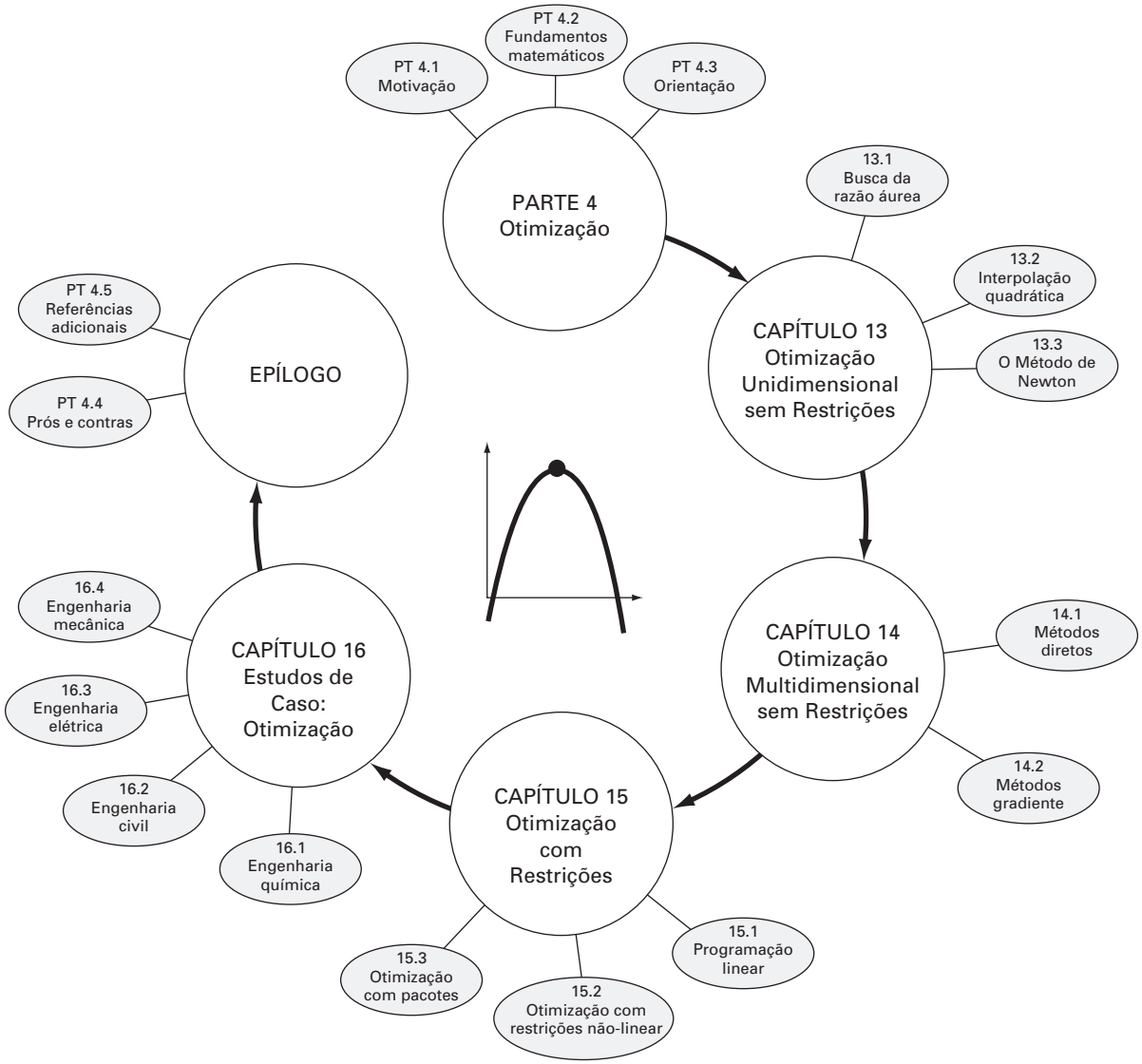


FIGURA PT4.5
Organização esquemática do material na Parte Quatro: Otimização

TABELA PT4.2 Objetivos de estudo específicos da Parte Quatro.

1. Entender onde e por que a otimização ocorre na resolução de problemas em engenharia.
2. Entender os elementos essenciais do problema geral de otimização: função objetivo, variáveis de decisão e restrições.
3. Ser capaz de distinguir entre otimização linear e não-linear e entre problemas com restrições e sem restrições.
4. Ser capaz de definir a razão áurea e entender como ela torna a otimização unidimensional eficiente.
5. Localizar o ponto ótimo de uma função de uma única variável com a busca da seção áurea, a interpolação quadrática ou o método de Newton. Além disso, reconhecer prós e contras entre essas abordagens, com atenção especial para as aproximações iniciais e a convergência.
6. Ser capaz de escrever um programa e determinar o ponto ótimo de uma função de várias variáveis usando busca aleatória.
7. Entender as idéias por trás da busca por padrões, das direções conjugadas e do método de Powel.
8. Ser capaz de definir e calcular o gradiente e a hessiana de uma função de várias variáveis tanto analiticamente quanto numericamente.
9. Calcular à mão o ponto ótimo de uma função de duas variáveis usando o método do aclive ou declive máximo.
10. Entender as idéias básicas por trás do gradiente conjugado e dos métodos de Newton, Marquardt e quasi-Newton. Em particular, entender os prós e contras entre as abordagens e reconhecer como cada uma delas melhora o aclive ou declive máximo.
11. Ser capaz de reconhecer e descrever um problema de programação linear para representar problemas de engenharia aplicáveis.
12. Ser capaz de resolver um problema bidimensional de programação linear com os métodos gráfico e simplex.
13. Entender as quatro possíveis saídas de um problema de programação linear.
14. Ser capaz de descrever e resolver um problema de otimização não-linear com restrições usando um pacote de software.

CAPÍTULO 13

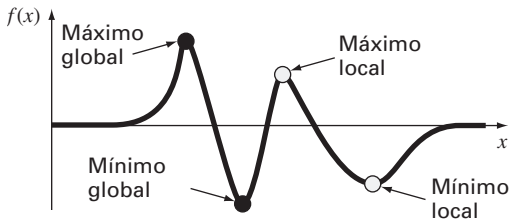
Otimização Unidimensional sem Restrições

Esta seção descreverá técnicas para encontrar o mínimo ou o máximo de uma função de uma única variável, $f(x)$. Uma imagem útil com respeito a isso é a montanha-russa — como a função descrita na Figura 13.1. Lembre-se, da Parte Dois, de que a localização das raízes era complicada pelo fato de que podiam ocorrer várias raízes para uma única função. Analogamente, podem ocorrer tanto pontos ótimos locais quanto globais na otimização. Tais casos são chamados *multimodais*. Em quase todos os exemplos, o interesse será encontrar o maior ou o menor valor global de uma função. Portanto, é preciso ser cuidadoso para não confundir um resultado local com o ponto ótimo global.

Distinguir o extremo local do global pode ser um problema muito difícil no caso geral. Usualmente, há três maneiras de abordar esse problema. Primeiro, algumas vezes pode-se obter informações sobre o comportamento de funções em dimensão baixa graficamente. Segundo, pode-se encontrar pontos ótimos com base em aproximações iniciais amplamente variadas e talvez geradas aleatoriamente, para depois selecionar o maior deles como global. Finalmente, pode-se perturbar o ponto de partida associado com um ponto ótimo local e verificar se a rotina retorna um ponto melhor ou retorna sempre ao mesmo ponto. Embora todas essas abordagens possam ser úteis, o fato é que, em alguns problemas (usualmente os maiores), não existe uma maneira prática de garantir que um ponto ótimo global foi localizado. Entretanto, embora sempre se deva prestar atenção a essas questões, felizmente há numerosos problemas de engenharia nos quais é possível localizar o ótimo global sem ambigüidades.

Assim como a localização de raízes, a otimização em uma dimensão pode ser dividida nos métodos intervalares ou abertos. Como descrito na próxima seção, a busca da razão áurea é um exemplo de método intervalar que depende das aproximações iniciais que delimitam um único ponto ótimo. Em seguida, vem uma abordagem intervalar um pouco mais sofisticada — a interpolação quadrática.

FIGURA 13.1
Uma função que se aproxima assintoticamente de zero em mais e menos infinito e tem dois pontos de máximo e dois de mínimo na vizinhança da origem. Os dois pontos à direita são pontos ótimos locais, enquanto os dois à esquerda são globais.



O último método descrito neste capítulo é um método aberto baseado na idéia de que o cálculo de mínimos e máximos pode ser feito resolvendo-se $f'(x) = 0$. Isso reduz o problema de otimização a encontrar as raízes de $f'(x)$ usando-se técnicas semelhantes às descritas na Parte Dois. Será mostrada uma versão dessa abordagem — o método de Newton.

13.1 BUSCA DA RAZÃO ÁUREA

Na determinação da raiz de uma única equação linear, o objetivo era achar o valor de uma variável x que fornece um *zero* da função $f(x)$. A *otimização em uma única variável* tem como objetivo encontrar o valor de x que fornece um extremo, seja ele um máximo ou um mínimo de $f(x)$.

A busca da razão áurea é uma simples técnica de busca, de propósito geral, em uma única variável. Ela é similar à abordagem da bissecção para localizar raízes, vista no Capítulo 5. Lembre-se de que a bissecção dependia da definição de um intervalo, especificado por uma aproximação inferior (x_l) e por uma aproximação superior (x_u), que delimita uma raiz. A presença de uma raiz entre esses limites era verificada determinando-se que $f(x_l)$ e $f(x_u)$ tinham sinais diferentes. A raiz era então estimada como o ponto médio desse intervalo,

$$x_r = \frac{x_l + x_u}{2}$$

O passo final na iteração da bissecção envolvia a determinação de um intervalo menor. Isso era feito substituindo-se qualquer uma das extremidades x_l ou x_u cujo valor da função tivesse o mesmo sinal que $f(x_r)$. Uma vantagem dessa abordagem era que o novo valor x_r substituía uma das extremidades antigas.

Agora, é possível desenvolver uma abordagem parecida para localizar o ponto ótimo de uma função unidimensional. Por simplicidade, o foco será o problema de achar o máximo. Na discussão sobre o algoritmo computacional, serão descritas as pequenas modificações necessárias para simular um mínimo.

Como na bissecção, pode-se começar definindo um intervalo que contenha uma única resposta. Assim, o intervalo deve conter um único máximo e, portanto, é chamado *unimodal*. Pode-se adotar a mesma nomenclatura que para a bissecção, onde x_l e x_u definiam as extremidades inferior e superior, respectivamente, de um intervalo. No entanto, em contraste com a bissecção, é necessária uma nova estratégia para achar o máximo nesse intervalo. Em vez de usar somente dois valores da função (o que é suficiente para determinar a mudança de sinal e, conseqüentemente, o zero), serão usados três valores de função para determinar a ocorrência do máximo. Assim, deve ser escolhido um ponto adicional no intervalo. Em seguida, é preciso escolher um quarto ponto. Então, o teste para o máximo pode ser aplicado para distinguir se o máximo ocorreu entre os três primeiros ou entre os três últimos pontos.

A chave para tornar a abordagem mais eficiente é uma escolha judiciosa dos pontos intermediários. Como na bissecção, o objetivo é minimizar cálculos da função substituindo os antigos valores por novos valores. O objetivo pode ser alcançado especificando-se que as duas condições seguintes são satisfeitas (Figura 13.2):

$$\ell_0 = \ell_1 + \ell_2 \tag{13.1}$$

$$\frac{\ell_1}{\ell_0} = \frac{\ell_2}{\ell_1} \tag{13.2}$$

A primeira condição especifica que a soma dos comprimentos dos dois subintervalos ℓ_1 e ℓ_2 deve ser igual ao comprimento do intervalo original. A segunda diz que a razão entre os comprimentos deve ser igual. A Equação (13.1) pode ser substituída na Equação (13.2),

$$\frac{\ell_1}{\ell_1 + \ell_2} = \frac{\ell_2}{\ell_1} \tag{13.3}$$

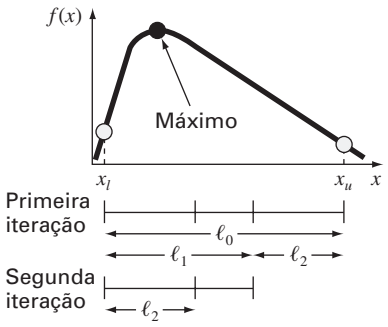


FIGURA 13.2
O passo inicial para o algoritmo de busca da razão áurea envolve a escolha de dois pontos interiores segundo a razão áurea.

Se for tomado o inverso e se fizer $R = \ell_2 / \ell_1$, chega-se a

$$1 + R = \frac{1}{R} \tag{13.4}$$

ou

$$R^2 + R - 1 = 0 \tag{13.5}$$

a qual pode ser resolvida para determinar a raiz positiva

$$R = \frac{-1 + \sqrt{1 - 4(-1)}}{2} = \frac{\sqrt{5} - 1}{2} = 0{,}61803\ldots \tag{13.6}$$

Esse valor, que é conhecido desde a Antiguidade, é chamado de *razão áurea* (ver Quadro 13.1). Por permitir que os pontos ótimos sejam encontrados de forma eficiente, ele é o elemento-chave do método da seção áurea que estamos desenvolvendo conceitualmente. Agora, vamos desenvolver um algoritmo para implementar essa abordagem em um computador.

Quadro 13.1 A Razão Áurea e os Números de Fibonacci

Em muitas culturas, são atribuídas qualidades a certos números. Por exemplo, o Ocidente está familiarizado com o "7 da sorte" e a "Sexta-feira 13". Os gregos antigos chamavam o seguinte número de "razão áurea":

$$\frac{\sqrt{5} - 1}{2} = 0{,}61803\ldots$$

Essa razão era empregada para vários propósitos, incluindo a determinação do retângulo na Figura 13.3. Essas proporções eram consideradas esteticamente agradáveis pelos gregos. Entre outras coisas, muitos dos seus templos seguiam tal forma.

A razão áurea está relacionada com uma importante série matemática conhecida como os *números de Fibonacci*, que são

0, 1, 1, 2, 3, 5, 8, 13, 21, 34, . . .

Assim, cada número depois dos dois primeiros representa a soma dos dois precedentes. Essa seqüência aparece em diversas áreas da ciência e da engenharia. No contexto da presente discussão, uma propriedade interessante da seqüência de Fibonacci está relacionada com a razão de números consecutivos na seqüência; isto é, $0/1 = 0$, $1/1 = 1$, $1/2 = 0{,}5$, $2/3 = 0{,}667$, $3/5 = 0{,}6$, $5/8 = 0{,}625$, $8/13 = 0{,}615$, e assim por diante. À medida que se continua o processo, a razão entre os números consecutivos se aproxima da razão áurea!

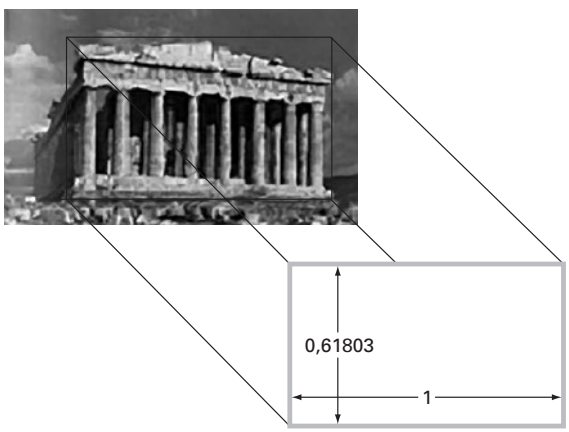


FIGURA 13.3
O Parthenon, em Atenas, Grécia, foi construído no século V a.C. As dimensões frontais se encaixam quase exatamente em um retângulo áureo.

Como mencionado anteriormente e descrito na Figura 13.4, o método começa com duas aproximações iniciais, x_l e x_u, que delimitem um extremo local de $f(x)$. Em seguida, dois pontos interiores x_1 e x_2 são escolhidos segundo a razão áurea,

$$d = \frac{\sqrt{5}-1}{2}(x_u - x_l)$$
$$x_1 = x_l + d$$
$$x_2 = x_u - d$$

A função é calculada nesses dois pontos interiores. Podem ocorrer dois resultados:

1. Se, como no caso da Figura 13.4, $f(x_1) > f(x_2)$, então o domínio de x à esquerda de x_2, de x_l a x_2, pode ser eliminado porque não contém o máximo. Nesse caso, x_2 se torna o novo valor de x_l para a próxima rodada.
2. Se $f(x_2) > f(x_1)$, então o domínio de x à direita de x_1 de x_1 a x_u teria sido eliminado. Nesse caso, x_1 se torna o novo x_u para a próxima rodada.

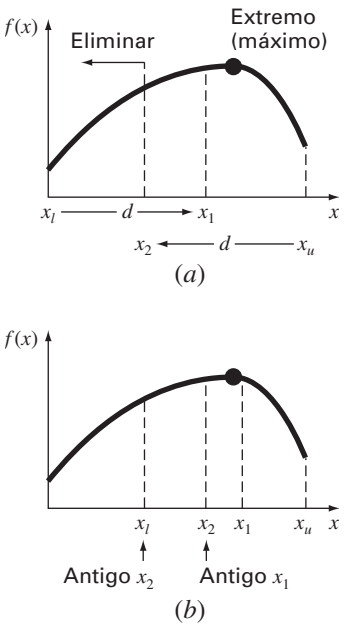
Agora, aqui está o real benefício do uso da razão áurea. Como os valores originais de x_1 e x_2 foram escolhidos usando-se a razão áurea, não é preciso recalcular todos os valores da função para a próxima iteração. Por exemplo, para o caso ilustrado na Figura 13.4, o antigo x_1 se torna o novo x_2. Isso significa que já se tem o valor para o novo $f(x_2)$, pois ele é o mesmo que o valor da função no antigo x_1.

Para completar o algoritmo, é necessário apenas determinar o novo x_1, o que é feito com a mesma proporcionalidade que antes,

$$x_1 = x_l + \frac{\sqrt{5}-1}{2}(x_u - x_l)$$

Uma abordagem análoga pode ser usada para o caso alternativo em que o ponto ótimo cai no subintervalo da esquerda.

FIGURA 13.4
(*a*) O passo inicial do algoritmo de busca da razão áurea envolve a escolha de dois pontos interiores segundo a razão áurea. (*b*) O segundo passo envolve a definição de um novo intervalo que inclui o ponto ótimo.



Conforme as iterações são repetidas, o intervalo contendo o ponto extremo é reduzido rapidamente. De fato, a cada rodada o intervalo é reduzido por um fator igual ao da razão áurea (cerca de 61,8%). Isso significa que, depois de 10 rodadas, o intervalo foi encolhido de cerca de $0{,}618^{10}$ ou 0,008 ou 0,8% do seu tamanho inicial. Depois de 20 rodadas, chega a 0,0066%. Isso não é tão bom quanto a redução conseguida com a bissecção, mas esse é um problema mais difícil.

EXEMPLO 13.1 Busca da Seção Áurea

Enunciado do Problema. Use a busca da seção áurea para achar o máximo de

$$f(x) = 2 \operatorname{sen} x - \frac{x^2}{10}$$

no intervalo entre $x_l = 0$ e $x_u = 4$.

Solução. Primeiro, a razão áurea é usada para criar os dois pontos interiores

$$d = \frac{\sqrt{5}-1}{2}(4-0) = 2{,}472$$

$$x_1 = 0 + 2{,}472 = 2{,}472$$

$$x_2 = 4 - 2{,}472 = 1{,}528$$

A função pode ser calculada nos pontos interiores

$$f(x_2) = f(1{,}528) = \frac{1{,}528^2}{10} - 2 \operatorname{sen}(1{,}528) = 1{,}765$$

$$f(x_1) = f(2{,}472) = 0{,}63$$

Uma vez que $f(x_2) > f(x_1)$, o máximo está no intervalo definido por x_l, x_2 e x_1. Portanto, para o novo intervalo, a extremidade inferior continua sendo $x_l = 0$, e x_1 se torna a extremidade superior, ou seja, $x_u = 2{,}472$. Além disso, o valor antigo x_2 se torna o novo x_1, isto é, $x_1 = 1{,}528$. Mais ainda, não é preciso recalcular $f(x_1)$ porque este foi determinado na iteração prévia como $f(1{,}528) = 1{,}765$.

Tudo o que resta fazer é calcular a nova razão áurea e x_2,

$$d = \frac{\sqrt{5}-1}{2}(2{,}472 - 0) = 1{,}528$$

$$x_2 = 2{,}4721 - 1{,}528 = 0{,}944$$

O valor da função em x_2 é $f(0{,}994) = 1{,}531$. Como esse valor é menor que o valor da função em x_1, o máximo está no intervalo prescrito por x_2, x_1 e x_u.

O processo pode ser repetido, com o resultado tabulado a seguir:

i	x_l	$f(x_l)$	x_2	$f(x_2)$	x_1	$f(x_1)$	x_u	$f(x_u)$	d
1	0	0	1,5279	1,7647	2,4721	0,6300	4,0000	−3,1136	2,4721
2	0	0	0,9443	1,5310	1,5279	1,7647	2,4721	0,6300	1,5279
3	0,9443	1,5310	1,5279	1,7647	1,8885	1,5432	2,4721	0,6300	0,9443
4	0,9443	1,5310	1,3050	1,7595	1,5279	1,7647	1,8885	1,5432	0,5836
5	1,3050	1,7595	1,5279	1,7647	1,6656	1,7136	1,8885	1,5432	0,3607
6	1,3050	1,7595	1,4427	1,7755	1,5279	1,7647	1,6656	1,7136	0,2229
7	1,3050	1,7595	1,3901	1,7742	1,4427	1,7755	1,5279	1,7647	0,1378
8	1,3901	1,7742	1,4427	1,7755	1,4752	1,7732	1,5279	1,7647	0,0851

Observe que o máximo corrente é destacado para cada iteração. Depois da oitava iteração, o máximo ocorre em $x = 1{,}4427$ com o valor da função de 1,7755. Portanto, o resultado está convergindo para o valor verdadeiro 1,7757 em $x = 1{,}4276$.

Lembre-se de que, para a bissecção (Seção 5.2.1), pode ser calculado um limite superior exato para o erro a cada iteração. Usando um raciocínio análogo, um limite superior para a busca da razão áurea pode ser deduzido da seguinte forma: uma vez que uma iteração é completada, o ponto ótimo cairá em um de dois intervalos. Se x_2 for o valor ótimo da função, ele estará no intervalo inferior (x_l, x_2, x_1). Se x_1 for o valor ótimo da função, estará no intervalo superior (x_2, x_1, x_u). Como os pontos interiores são simétricos, qualquer um dos casos pode ser usado para definir o erro.

Olhando para o intervalo superior, se o valor verdadeiro estiver na extremidade esquerda, a distância máxima do valor estimado será

$$\begin{aligned}\Delta x_a &= x_1 - x_2 \\ &= x_l + R(x_u - x_l) - x_u + R(x_u - x_l) \\ &= (x_l - x_u) + 2R(x_u - x_l) \\ &= (2R - 1)(x_u - x_l)\end{aligned}$$

ou $0{,}236(x_u - x_l)$.

Se o valor verdadeiro estiver na extremidade direita, a distância máxima do valor estimado será

$$\begin{aligned}\Delta x_b &= x_u - x_1 \\ &= x_u - x_l - R(x_u - x_l) \\ &= (1 - R)(x_u - x_l)\end{aligned}$$

ou $0{,}382(x_u - x_l)$. Portanto, esse caso representa o erro máximo. O resultado pode então ser normalizado para o valor ótimo para a iteração, x_{opt}, o que fornece

$$\varepsilon_a = (1 - R)\left|\frac{x_u - x_l}{x_{\text{opt}}}\right| 100\%$$

Essa estimativa proporciona uma base para o término das iterações.

O pseudocódigo para o algoritmo da busca da razão áurea para maximização é apresentado na Figura 13.5*a*. As pequenas modificações para converter o algoritmo para a minimização estão listadas na Figura 13.5*b*. Em ambas as versões, o valor de x para o ponto ótimo é retornado como o valor da função (*aurea*). Além disso, o valor de $f(x)$ no ponto ótimo é retornado como a variável $f(x)$.

Você deve estar imaginando por que foram enfatizados o número reduzido de cálculos da função na busca da razão áurea. Naturalmente, para resolver uma única otimização, o ganho em velocidade seria desprezível. Entretanto, existem dois importantes contextos nos quais a minimização do número de cálculos da função pode ser importante. São eles:

1. *Muitos cálculos.* Há casos em que o algoritmo de busca da razão áurea pode ser uma parte de um cálculo muito maior. Nessa situação, o algoritmo pode ser chamado muitas vezes. Portanto, manter o cálculo da função em um número mínimo pode trazer grandes dividendos para tais casos.
2. *Cálculos demorados.* Por razões pedagógicas, foram utilizadas funções simples na maioria dos exemplos. É preciso entender que uma função pode ser muito complexa e levar muito tempo para ser calculada. Por exemplo, em uma parte posterior deste livro, será descrito como a otimização pode ser utilizada para estimar os parâmetros de um modelo que consiste em um sistema de equações diferenciais. Para esses casos, a "função" envolve a integração do modelo, que consome tempo. Qualquer método que minimize tais cálculos seria vantajoso.

```
FUNCTION Aurea (xlow, xhigh, maxit, es, fx)
R = (5^0.5 − 1)/2
xℓ = xlow; xu = xhigh
iter = 1
d  = R * (xu − xℓ)
x1 = xℓ + d; x2 = xu − d
f1 = f(x1)
f2 = f(x2)
IF f1 > f2 THEN                                   IF f1 < f2 THEN
  xopt = x1
  fx = f1
ELSE
  xopt = x2
  fx = f2
END IF
DO
  d = R*d
  IF f1 > f2 THEN                                 IF f1 < f2 THEN
    xℓ = x2
    x2 = x1
    x1 = xℓ+d
    f2 = f1
    f1 = f(x1)
  ELSE
    xu = x1
    x1 = x2
    x2 = xu−d
    f1 = f2
    f2 = f(x2)
  END IF
  iter = iter+1
  IF f1 > f2 THEN                                 IF f1 < f2 THEN
    xopt = x1
    fx = f1
  ELSE
    xopt = x2
    fx = f2
  END IF
  IF xopt ≠ 0. THEN
    ea = (1.−R) *ABS((xu − xℓ)/xopt) * 100.
  END IF
  IF ea ≤ es OR iter ≥ maxit EXIT
END DO
Aurea = xopt
END Aurea
```

(*a*) **Maximização** (*b*) **Minimização**

FIGURA 13.5
Algoritmo para a busca da seção áurea.

13.2 INTERPOLAÇÃO QUADRÁTICA

A interpolação quadrática aproveita-se do fato de que um polinômio de segundo grau freqüentemente fornece uma boa aproximação para a forma de $f(x)$ nas proximidades do ponto ótimo (Figura 13.6).

Assim como existe somente uma reta ligando dois pontos, há apenas uma função quadrática ou parábola ligando três pontos. Então, se houver três pontos que conjuntamente delimitem um ponto ótimo, pode-se ajustar uma parábola aos três pontos. É possível,

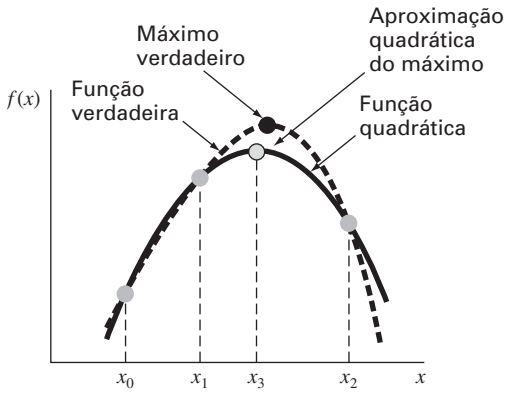


FIGURA 13.6
Descrição gráfica da interpolação quadrática.

desse modo, derivá-la, igualar o resultado a zero, e resolver para determinar uma estimativa do ponto ótimo x. Pode-se mostrar mediante um pouco de manipulação algébrica que o resultado é

$$x_3 = \frac{f(x_0)\left(x_1^2 - x_2^2\right) + f(x_1)\left(x_2^2 - x_0^2\right) + f(x_2)\left(x_0^2 - x_1^2\right)}{2f(x_0)(x_1 - x_2) + 2f(x_1)(x_2 - x_0) + 2f(x_2)(x_0 - x_1)} \tag{13.7}$$

onde x_0, x_1 e x_2 são as aproximações iniciais e x_3 é o valor de x que corresponde ao máximo valor da função quadrática ajustada às aproximações.

EXEMPLO 13.2 Interpolação Quadrática

Enunciado do Problema. Use a interpolação quadrática para aproximar o ponto de máximo de

$$f(x) = 2 \operatorname{sen} x - \frac{x^2}{10}$$

com aproximações iniciais $x_0 = 0$, $x_1 = 1$ e $x_2 = 4$.

Solução. Os valores da função nas três aproximações são dados por

$$\begin{aligned} x_0 &= 0 \qquad f(x_0) = 0 \\ x_1 &= 1 \qquad f(x_1) = 1{,}5829 \\ x_2 &= 4 \qquad f(x_2) = -3{,}1136 \end{aligned}$$

o que, substituindo na Equação (13.7), dá

$$x_3 = \frac{0(1^2 - 4^2) + 1{,}5829(4^2 - 0^2) + (-3{,}1136)(0^2 - 1^2)}{2(0)(1 - 4) + 2(1{,}5829)(4 - 0) + 2(-3{,}1136)(0 - 1)} = 1{,}5055$$

o qual tem o valor da função de $f(1{,}5055) = 1{,}7691$.

Em seguida, uma estratégia parecida àquela da busca da razão áurea pode ser empregada para determinar qual ponto deve ser descartado. Uma vez que o valor da função para o novo ponto é maior do que para o ponto intermediário (x_1) e o novo valor de x está à direita do ponto intermediário, a aproximação inferior (x_0) é descartada. Portanto, para a próxima iteração,

$$\begin{aligned} x_0 &= 1 \qquad\quad f(x_0) = 1{,}5829 \\ x_1 &= 1{,}5055 \qquad f(x_1) = 1{,}7691 \\ x_2 &= 4 \qquad\quad f(x_2) = -3{,}1136 \end{aligned}$$

o que pode ser substituído na Equação (13.7) para dar

$$x_3 = \frac{1{,}5829(1{,}5055^2 - 4^2) + 1{,}7691(4^2 - 1^2) + (-3{,}1136)(1^2 - 1{,}5055^2)}{2(1{,}5829)(1{,}5055 - 4) + 2(1{,}7691)(4 - 1) + 2(-3{,}1136)(1 - 1{,}5055)}$$
$$= 1{,}4903$$

o qual tem o valor da função de $f(1{,}4903) = 1{,}7714$.

O processo pode ser repetido, com os resultados tabulados a seguir:

i	x_0	$f(x_0)$	x_1	$f(x_1)$	x_2	$f(x_2)$	x_3	$f(x_3)$
1	0,0000	0,0000	1,0000	1,5829	4,0000	−3,1136	1,5055	1,7691
2	1,0000	1,5829	1,5055	1,7691	4,0000	−3,1136	1,4903	1,7714
3	1,0000	1,5829	1,4903	1,7714	1,5055	1,7691	1,4256	1,7757
4	1,0000	1,5829	1,4256	1,7757	1,4903	1,7714	1,4266	1,7757
5	1,4256	1,7757	1,4266	1,7757	1,4903	1,7714	1,4275	1,7757

Assim, em cinco iterações, o resultado está convergindo rapidamente para o valor verdadeiro 1,7757 em $x = 1{,}4276$.

Deve-se mencionar que, do mesmo modo como no método da falsa posição, a interpolação quadrática pode ficar presa, com apenas uma extremidade do intervalo convergindo. Portanto, a convergência pode ser lenta. Por exemplo, observe que, no exemplo, 1,0000 era uma extremidade para a maioria das iterações.

Esse método, bem como outros que usam polinômios de terceiro grau, pode ser formulado em um algoritmo que contenha testes de convergência, estratégias cuidadosas de seleção dos pontos a serem retidos em cada iteração e tentativas de minimizar o erro de arredondamento acumulado. Em particular, ver o método de Brent em Press et al. (1992).

13.3 O MÉTODO DE NEWTON

Lembre-se de que o método de Newton-Raphson do Capítulo 6 é um método aberto que acha a raiz x de uma função, $f(x) = 0$. O método é sintetizado como

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

Uma abordagem aberta análoga pode ser usada para achar o ponto ótimo de $f(x)$, definindo uma nova função, $g(x) = f'(x)$. Assim, uma vez que o mesmo ponto ótimo x^* satisfaz ambos

$$f'(x^*) = g(x^*) = 0$$

pode-se usar o seguinte,

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)} \tag{13.8}$$

como uma técnica para encontrar o mínimo ou o máximo de $f(x)$. Deve-se observar que essa equação pode também ser deduzida escrevendo-se uma série de Taylor de segunda ordem para $f(x)$ e fazendo a derivada da série igual a zero. O método de Newton é um método aberto, análogo ao método de Newton-Raphson, porque não exige aproximações iniciais que delimitem o ponto ótimo. Além disso, ele também compartilha a desvantagem de poder ser divergente. Finalmente, em geral é uma boa idéia verificar se a segunda derivada tem o sinal correto para confirmar que a técnica está convergindo para o resultado desejado.

EXEMPLO 13.3 Método de Newton

Enunciado do Problema. Use o método de Newton para achar o máximo de

$$f(x) = 2 \operatorname{sen} x - \frac{x^2}{10}$$

com a aproximação inicial $x_0 = 2{,}5$.

Solução. A primeira e a segunda derivadas da função podem ser calculadas como

$$f'(x) = 2\cos x - \frac{x}{5}$$

$$f''(x) = -2 \operatorname{sen} x - \frac{1}{5}$$

as quais podem ser substituídas na Equação (13.8) para dar

$$x_{i+1} = x_i - \frac{2\cos x_i - x_i/5}{-2 \operatorname{sen} x_i - 1/5}$$

Substituindo a estimativa inicial, obtém-se

$$x_1 = 2{,}5 - \frac{2\cos 2{,}5 - 2{,}5/5}{-2 \operatorname{sen} 2{,}5 - 1/5} = 0{,}99508$$

que tem um valor da função de 1,57859. A segunda iteração dá

$$x_1 = 0{,}995 - \frac{2\cos 0{,}995 - 0{,}995/5}{-2 \operatorname{sen} 0{,}995 - 1/5} = 1{,}46901$$

que tem um valor da função de 1,77385.

O processo pode ser repetido, retornando os resultados tabulados a seguir:

i	*x*	*f*(*x*)	*f*′(*x*)	*f*″(*x*)
0	2,5	0,57194	−2,10229	−1,39694
1	0,99508	1,57859	0,88985	−1,87761
2	1,46901	1,77385	−0,09058	−2,18965
3	1,42764	1,77573	−0,00020	−2,17954
4	1,42755	1,77573	0,00000	−2,17952

Assim, em quatro iterações, o resultado converge rapidamente para o valor verdadeiro.

Embora o método de Newton funcione bem em alguns casos, é impraticável para aqueles em que as derivadas não podem ser calculadas convenientemente. Para tais casos, outras abordagens que não envolvem o cálculo das derivadas estão disponíveis. Por exemplo, uma versão do tipo secante do método de Newton pode ser desenvolvida utilizando-se aproximações por diferenças finitas para o cálculo das derivadas.

Uma reserva ainda maior com respeito a essa abordagem é que ela pode divergir dependendo da natureza da função e da qualidade da estimativa inicial. Portanto, ela é empregada usualmente apenas quando estamos próximos do valor ótimo. Técnicas híbridas que usam *abordagens intervalares* longe do ponto ótimo e *métodos abertos* próximo ao ponto ótimo tentam explorar os pontos fortes de ambas as abordagens.

Isso conclui nossos tratamentos dos métodos para determinar o ponto ótimo de uma função de uma única variável. Alguns exemplos da engenharia serão apresentados no Capítulo 16. Além disso, as técnicas descritas aqui são um elemento importante em alguns dos procedimentos para otimizar funções de muitas variáveis, como discutido no próximo capítulo.

PROBLEMAS

13.1 Dada a fórmula

$$f(x) = -x^2 + 8x - 12$$

(a) Determine analiticamente (isto é, usando derivação) o máximo e o valor correspondente de x para essa função.
(b) Verifique que a Equação (13.7) fornece o mesmo resultado, com base nas aproximações iniciais $x_0 = 0$, $x_1 = 2$ e $x_2 = 6$.

13.2 Dado

$$f(x) = -1{,}5x^6 - 2x^4 + 12x$$

(a) Trace a função.
(b) Use métodos analíticos para demonstrar que a função é côncava para todos os valores de x.
(c) Derive a função e então use um método da localização de raízes para achar o máximo de $f(x)$ e o valor correspondente de x.

13.3 Ache o valor de x que maximiza $f(x)$ no Problema 13.2 utilizando a busca da razão áurea. Use como aproximações iniciais $x_l = 0$ e $x_u = 2$ e faça três iterações.

13.4 Repita o Problema 13.3, utilizando agora a interpolação quadrática. Use como aproximações iniciais $x_0 = 0$, $x_1 = 1$ e $x_2 = 2$ e faça três iterações.

13.5 Repita o Problema 13.3, utilizando agora o método de Newton. Use como aproximação inicial $x_0 = 2$ e faça três iterações.

13.6 Discuta as vantagens e as desvantagens da busca da razão áurea, da interpolação quadrática e do método de Newton para localizar o valor ótimo em uma dimensão.

13.7 Use os seguintes métodos para achar o máximo de

$$f(x) = 4x - 1{,}8x^2 + 1{,}2x^3 - 0{,}3x^4$$

(a) Busca da razão áurea ($x_l = -2$, $x_u = 4$, $\varepsilon_s = 1\%$).
(b) Interpolação quadrática ($x_0 = 1{,}75$, $x_1 = 2$, $x_2 = 2{,}5$, iterações $= 4$).
(c) Método de Newton ($x_0 = 3$, $\varepsilon_s = 1\%$).

13.8 Considere a seguinte função:

$$f(x) = -x^4 - 2x^3 - 8x^2 - 5x$$

Use métodos analíticos e gráficos para mostrar que a função tem um máximo para algum valor de x no intervalo $-2 \le x \le 1$.

13.9 Use os seguintes métodos para achar o máximo da função do Problema 13.8:
(a) Busca da razão áurea ($x_l = -2$, $x_u = 1$, $\varepsilon_s = 1\%$).
(b) Interpolação quadrática ($x_0 = -2$, $x_1 = -1$, $x_2 = 1$, iterações $= 4$).
(c) Método de Newton ($x_0 = -1$, $\varepsilon_s = 1\%$).

13.10 Considere a seguinte função:

$$f(x) = 2x + \frac{3}{x}$$

Faça 10 iterações da interpolação quadrática para localizar o mínimo. Comente sobre a convergência dos seus resultados ($x_0 = 0{,}1$, $x_1 = 0{,}5$, $x_2 = 5$).

13.11 Considere a seguinte função:

$$f(x) = 3 + 6x + 5x^2 + 3x^3 + 4x^4$$

Localize o mínimo determinando as raízes da derivada dessa função. Use a bissecção com as aproximações iniciais $x_l = -2$ e $x_u = 1$.

13.12 Determine o mínimo da função do Problema 13.11 com os seguintes métodos:
(a) Método de Newton ($x_0 = -1$, $\varepsilon_s = 1\%$).
(b) Método de Newton, mas utilizando uma aproximação por diferenças finitas para as estimativas das derivadas

$$f'(x) = \frac{f(x_i + \delta x_i) - f(x_i - \delta x_i)}{2\delta x_i}$$

$$f''(x) = \frac{f(x_i + \delta x_i) - 2f(x_i) - f(x_i - \delta x_i)}{(\delta x_i)^2}$$

onde δ é uma fração de perturbação ($= 0{,}01$). Use uma aproximação inicial $x_0 = -1$ e itere até $\varepsilon_s = 1\%$.

13.13 Desenvolva um programa usando uma linguagem de programação ou de macro para implementar o algoritmo da busca da razão áurea. Faça o programa de modo que ele seja expressamente projetado para localizar um máximo. A sub-rotina deve ter as seguintes propriedades:

- Iterar até que o erro relativo fique abaixo do critério de parada ou exceda um número máximo de iterações.
- Retornar tanto o ponto ótimo x quanto o valor de $f(x)$.
- Minimizar o número de cálculos do valor da função.

Teste o seu programa com o mesmo problema do Exemplo 13.1.

13.14 Desenvolva um programa como descrito no Problema 13.13, mas faça de modo que o usuário possa escolher a maximização ou a minimização.

13.15 Desenvolva um programa usando uma linguagem de programação ou de macro para implementar o algoritmo da interpolação quadrática. Faça o programa de modo que ele seja expressamente projetado para localizar o máximo. A sub-rotina deve ter as seguintes características:

- Ser baseada em duas aproximações iniciais e fazer o programa gerar o terceiro valor inicial no ponto médio do intervalo.
- Verificar se as aproximações delimitam um máximo. Em caso negativo, a sub-rotina não deve implementar o algoritmo e deve retornar uma mensagem de erro.
- Iterar até que o erro relativo fique abaixo de um critério de parada ou exceda um número máximo de iterações.
- Retornar tanto o ponto ótimo x quanto o valor de $f(x)$.
- Minimizar o número de cálculos do valor da função.

Teste o seu programa com o mesmo problema do Exemplo 13.2.

13.16 Desenvolva um programa usando uma linguagem de programação ou de macro para implementar o método de Newton. A sub-rotina deve ter as seguintes características:

- Iterar até que o erro relativo fique abaixo de um critério de parada ou exceda um número máximo de iterações.
- Retornar tanto o ótimo x quanto o valor de $f(x)$.

Teste o seu programa com o mesmo problema do Exemplo 13.3.

13.17 São feitas medidas de pressão em certos pontos atrás de um aerofólio durante um certo tempo. Os dados se ajustam à curva $y = 6 \cos x - 1{,}5 \operatorname{sen} x$ de $x = 0$ a 6 s. Faça quatro iterações do método da razão áurea para achar a pressão mínima. Use $x_l = 2$ e $x_u = 4$.

13.18 A trajetória de uma bola pode ser calculada como

$$y = (\operatorname{tg} \theta_0)x - \frac{g}{2v_0^2 \cos^2 \theta_0}x^2 + y_0$$

onde y é a altura (m), θ_0 é o ângulo inicial (em radianos), v_0 é a velocidade inicial (m/s), $g = 9{,}81$ m/s^2 é a constante gravitacional e y_0 é a altura inicial (m). Use a busca da razão áurea para determinar a altura máxima, dados $y_0 = 1$ m, $v_0 = 25$ m/s e $\theta_0 = 50^o$. Itere até que o erro aproximado caia abaixo de $\varepsilon_s = 1\%$ usando aproximações iniciais $x_l = 0$ e $x_u = 60$ m.

13.19 A deflexão de uma viga uniforme submetida a uma carga crescente e linearmente distribuída pode ser escrita como

$$y = \frac{w_0}{120EIL}(-x^5 + 2L^2x^3 - L^4x)$$

Dado que $L = 600$ cm, $E = 50.000$ kN/cm^2, $I = 30.000$ cm^4 e $w_0 = 2{,}5$ kN/cm, determine o ponto de máxima deflexão (**a**) graficamente, (**b**) usando a busca da razão áurea até que o erro aproximado fique abaixo de $\varepsilon_s = 1\%$ com aproximações iniciais $x_l = 0$ e $x_u = L$.

13.20 Um objeto com uma massa de 100 kg é arremessado para cima a partir de um ponto na superfície da terra, com uma velocidade de 50 m/s. Se o objeto está submetido a um arrasto linear ($c = 15$ kg/s), use a busca da razão áurea para determinar qual a máxima altura que o objeto atinge (*Sugestão*: lembre-se da Seção PT4.1.2).

13.21 A distribuição normal é uma curva em forma de sino, definida por

$$y = e^{-x^2}$$

Use a busca da razão áurea para determinar a posição do ponto de inflexão nessa curva para x positivo.

CAPÍTULO 14

Otimização Multidimensional sem Restrições

Este capítulo descreve técnicas para encontrar o mínimo ou máximo de uma função de muitas variáveis. Lembre-se, do Capítulo 13, de que nossa imagem pictórica de uma busca unidimensional era uma montanha-russa. Para o caso bidimensional, a imagem se tornou a de montanhas e vales (Figura 14.1). Para problemas de maior dimensão, não é possível ter imagens convenientes.

Este capítulo foi limitado ao caso bidimensional. Adotou-se essa abordagem porque as propriedades essenciais das buscas multidimensionais são, em geral, mais bem comunicadas visualmente.

Técnicas para a otimização multidimensional sem restrições podem ser classificadas de diversas maneiras. Para o propósito da presente discussão, elas serão divididas dependendo do fato de exigirem ou não o cálculo da derivada. As abordagens que não exigem o cálculo da derivada são chamadas de *não-gradientes* ou de *métodos diretos*. Aquelas que exigem derivadas são chamadas *métodos gradientes* ou de *declive* (ou *aclive*).

14.1 MÉTODOS DIRETOS

Esses métodos variam de abordagens de simples força bruta até técnicas mais elegantes que tentam explorar a natureza da função. A discussão se inicia com uma abordagem por força bruta.

14.1.1 Busca Aleatória

Um exemplo simples de uma abordagem por força bruta é o *método de busca aleatória*. Como o nome indica, ele calcula repetidamente a função em valores das variáveis independentes escolhidos aleatoriamente. Se um número suficiente de amostras for estudado, o valor ótimo será eventualmente localizado.

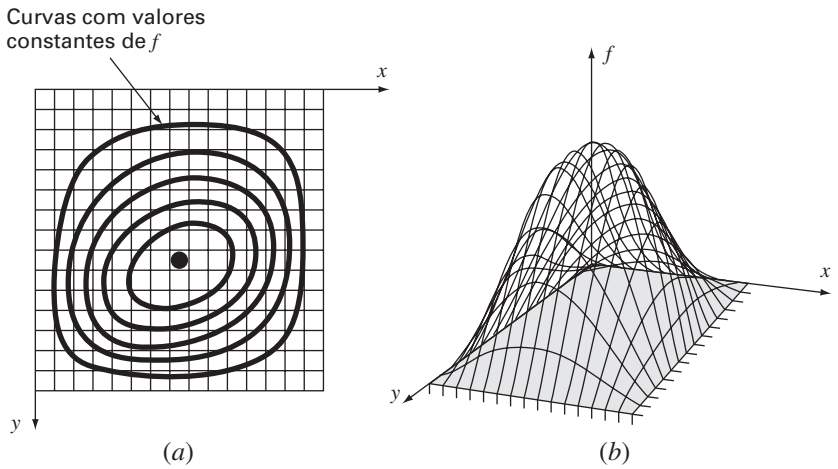


FIGURA 14.1
A maneira mais palpável de visualizar buscas bidimensionais é no contexto da subida de uma montanha (maximização) ou na descida para um vale (minimização). (*a*) Um mapa topográfico em 2-D que corresponde à montanha em 3-D em (*b*).

EXEMPLO 14.1

Método da Busca Aleatória

Enunciado do Problema. Use um gerador de números aleatórios para localizar o máximo de

$$f(x, y) = y - x - 2x^2 - 2xy - y^2 \tag{E14.1.1}$$

no domínio limitado por $x = -2$ a 2 e $y = 1$ a 3. Esse domínio é descrito na Figura 14.2. Observe que um único máximo de 1,5 ocorre em $x = -1$ e $y = 1{,}5$.

Solução. Geradores de números aleatórios tipicamente geram valores entre 0 e 1. Se tal número for designado por r, a seguinte fórmula pode ser usada para gerar valores de x aleatoriamente dentro de um intervalo entre x_l e x_u:

$$x = x_l + (x_u - x_l)r$$

Para a presente aplicação, $x_l = -2$ e $x_u = 2$, e a fórmula será

$$x = -2 + (2 - (-2))r = -2 + 4r$$

Isso pode ser testado substituindo-se 0 e 1 e obtendo-se -2 e 2, respectivamente.

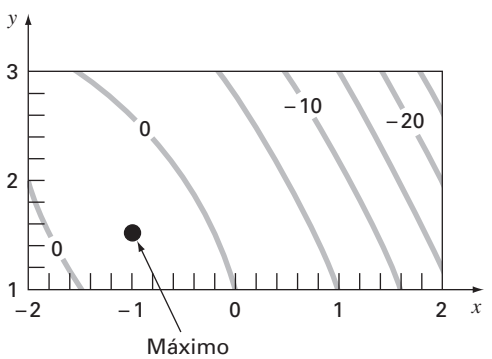
Analogamente para y, a fórmula para esse exemplo pode ser desenvolvida como

$$y = y_l + (y_u - y_l)r = 1 + (3 - 1)r = 1 + 2r$$

O seguinte macrocódigo Excel VBA usa a função de número aleatório VBA Rnd, para gerar pares (x, y). Estes são então substituídos na Equação (E14.1.1). O máximo valor entre essas tentativas aleatórias é armazenado na variável *maxf*, e os valores correspondentes de x e y em *maxx* e *maxy*, respectivamente.

```
maxf = -1E-9
For j = 1 To n
  x = -2 + 4 * Rnd
  y = 1 + 2 * Rnd
  fn = y - x - 2 * x ^ 2 - 2 * x * y - y ^ 2
  If fn > maxf Then
    maxf = fn
    maxx = x
    maxy = y
  End If
Next j
```

FIGURA 14.2
Equação (E14.1.1) mostrando o máximo em $x = -1$ e $y = 1{,}5$.



Diversos números de iterações resultam em

Iterações	x	y	$f(x, y)$
1000	−0.9886	1.4282	1.2462
2000	−1.0040	1.4724	1.2490
3000	−1.0040	1.4724	1.2490
4000	−1.0040	1.4724	1.2490
5000	−1.0040	1.4724	1.2490
6000	−0.9837	1.4936	1.2496
7000	−0.9960	1.5079	1.2498
8000	−0.9960	1.5079	1.2498
9000	−0.9960	1.5079	1.2498
10000	−0.9978	1.5039	1.2500

Os resultados indicam que a técnica atinge o máximo verdadeiro.

Essa simples abordagem por força bruta funciona mesmo para funções descontínuas e não-diferenciáveis. Além disso, ela sempre acha o ponto ótimo global em vez de um ponto ótimo local. Sua principal desvantagem é que, conforme o número de variáveis independentes cresce, o esforço de implementação necessário pode ficar pesado. Além disso, não é muito eficiente porque não leva em conta o comportamento da função subjacente. O restante das abordagens descritas neste capítulo leva em conta o comportamento da função bem como os resultados das tentativas anteriores para aumentar a velocidade de convergência. Assim, embora a busca aleatória possa certamente se mostrar útil em contextos de problemas específicos, os métodos a seguir são de utilidade mais geral e quase sempre levam a um processo de convergência mais eficiente.

Deve ser observado que estão disponíveis técnicas mais sofisticadas — abordagens heurísticas que foram desenvolvidas para lidar com problemas não-lineares e/ou descontínuos, que a otimização clássica não pode tratar eficientemente, se é que pode tratar de alguma forma. Alguns exemplos são têmpera simulada, busca tabu, redes neurais artificiais e algoritmo genético. O *algoritmo genético* é o de aplicação mais ampla, tendo alguns pacotes comerciais já disponíveis. Holland (1975) é o pioneiro na abordagem do algoritmo genético, e Davis (1991) e Goldberg (1989) fornecem uma boa visão geral da teoria e das aplicações desse método.

14.1.2 Busca Unidimensional e por Padrões

Ter uma abordagem de otimização eficiente que não exija o cálculo das derivadas é uma idéia muito atraente. O método de busca aleatória descrito não exige o cálculo da derivada, mas não é muito eficiente. Esta seção descreve uma abordagem, o método de busca unidimensional, que é mais eficiente e também não requer o cálculo da derivada.

A estratégia básica por trás do *método de busca unidimensional* é mudar uma variável por vez para melhorar a aproximação enquanto as outras variáveis são mantidas constantes. Já que somente uma variável é modificada, o problema se reduz a uma seqüência de buscas unidimensionais que podem ser resolvidas usando-se vários métodos (incluindo aqueles descritos no Capítulo 13).

Vamos executar uma busca unidimensional graficamente, como mostrado na Figura 14.3. Comece no ponto 1 e mova-se ao longo do eixo x com y constante até o ponto 2. Você pode ver que o ponto 2 é um máximo observando que a trajetória ao longo do eixo x apenas toca uma curva de nível nesse ponto. Em seguida, mova-se ao longo do eixo y com x constante até o ponto 3. Continue com esse processo gerando os pontos 4, 5, 6 etc.

Embora estejamos movendo-nos gradualmente em direção ao máximo, a busca se torna menos eficiente conforme nos movemos ao longo da estreita crista em direção ao máximo. Entretanto, observe também que as retas unindo pontos alternados como 1-3, 3-5 ou 2-4, 4-6 apontam na direção geral do máximo. Essas trajetórias representam uma oportunidade de arremeter diretamente ao longo da crista em direção ao máximo. Tais trajetórias são chamadas *direções padrão*.

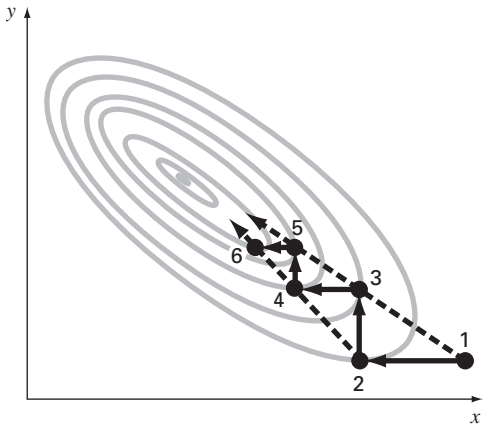


FIGURA 14.3
Representação gráfica de como uma busca unidimensional pode ser conduzida.

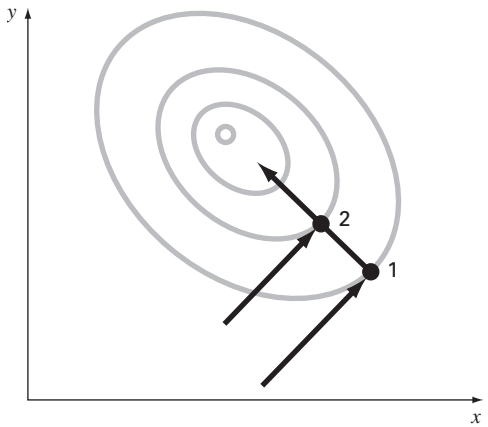


FIGURA 14.4
Direções conjugadas.

Existem algoritmos formais que capitalizam a idéia das direções padrão para encontrar o valor ótimo eficientemente. O mais bem conhecido deles é chamado *método de Powell*. Ele é baseado na observação (ver Figura 14.4) de que os pontos 1 e 2 são obtidos por buscas unidimensionais na mesma direção mas de diferentes pontos de partida — portanto, a reta determinada por 1 e 2 irá na direção do máximo. Tais retas são chamadas de *direções conjugadas*.

De fato, pode ser demonstrado que, se $f(x, y)$ for uma função quadrática, uma busca seqüencial ao longo das direções conjugadas convergirá exatamente em um número finito de passos, independentemente do ponto de partida. Uma vez que uma função não-linear, em geral, pode ser razoavelmente aproximada por uma função quadrática, métodos baseados nas direções conjugadas são geralmente bastante eficientes e são, de fato, quadraticamente convergentes quando se aproximam do ponto ótimo.

Vamos implementar graficamente uma versão simplificada do método de Powell para encontrar o máximo de

$$f(x, y) = c - (x - a)^2 - (y - b)^2$$

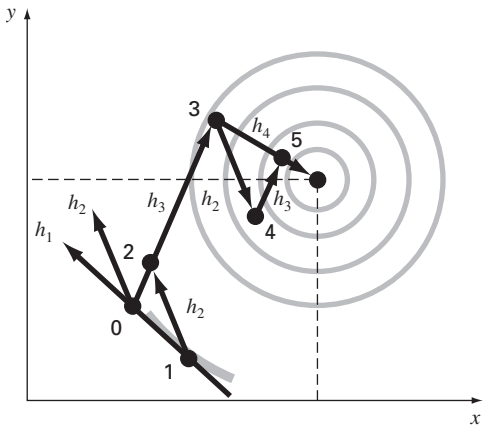


FIGURA 14.5
Método de Powell.

onde a, b e c são constantes positivas. Essa equação resulta em curvas de nível circulares no plano x, y como mostrado na Figura 14.5.

Comece a busca no ponto 0 com direções iniciais h_1 e h_2. Observe que h_1 e h_2 não são necessariamente direções conjugadas. Do zero, mova-se ao longo de h_1 até que um máximo seja localizado no ponto 1. Então, faça a busca a partir do ponto 1 na direção h_2 até encontrar o ponto 2. Em seguida, forme uma nova direção de busca h_3 entre os pontos 0 e 2. Faça a busca ao longo dessa direção até que o máximo no ponto 3 seja localizado. Então, faça a busca desde o ponto 3 na direção h_2 até que o máximo no ponto 4 seja localizado. Do ponto 4 se chega ao ponto 5 buscando novamente ao longo da direção h_3. Agora, observe que os pontos 5 e 3 foram localizados buscando na direção h_3, a partir de dois pontos diferentes. Powell mostrou que h_4 (formada pelos pontos 3 e 5) e h_3 são direções conjugadas. Assim, fazer uma busca a partir do ponto 5 na direção h_4 leva diretamente para o máximo.

O método de Powell pode ser refinado para se tornar mais eficiente, mas o algoritmo formal está além do escopo deste texto. Entretanto, é um método eficiente, quadraticamente convergente, que não exige o cálculo das derivadas.

14.2 MÉTODOS GRADIENTE

Como o nome implica, os *métodos gradiente* usam explicitamente informação sobre a derivada para gerar um algoritmo eficiente para localizar o ponto ótimo. Antes de descrever abordagens específicas, deve-se primeiro rever alguns conceitos e operações matemáticos chave.

14.2.1 GRADIENTES E HESSIANAS

Do cálculo, lembre-se de que a primeira derivada de uma função unidimensional fornece a inclinação da função ou a tangente à função que está sendo derivada. Do ponto de vista da otimização, essa é uma informação útil. Por exemplo, se a inclinação for positiva, isso diz que o aumento da variável independente levará a um valor maior da função que está sendo explorada.

Ainda do cálculo, lembre-se de que a primeira derivada pode dizer quando se atinge um ponto ótimo, uma vez que esse é um ponto em que a derivada é nula. Além disso, o sinal da segunda derivada diz se foi atingido um ponto de mínimo (derivada segunda positiva) ou um de máximo (derivada segunda negativa).

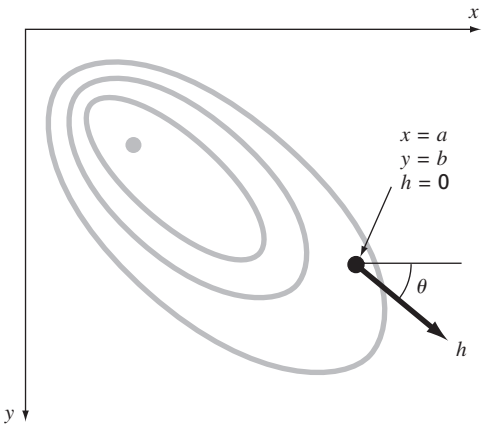


FIGURA 14.6
O gradiente direcional é definido ao longo de um eixo h que forma um ângulo θ com o eixo x.

Essas idéias foram úteis nos algoritmos de busca unidimensional explorados no capítulo anterior. Entretanto, para entender completamente a busca multidimensional, deve-se primeiro entender como a primeira e a segunda derivadas são expressas no contexto multidimensional.

O Gradiente. Suponha que tenhamos uma função bidimensional $f(x, y)$. Um exemplo pode ser sua altura em uma montanha como função da sua posição. Suponha que você está em um local específico (a, b) da montanha e que quer saber a inclinação em uma direção arbitrária. Uma maneira de definir a direção é ao longo de um novo eixo h que forma um ângulo θ com o eixo x (Figura 14.6). A altura ao longo desse novo eixo pode ser considerada como uma nova função $g(h)$. Se você define sua posição como sendo a origem desse eixo (isto é, $h = 0$), a inclinação nessa direção será determinada por $g'(0)$. A inclinação, que é chamada de *derivada direcional*, pode ser calculada a partir das derivadas parciais ao longo de x e y por

$$g'(0) = \frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\text{sen}\,\theta \tag{14.1}$$

onde as derivadas parciais são calculadas em $x = a$ e $y = b$.

Supondo que o seu objetivo seja atingir a maior altura com o próximo passo, a próxima questão lógica será: qual a direção de aclive máximo? A resposta é suprida naturalmente pelo que se chama matematicamente de *gradiente*, que é definido por

$$\nabla f = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} \tag{14.2}$$

Esse vetor é também conhecido como "del f". Ele representa a derivada direcional de $f(x, y)$ no ponto $x = a$ e $y = b$.

A notação vetorial fornece uma maneira concisa de generalizar o gradiente para n dimensões, como

$$\nabla f(\mathbf{x}) = \begin{Bmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{x}) \\ \dfrac{\partial f}{\partial x_2}(\mathbf{x}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\mathbf{x}) \end{Bmatrix}$$

Como usar o gradiente? Para o problema da escalada da montanha, se o interesse é ganhar altura tão rapidamente quanto possível, o gradiente diz em qual direção mover-se localmente e quanta altura será ganha ao se fazer isso. Observe, entretanto, que essa estratégia não leva necessariamente a um caminho direto ao pico! Essas idéias serão discutidas com maior profundidade mais adiante neste capítulo.

EXEMPLO 14.2 Usando o Gradiente para Encontrar o Caminho de Aclive Máximo

Enunciado do Problema. Use o gradiente para calcular a direção de aclive máximo para a função

$$f(x, y) = xy^2$$

no ponto (2, 2). Suponha que o eixo x positivo aponta para o leste e que o eixo y positivo aponta para o norte.

Solução. Primeiro, nossa altura pode ser determinada por

$$f(4, 2) = 2(2)^2 = 8$$

Em seguida, as derivadas parciais podem ser calculadas

$$\frac{\partial f}{\partial x} = y^2 = 2^2 = 4$$

$$\frac{\partial f}{\partial y} = 2xy = 2(2)(2) = 8$$

o que pode ser usado para determinar o gradiente,

$$\nabla f = 4\mathbf{i} + 8\mathbf{j}$$

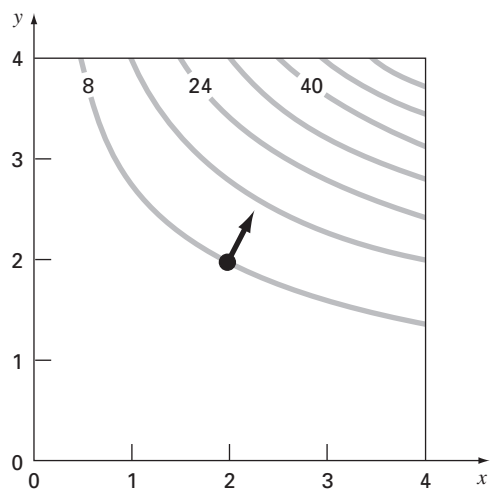
Esse vetor pode ser esboçado em um mapa topográfico da função, como mostrado na Figura 14.7. Isso imediatamente diz que a direção que se deve tomar é

$$\theta = \text{tg}^{-1}\left(\frac{8}{4}\right) = 1{,}107 \text{ radianos } (= 63{,}4^\circ)$$

relativo ao eixo x. A inclinação nessa direção, que é o módulo de ∇f, pode ser calculada por

$$\sqrt{4^2 + 8^2} = 8{,}944$$

FIGURA 14.7
A flecha segue a direção de aclive máximo calculado com o gradiente.



Então, durante o primeiro passo, serão ganhas inicialmente 8,944 unidades de aumento na elevação para uma unidade de distância avançada ao longo do caminho de maior inclinação. Observe que a Equação (14.1) fornece o mesmo resultado,

$$g'(0) = 4\ \cos(1{,}107) + 8\ \text{sen}(1{,}107) = 8{,}944$$

Observe que, para qualquer outra direção, como $\theta = 1{,}107/2 = 0{,}5235$, $g'(0) = 4\ \cos(0{,}5235) + 8\ \text{sen}(0{,}5235) = 7{,}608$, o que é menor.

Ao se mover para a frente, tanto a direção quanto o módulo da maior inclinação mudam. Essas mudanças podem ser quantificadas a cada passo usando-se o gradiente, e sua direção de escalada, modificada de acordo com ele.

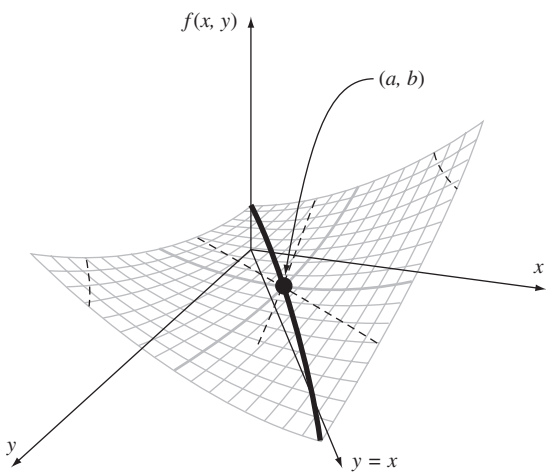
Uma percepção final pode ser adquirida inspecionando-se a Figura 14.7. Como indicado, a direção de aclive máximo é perpendicular, ou *ortogonal*, à curva de nível da elevação nas coordenadas (2, 2). Essa é uma característica geral do gradiente.

Além de definir o caminho de maior inclinação, a primeira derivada pode também ser usada para distinguir se um ponto ótimo foi atingido. Como no caso de uma função unidimensional, se as derivadas parciais com relação a x e y são ambas nulas, é possível que um ponto ótimo bidimensional tenha sido atingido.

A Hessiana. Para problemas unidimensionais, tanto a primeira quanto a segunda derivada fornecem informações valiosas para a busca do ponto ótimo. A primeira derivada (a) mostra a trajetória de maior inclinação de uma função e (b) diz que foi atingido um ponto ótimo. Uma vez atingido o ponto ótimo, a segunda derivada diz se você está em um máximo [$f''(x)$ negativa] ou em um mínimo [$f''(x)$ positivo]. Nos parágrafos anteriores, foi ilustrado como o gradiente fornece a melhor trajetória local para problemas multidimensionais. Agora se examinará como a segunda derivada é usada em tais contextos.

Você poderia esperar que, se as segundas derivadas parciais com relação a x e a y forem ambas negativas, então deve ter sido atingido um máximo. A Figura 14.8 mostra uma função na qual isso não é verdade. O ponto (a, b) desse gráfico parece ser um mínimo quando observado ao longo da dimensão x ou da dimensão y. Em ambos os casos, a segunda derivada parcial é positiva. Entretanto, se a função for observada ao longo da reta $y = x$, pode ser visto que ocorre um máximo nesse mesmo ponto. Essa forma é chamada *sela* e claramente não ocorrem nem um máximo nem um mínimo nesse ponto.

FIGURA 14.8
Um ponto de sela ($x = a$ e $y = b$). Observe que, quando se observa uma curva ao longo das direções x e y, a função parece atingir um mínimo (derivada segunda positiva), enquanto que, ao longo do eixo $x = y$, ela é côncava para baixo (derivada segunda negativa).



A ocorrência de um máximo ou de um mínimo envolve não apenas as derivadas parciais com relação a x e y, mas também a segunda derivada parcial com relação a x e y. Supondo que as derivadas parciais sejam contínuas na vizinhança do ponto sendo estudado, a seguinte quantidade pode ser calculada:

$$|H| = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 \tag{14.3}$$

Três casos, entre outros, podem ocorrer:

- Se $|H| > 0$ e $\partial^2 f/\partial x^2 > 0$, então $f(x, y)$ tem um mínimo local.
- Se $|H| > 0$ e $\partial^2 f/\partial x^2 < 0$, então $f(x, y)$ tem um máximo local.
- Se $|H| < 0$, então $f(x, y)$ tem um ponto de sela.

A quantidade $|H|$ é igual ao determinante da matriz formada pelas segundas derivadas,[1]

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix} \tag{14.4}$$

Essa matriz é formalmente conhecida como hessiana de f.

Além de fornecer uma maneira de discernir se uma função multidimensional atingiu um ponto ótimo, a hessiana tem outros usos em otimização (por exemplo, para a forma multidimensional do método de Newton). Em particular, ela permite que a busca inclua a curvatura de segunda ordem para se atingir resultados melhores.

Aproximação por Diferenças Finitas. Deve-se mencionar que, para os casos nos quais é difícil ou inconveniente calcular analiticamente, o gradiente e o determinante da hessiana podem ser calculados numericamente. Em muitos casos, é utilizada a abordagem introduzida na Seção 6.3.3 para o método da secante modificada. Isto é, as variáveis independentes podem ser perturbadas ligeiramente para gerar as derivadas parciais necessárias. Por exemplo, se abordagem baseada em diferenças centradas for adotada, elas podem ser calculadas como

$$\frac{\partial f}{\partial x} = \frac{f(x + \delta x, y) - f(x - \delta x, y)}{2\delta x} \tag{14.5}$$

$$\frac{\partial f}{\partial y} = \frac{f(x, y + \delta y) - f(x, y - \delta y)}{2\delta y} \tag{14.6}$$

$$\frac{\partial^2 f}{\partial x^2} = \frac{f(x + \delta x, y) - 2f(x, y) - f(x - \delta x, y)}{\delta x^2} \tag{14.7}$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{f(x, y + \delta y) - 2f(x, y) - f(x, y - \delta y)}{\delta y^2} \tag{14.8}$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{f(x + \delta x, y + \delta y) - f(x + \delta x, y - \delta y) - f(x - \delta x, y + \delta y) + f(x - \delta x, y - \delta y)}{4\delta x \delta y} \tag{14.9}$$

onde δ é um pequeno valor fracionário.

Observe que os métodos utilizados em pacotes de software comerciais também empregam diferenças progressivas. Além disso, eles são usualmente mais complicados que

[1] Observe que $\partial^2 f/(\partial x \partial y) = \partial^2 f/(\partial y \partial x)$ quando estas funções forem conínuas.

as aproximações listadas nas Equações (14.5) a (14.9). Por exemplo, a biblioteca IMSL baseia a perturbação no épsilon de máquina. Dennis e Schnabel (1996) fornecem mais detalhes sobre essa abordagem.

Independentemente de como a aproximação é implementada, o ponto importante é que você pode ter a opção de calcular analiticamente o gradiente e/ou a hessiana. Algumas vezes isso é uma tarefa árdua, mas o desempenho do algoritmo se beneficia o suficiente para fazer que o seu esforço valha a pena. As derivadas na forma fechada serão exatas, porém, ainda mais importante, você reduzirá o número de cálculos da função. Esse último ponto pode ter um impacto crítico no tempo de execução.

Por outro lado, você exercitará freqüentemente a opção de ter as quantidades computadas internamente usando abordagens numéricas. Em muitos casos, o desempenho será bastante adequado e você se livrará da dificuldade de derivações parciais numerosas. Este é o caso de otimizadores usados em certas planilhas e pacotes de software matemáticos (por exemplo, o Excel). Em tais casos, nem mesmo é dada a opção de entrar com o gradiente ou a hessiana calculados analiticamente. Entretanto, para problemas de tamanho pequeno ou moderado, essa não é uma grande limitação.

14.2.2 Método do Aclive Máximo

Uma estratégia óbvia para escalar o morro seria determinar a inclinação máxima no ponto de partida e então começar a andar nessa direção. Mas surge claramente outro problema quase de imediato. Exceto se você tiver realmente muita sorte e começar em uma crista que aponta diretamente para o pico, assim que se mover, seu caminho divergirá da direção de aclive máximo.

Reconhecendo esse fato, você poderá adotar a seguinte estratégia. Caminhe uma distância curta na direção do gradiente. Então pare, recalcule o gradiente e ande novamente uma distância curta. Repetindo esse processo, você pode eventualmente chegar ao topo do monte.

Embora a estratégia pareça superficialmente correta, não é muito prática. Em particular, os cálculos contínuos do gradiente podem ser computacionalmente exigentes. Uma abordagem preferível envolve o movimento ao longo de um caminho fixo na direção do gradiente até que $f(x, y)$ pare de crescer, isto é, fique nivelada na direção da caminhada. Esse ponto de parada será o ponto de partida onde ∇f é recalculado e uma nova direção será seguida. O processo se repete até que o pico seja atingido. Esse processo é chamado de *método do aclive máximo*.[2] Essa é a mais direta das técnicas de busca baseadas no gradiente. A idéia básica por trás de tal abordagem é mostrada na Figura 14.9.

O começo é em um ponto inicial (x_0, y_0) rotulado como "0" na figura. Nesse ponto, determina-se a direção do aclive máximo, isto é, o gradiente. Então, faz-se uma busca na direção do gradiente, h_0, até se encontrar um máximo, que é rotulado como "1" na figura. O processo é então repetido.

Assim, o problema se reduz simplesmente a duas partes: (1) determinar a "melhor" direção de busca e (2) determinar o "melhor valor" ao longo dessa direção de busca. Como se verá, a eficácia dos vários algoritmos descritos nas páginas que virão depende do quão perspicaz se é em ambas as partes.

Por enquanto, o método do aclive máximo usa o gradiente para escolher a "melhor" direção. Já se mostrou como o gradiente é calculado no Exemplo 14.1. Agora, antes de examinar como o algoritmo procede para localizar o máximo ao longo da direção de maior inclinação, deve-se parar para explorar como transformar a função de x e y em uma função de h ao longo da direção do gradiente.

Começando em x_0, y_0 as coordenadas de um ponto qualquer na direção do gradiente podem ser expressas como

$$x = x_0 + \frac{\partial f}{\partial x} h \tag{14.10}$$

[2] Por causa da nossa ênfase na maximização, usamos a terminologia de aclive máximo. A mesma abordagem pode também ser usada para minimização – nesse caso, a terminologia utilizada seria declive máximo.

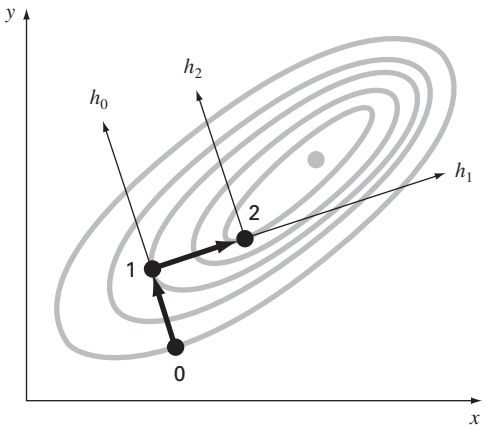


FIGURA 14.9
Uma descrição gráfica do método do máximo aclive.

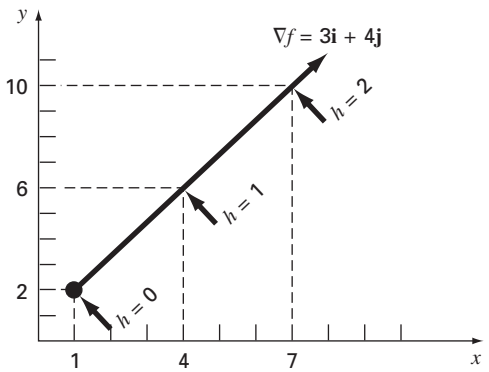


FIGURA 14.10
A relação entre uma direção arbitrária h e as coordenadas x e y.

$$y = y_0 + \frac{\partial f}{\partial y}h \tag{14.11}$$

onde h é uma distância ao longo do eixo h. Por exemplo, suponha que $x_0 = 1$, $y_0 = 2$ e $\nabla f = 3\mathbf{i} + 4\mathbf{j}$, como mostrado na Figura 14.10. As coordenadas de um ponto ao longo do eixo h são dadas por

$$x = 1 + 3h \tag{14.12}$$

$$y = 2 + 4h \tag{14.13}$$

O exemplo seguinte ilustra como é possível usar essa transformação para converter uma função bidimensional em x e y em uma função unidimensional em h.

EXEMPLO 14.3 Desenvolvendo uma Função 1-D ao Longo da Direção do Gradiente

Enunciado do Problema. Suponha que se tenha a seguinte função bidimensional

$$f(x, y) = 2xy + 2x - x^2 - 2y^2$$

Desenvolva uma versão unidimensional dessa equação ao longo da direção do gradiente no ponto $x = -1$ e $y = 1$.

Solução. As derivadas parciais podem ser calculadas no ponto $(-1, 1)$,

$$\frac{\partial f}{\partial x} = 2y + 2 - 2x = 2(1) + 2 - 2(-1) = 6$$

$$\frac{\partial f}{\partial y} = 2x - 4y = 2(-1) - 4(1) = -6$$

Portanto, o vetor gradiente é

$$\nabla f = 6\mathbf{i} - \mathbf{6j}$$

Para encontrar o máximo, deve-se procurar ao longo da direção do gradiente, isto é, ao longo do eixo h, caminhando na direção desse vetor. A função pode ser expressa ao longo desse eixo como

$$f\left(x_0 + \frac{\partial f}{\partial x}h, y_0 + \frac{\partial f}{\partial y}h\right) = f(-1 + 6h, 1 - 6h)$$

$$= 2(-1 + 6h)(1 - 6h) + 2(-1 + 6h) - (-1 + 6h)^2 - 2(1 - 6h)^2$$

onde as derivadas parciais são calculadas em $x = -1$ e $y = 1$.

Combinando esses termos, tem-se uma função unidimensional $g(h)$ que descreve $f(x, y)$ ao longo do eixo h,

$$g(h) = -180h^2 + 72h - 7$$

Agora que foi desenvolvida uma função ao longo do caminho do máximo aclive, pode-se explorar como responder à segunda questão – ou seja, quanto ao longo desse caminho deve-se andar? Uma abordagem poderia ser mover-se ao longo desse caminho até encontrar o máximo dessa função. Chamaremos a posição desse máximo de h^*. Esse é o valor do passo que maximiza g (e, portanto, f) na direção do gradiente. O problema é equivalente a encontrar o máximo de uma função de uma única variável h. Isso pode ser feito usando-se diferentes técnicas de busca unidimensionais como as discutidas no Capítulo 13. Assim, a procura por um ponto ótimo de uma função bidimensional é convertida em uma busca unidimensional na direção do gradiente.

Esse método é chamado aclive *máximo* quando um tamanho de passo arbitrário h for usado. Se for encontrado o valor de um único passo h^* que leve diretamente a um máximo ao longo da direção do gradiente, esse método é chamado de *aclie máximo ótimo.*

EXEMPLO 14.4 Aclive Máximo Ótimo

Enunciado do Problema. Maximize a seguinte função:

$$f(x, y) = 2xy + 2x - x^2 - 2y^2$$

usando as aproximações iniciais $x = -1$ e $y = 1$.

Solução. Pode-se em primeiro lugar gerar uma solução analítica, já que a função é tão simples. Para fazer isso, as derivadas parciais podem ser calculadas por

$$\frac{\partial f}{\partial x} = 2y + 2 - 2x = 0$$

$$\frac{\partial f}{\partial y} = 2x - 4y = 0$$

Esse par de equações pode ser resolvido para se determinar o ponto ótimo, $x = 2$ e $y = 1$. As derivadas parciais de segunda ordem podem ser determinadas e calculadas no ponto ótimo,

$$\frac{\partial^2 f}{\partial x^2} = -2$$

$$\frac{\partial^2 f}{\partial y^2} = -4$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 2$$

e o determinante da hessiana é calculado por [Equação (14.3)]

$$|H| = -2(-4) - 2^2 = 4$$

Portanto, como $|H| > 0$ e $\partial^2 f/\partial x^2 < 0$, o valor da função $f(2, 1)$ é um máximo.

Agora vamos implementar o aclive máximo. Lembre-se de que, no final do Exemplo 14.3, já tinham sido implementados os passos iniciais do problema gerando-se

$$g(h) = -180h^2 + 72h - 7$$

Agora, como se trata de uma simples parábola, pode-se localizar diretamente o máximo (isto é, $h = h^*$) resolvendo-se o problema,

$$g'(h^*) = 0$$
$$-360h^* + 72 = 0$$
$$h^* = 0{,}2$$

Isso significa que, andando-se ao longo do eixo x, $g(h)$ atinge um valor mínimo quando $h = h^* = 0{,}2$. Tal resultado pode ser colocado de volta nas Equações (14.10) e (14.11) para determinar as coordenadas (x, y) correspondentes a esse ponto.

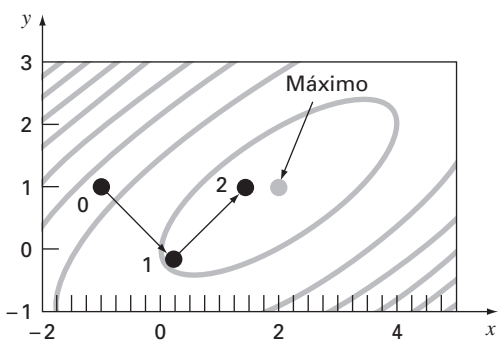
$$x = -1 + 6(0{,}2) = 0{,}2$$

$$y = 1 - 6(0{,}2) = -0{,}2$$

Esse passo é mostrado na Figura 14.11 como o movimento do ponto 0 para o ponto 1.

O segundo passo é implementado simplesmente repetindo-se esse procedimento.

FIGURA 14.11
O método do aclive máximo ótimo.



Primeiro, as derivadas parciais podem ser calculadas nesse novo ponto de partida (0,2, −0,2), resultando

$$\frac{\partial f}{\partial x} = 2(-0,2) + 2 - 2(0,2) = 1,2$$

$$\frac{\partial f}{\partial y} = 2(0,2) - 4(-0,2) = 1,2$$

Portanto, o vetor gradiente é

$$\nabla f = 1,2\,\mathbf{i} + 1,2\,\mathbf{j}$$

Isso significa que a direção de máxima inclinação agora aponta para cima e para a direita, em um ângulo de 45° com o eixo x (ver Figura 14.11). As coordenadas ao longo desse novo eixo h podem ser expressas como

$$x = 0,2 + 1,2h$$

$$y = -0,2 + 1,2h$$

Substituindo esses valores na função, obtém-se

$$f(0,2 + 1,2h, -0,2 + 1,2h) = g(h) = -1,44h^2 + 2,88h + 0,2$$

O passo h^* que leva ao máximo ao longo da direção de busca pode então ser calculado como

$$g'(h^*) = -2,88h^* + 2,88 = 0$$
$$h^* = 1$$

Esses resultados podem ser colocados de volta nas Equações (14.10) e (14.11) para se determinar as coordenadas (x, y) correspondentes a esse novo ponto

$$x = 0,2 + 1,2(1) = 1,4$$

$$y = -0,2 + 1,2(1) = 1$$

Como mostrado na Figura 14.11, movemo-nos para novas coordenadas, rotuladas como ponto 2 no desenho, e fazendo isso nos movemos para mais perto do máximo. Essa abordagem pode ser repetida com o resultado final convergindo para a solução analítica $x = 2$ e $y = 1$.

Pode ser mostrado que o método do aclive máximo é linearmente convergente. Ademais, tende a se mover muito devagar ao longo de cristas longas e estreitas. Isso ocorre porque o novo gradiente em cada ponto de máximo será perpendicular à direção original. Assim, a técnica precisa de muitos passos pequenos atravessando a rota direta para o pico. Logo, embora seja confiável, há outras abordagens que convergem muito mais rapidamente, particularmente na vizinhança de um ótimo. O restante dessa seção é dedicado a tais métodos.

14.2.3 Abordagens Gradiente Avançadas

Método do Gradiente Conjugado (Fletcher-Reeves). Na Seção 14.1.2, foi visto como as direções conjugadas do método de Powell melhoraram a eficiência da busca unidimensional. De maneira similar, pode-se também melhorar a convergência linear do aclive máximo ascendente usando-se gradientes conjugados. Na realidade, é possível mostrar que um método de otimização que utiliza gradientes conjugados para definir as direções de busca é quadraticamente convergente. Isso também garante que o método otimizará uma função quadrática exatamente em um número finito de passos, independentemente do ponto inicial. Já que muitas funções bem comportadas podem ser aproximadas razoavelmente bem por uma função quadrática nas vizinhanças do ponto ótimo,

uma abordagem quadraticamente convergente é normalmente muito eficiente próximo a um ponto ótimo.

Viu-se como, começando com duas direções arbitrárias de busca, o método de Powell produziu novas direções de busca conjugadas. Esse método é quadraticamente convergente e não requer informação sobre o gradiente. Por outro lado, se o cálculo das derivadas for prático, pode-se pensar em algoritmos que combinem as idéias do declive máximo e direções conjugadas para atingir um desempenho inicial robusto e uma convergência rápida quando a técnica gravita em direção ao ponto ótimo. O *algoritmo de gradiente conjugado de Fletcher-Reeves* modifica o método do aclive máximo impondo a condição de que sucessivas direções de busca do gradiente sejam mutuamente conjugadas. A demonstração e o algoritmo estão além do escopo deste livro, mas são descritos por Rao (1996).

Método de Newton. O método de Newton para uma única variável (ver Seção 13.3) pode ser estendido para o caso de várias variáveis. Escreva uma série de Taylor de segunda ordem para $f(\mathbf{x})$ próximo de $\mathbf{x} = \mathbf{x}_i$,

$$f(\mathbf{x}) = f(\mathbf{x}_i) + \nabla f^T(\mathbf{x}_i)(\mathbf{x} - \mathbf{x}_i) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_i)^T H_i (\mathbf{x} - \mathbf{x}_i)$$

onde H_i é a matriz hessiana. No ponto de mínimo,

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_j} = 0 \qquad \text{para } j = 1, 2, \ldots, n$$

Assim,

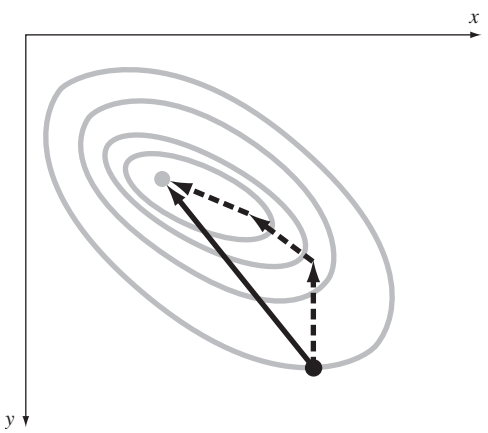
$$\nabla f = \nabla f(\mathbf{x}_i) + H_i(\mathbf{x} - \mathbf{x}_i) = 0$$

Se H for não singular

$$\mathbf{x}_{i+1} = \mathbf{x}_i - H_i^{-1} \nabla f \tag{14.14}$$

o que pode ser mostrado que converge quadraticamente próximo ao ponto ótimo. Esse método tem melhor desempenho que o do aclive máximo (ver Figura 14.12). No entanto, observe que ele requer tanto o cálculo das derivadas de segunda ordem quanto a inversão de matrizes a cada iteração. Assim, não é muito útil, na prática, para funções com um número grande de variáveis. Além disso, o método de Newton pode não ser convergente se o ponto de partida não estiver próximo do ponto ótimo.

FIGURA 14.12
Quando o ponto de partida está próximo do ponto de ótimo, seguir o gradiente pode ser ineficiente. O método de Newton tenta a busca ao longo de um caminho que vai diretamente ao ponto ótimo (linha sólida).



Método de Marquardt. O método do aclive máximo aumenta o valor da função mesmo quando o ponto de partida está longe do ótimo. Por outro lado, o método de Newton que acabamos de descrever, converge rapidamente próximo do máximo. O *método de Marquardt* usa o método do declive máximo quando $\mathbf{x}$ estiver longe de $\mathbf{x}^*$, e o método de Newton quando x estiver perto de um ponto ótimo. Isso se consegue modificando-se a diagonal da hessiana na Equação (14.14),

$$\tilde{H}_i = H_i + \alpha_i I$$

onde α_i é uma constante positiva e I é a matriz identidade. No início do procedimento, α_i é suposto ser grande e

$$\tilde{H}_i^{-1} \approx \frac{1}{\alpha_i} I$$

o que reduz a Equação (14.14) ao método da máxima inclinação ascendente. Conforme a iteração prossegue, α_i se aproxima de zero e o método se torna o método de Newton.

Portanto, o método de Marquardt oferece o melhor dos dois mundos: caminha de forma lenta, mas confiável, para valores iniciais insatisfatórios, mas acelera rapidamente quando se aproxima do ponto ótimo. Infelizmente, o método ainda requer o cálculo da hessiana e a inversão de matrizes a cada iteração.

Deve-se observar que o método de Marquardt é usado primordialmente para problemas de mínimos quadrados não-lineares. Por exemplo, a biblioteca IMSL tem uma subrotina com esse propósito.

Métodos Quasi-Newton. Os *métodos quasi-Newton*, ou *de variável métrica*, procuram fazer uma estimativa de caminho direto ao ponto ótimo de forma parecida com o método de Newton. Entretanto, observe que a matriz hessiana na Equação (14.14) é composta das derivadas de segunda ordem de f que variam de passo a passo. Os métodos quasi-Newton tentam evitar essas dificuldades aproximando H por uma outra matriz A usando apenas as derivadas parciais de primeira ordem de f. A abordagem envolve começar com uma aproximação inicial de H^{-1} e atualizá-la e melhorá-la a cada iteração. Os métodos são chamados quasi-Newton porque não é usada a verdadeira hessiana, mas, sim, uma aproximação. Assim, têm-se duas aproximações sendo feitas simultaneamente: (1) aproximação original pela série de Taylor e (2) a aproximação da hessiana.

Existem dois métodos principais nessa categoria: os algoritmos de *Davidson-Fletcher-Powell* (DFP) e de *Broyden-Fletcher-Goldfarb-Shanno* (BFGS). Eles são similares, exceto por detalhes relativos a como lidar com erros de arredondamento e aspectos de convergência. O BFGS é geralmente reconhecido como sendo superior na maioria dos casos. Rao (1996) traz detalhes e enunciados formais tanto para o algoritmo DFP quanto para o BFGS.

PROBLEMAS

14.1 Repita o Exemplo 14.2 para a seguinte função, no ponto (0,8, 1,2).

$$f(x, y) = 2xy + 1{,}5y - 1{,}25x^2 - 2y^2 + 5$$

14.2 Encontre a derivada direcional de

$$f(x, y) = x^2 + 2y^2$$

em $x = 2$ e $y = 2$ na direção $h = 2i + 3j$.

14.3 Encontre o vetor gradiente e a matriz hessiana para cada uma das seguintes funções:

(a) $f(x, y) = 3xy^2 + 2e^{xy}$

(b) $f(x, y, z) = 2x^2 + y^2 + z^2$

(c) $f(x, y) = \ln(x^2 + 3xy + 2y^2)$

14.4 É dada a função

$$f(x, y) = 2{,}25xy + 1{,}75y - 1{,}5x^2 - 2y^2$$

Determine e resolva um sistema de equações algébricas lineares que maximiza $f(x)$. Observe que isso é feito igualando-se as derivadas parciais de f com relação a x e a y a zero.

14.5

(a) Comece com uma aproximação inicial $x = 1$ e $y = 1$ e faça duas aplicações do método do aclive máximo na função $f(x, y)$ do Problema 14.4.

(b) Construa um gráfico a partir dos resultados de **(a)** mostrando o caminho de busca.

14.6 Encontre o valor mínimo de

$$f(x, y) = (x - 3)^2 + (y - 2)^2$$

começando em $x = 1$ e $y = 1$, usando o método do declive máximo, com o critério de parada de $\varepsilon_s = 1\%$. Explique seus resultados.

14.7 Faça uma iteração do método do aclive máximo para localizar o máximo de

$$f(x, y) = 4x + 2y + x^2 - 2x^4 + 2xy - 3y^2$$

usando as aproximações iniciais $x = 0$ e $y = 0$. Use a bissecção para encontrar o tamanho do passo ótimo na direção de busca do gradiente.

14.8 Faça uma iteração do método do gradiente ótimo no declive máximo para localizar o mínimo de

$$f(x, y) = -8x + x^2 + 12y + 4y^2 - 2xy$$

usando aproximações iniciais $x = 0$ e $y = 0$.

14.9 Desenvolva um programa usando uma linguagem de programação ou linguagem de macro para implementar o método da busca aleatória. Projete o subprograma de modo que ele seja expressamente preparado para localizar o máximo. Teste o programa com $f(x, y)$ do Problema 14.7. Use o intervalo de -2 a 2 para x e y.

14.10 A busca reticulada é outra abordagem por força bruta para otimização. A versão bidimensional é mostrada na Figura P14.10. As dimensões x e y são divididas em incrementos para criar um reticulado. A função é então calculada em cada nó do reticulado. Quanto maior a densidade do reticulado, maior é a probabilidade de localizar o ponto ótimo.

Desenvolva um programa usando uma linguagem de programação ou uma linguagem de macro para implementar o método da busca reticulada. Projete um programa que seja expressamente preparado para localizar o máximo. Teste o programa com o mesmo problema do Exemplo 14.1.

14.11 Desenvolva uma equação unidimensional na direção do gradiente de pressão, no ponto (4, 2). A função que dá a pressão é

$$f(x, y) = 6x^2y - 9y^2 - 8x^2$$

14.12 A função temperatura é

$$f(x, y) = 2x^3y^2 - 7xy + x^2 + 3y$$

Desenvolva uma função unidimensional na direção do gradiente de temperatura, no ponto (1, 1).

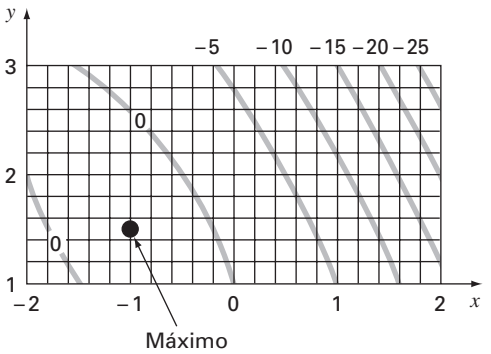


Figura P14.10
A busca reticulada.

CAPÍTULO 15

Otimização com Restrições

Este capítulo trata dos problemas da otimização nos quais restrições entram em jogo. Em primeiro lugar serão abordados problemas em que tanto a função objetivo quanto as restrições são lineares. Para esses casos, existem métodos especiais que exploram a linearidade das funções envolvidas — eles são chamados métodos de programação linear, e os algoritmos associados resolvem com bastante eficiência problemas de grande porte, com milhares de variáveis e restrições. São usados em uma grande variedade de problemas de engenharia e de gerenciamento.

Então, retornaremos brevemente ao problema mais geral da otimização com restrições não-lineares. Finalmente, será apresentada uma visão geral de como os pacotes e bibliotecas podem ser utilizados na otimização.

15.1 PROGRAMAÇÃO LINEAR

A *programação linear* (PL) é uma abordagem para otimização que trabalha para encontrar um objetivo desejado, como a maximização de lucros ou a minimização de custos, na presença de restrições tais como recursos limitados. O termo *linear* indica que as funções que representam tanto o objetivo quanto as restrições são lineares. O termo *programação* não significa "programação de computadores", mas, em vez disso, indica "planejamento" ou "agendamento" (Revelle et al., 1997).

15.1.1 Forma Padrão

O problema básico da programação linear consiste em duas partes principais: a função objetivo e um conjunto de restrições. Para um problema de maximização, a função objetivo é geralmente expressa como

$$\text{Maximize } Z = c_1x_1 + c_2x_2 + \cdots + c_nx_n \tag{15.1}$$

onde c_j é o lucro de cada unidade da j-ésima atividade que foi empreendida e x_j é a quantidade da j-ésima atividade. Assim, o valor da função objetivo, Z, é o lucro total devido ao número total de atividades, n.

As restrições podem ser representadas em geral por

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i \tag{15.2}$$

onde a_{ij} é a quantidade do i-ésimo recurso que é consumido por cada unidade da j-ésima atividade e b_i é a quantidade do i-ésimo recurso disponível. Isto é, os recursos são limitados.

O segundo tipo geral de restrições especifica que todas as atividades devem ter um valor positivo,

$$x_i \geq 0 \tag{15.3}$$

Nesse contexto, isso expressa a noção realística de que, para alguns problemas, uma atividade negativa é fisicamente impossível (por exemplo, não é possível produzir um número negativo de artefatos).

Conjuntamente, a função objetivo e as restrições especificam o problema de programação linear. Elas dizem que estamos tentando maximizar o lucro para um número de atividades sob a restrição de que tais atividades utilizam uma quantidade finita de recursos. Antes de mostrar como esse resultado pode ser obtido, primeiro será desenvolvido um exemplo.

EXEMPLO 15.1 Preparando um Problema de PL

Enunciado do Problema. O problema a seguir é desenvolvido a partir da área da engenharia química ou de petróleo. Entretanto, ele é relevante para todas as áreas da engenharia que lidam com a produção de produtos com recursos limitados.

Suponha que uma indústria de processamento de gás receba uma quantidade fixa de gás bruto por semana. Esse gás bruto é processado em dois tipos de gás para aquecimento, com qualidade regular e premium. Esses tipos de gás estão em alta demanda (isto é, têm garantia de venda) e fornecem lucros diferentes para a companhia. No entanto, sua produção envolve restrições de tempo e armazenamento. Por exemplo, apenas um tipo de gás pode ser produzido por vez e as ferramentas de processamento estão disponíveis apenas por 80 horas/semana. Além disso, existe restrição de armazenamento local para os produtos. Todos esses fatores são listados a seguir (observe que uma tonelada métrica, ou ton, é igual a 1.000 kg):

	Produto		
Recurso	**Regular**	**Premium**	**Disponibilidade do Recurso**
Gás bruto	7 m^3/ton	11 m^3/ton	77 m^3/semana
Tempo de produção	10 h/ton	8 hr/ton	80 h/semana
Armazenamento	9 tons	6 tons	
Lucro	150/ton	175/ton	

Desenvolva uma formulação em programação linear para maximizar os lucros dessa operação.

Solução. O engenheiro operando essa indústria deve decidir quanto de cada tipo de gás deve ser produzido para maximizar os lucros. Se a quantidade de gás regular ou premium produzida por semana for denotada por x_1 e x_2, respectivamente, o lucro total por semana pode ser calculado como

$$\text{Lucro total} = 150x_1 + 175x_2$$

ou, escrito como uma função objetivo de programação linear,

$$\text{Maximize } Z = 150x_1 + 175x_2$$

As restrições podem ser desenvolvidas de forma similar. Por exemplo, a quantidade total de gás bruto utilizada pode ser calculada como

$$\text{Total de gás utilizado} = 7x_1 + 11x_2$$

Esse total não pode exceder o estoque disponível de 77 m^3/semana, de modo que a restrição pode ser representada por

$$7x_1 + 11x_2 \leq 77$$

As restrições restantes podem ser desenvolvidas de forma parecida, com o resultado da formulação total do problema de PL dada por

$$\text{Maximize } Z = 150x_1 + 175x_2 \quad \text{(maximize o lucro)}$$

sujeito a

$$7x_1 + 11x_2 \leq 77 \quad \text{(restrição de material)}$$
$$10x_1 + 8x_2 \leq 80 \quad \text{(restrição de tempo)}$$
$$x_1 \leq 9 \quad \text{(restrição de armazenamento do tipo “regular”)}$$
$$x_2 \leq 6 \quad \text{(restrição de armazenamento do tipo “premium”)}$$
$$x_1, x_2 \geq 0 \quad \text{(restrições de sinal positivo)}$$

Observe que essas equações constituem a formulação total da PL. As explicações entre parênteses do lado direito foram acrescentadas para esclarecer o significado de cada termo.

15.1.2 Solução Gráfica

Como estão limitadas a duas ou três dimensões, as soluções gráficas têm utilidade prática limitada. Entretanto, são muito usadas para ilustrar alguns conceitos básicos subjacentes às técnicas algébricas usadas para resolver problemas de grande dimensão com o auxílio do computador.

Para um problema bidimensional, como o Exemplo 15.1, o espaço da solução é definido como um plano com x_1 medido ao longo da abscissa e x_2 ao longo das ordenadas. Já que são lineares, as restrições podem ser representadas nesse plano como retas. Se o problema de PL foi formulado adequadamente (isto é, tem solução), essas retas de restrição demarcam uma região, chamada *espaço das soluções viáveis*, compreendendo todas as possíveis combinações de x_1 e x_2 que obedecem às restrições e, portanto, representam soluções viáveis. A função objetivo para um valor particular de Z pode então ser desenhada como outra reta e superposta nesse espaço. O valor de Z pode ser ajustado até que ele esteja em um valor máximo e ainda toque o espaço viável. Esse valor de Z representa a solução ótima. Os valores correspondentes de x_1 e x_2, onde Z toca o espaço das soluções viáveis, representam os valores ótimos para as atividades. O exemplo a seguir deve ajudar a esclarecer essa abordagem.

EXEMPLO 15.2 Solução Gráfica

Enunciado do Problema. Desenvolva a solução gráfica para o problema de processamento de gás descrito previamente no Exemplo 15.1:

Maximize $Z = 150x_1 + 175x_2$

sujeito a

$$7x_1 + 11x_2 \leq 77 \quad (1)$$
$$10x_1 + 8x_2 \leq 80 \quad (2)$$
$$x_1 \leq 9 \quad (3)$$
$$x_2 \leq 6 \quad (4)$$
$$x_1 \geq 0 \quad (5)$$
$$x_2 \geq 0 \quad (6)$$

As restrições foram numeradas para que sejam identificadas na solução gráfica a seguir.

Solução. Primeiro, as restrições devem ser traçadas no espaço de soluções. Por exemplo, a primeira restrição pode ser reformulada como uma reta substituindo-se a desigualdade por um sinal de igualdade e isolando x_2:

$$x_2 = -\frac{7}{11}x_1 + 7$$

Assim, na Figura 15.1*a*, os possíveis valores para x_1 e x_2 que obedecem a essa restrição caem abaixo dessa reta (a direção designada no desenho por uma pequena flecha). As outras restrições podem ser tratadas de forma análoga e superpostas na Figura 15.1*a*. Observe como elas englobam a região onde todas são satisfeitas. Esse é o espaço das soluções viáveis (a área *ABCDE* na figura).

Além de definir o espaço viável, a Figura 15.1*a* também fornece uma visão intuitiva da solução. Em particular, pode-se ver que a restrição 3 (armazenamento do gás regular) é "redundante". Isto é, o espaço das soluções viáveis não é afetado se essa restrição for removida.

Em seguida, a função objetivo pode ser acrescentada ao desenho. Para fazer isso, um valor de Z deve ser escolhido. Por exemplo, para $Z = 0$, a função objetivo se torna

$$0 = 150x_1 + 175x_2$$

ou, isolando x_2, encontramos a reta

$$x_2 = -\frac{150}{175}x_1$$

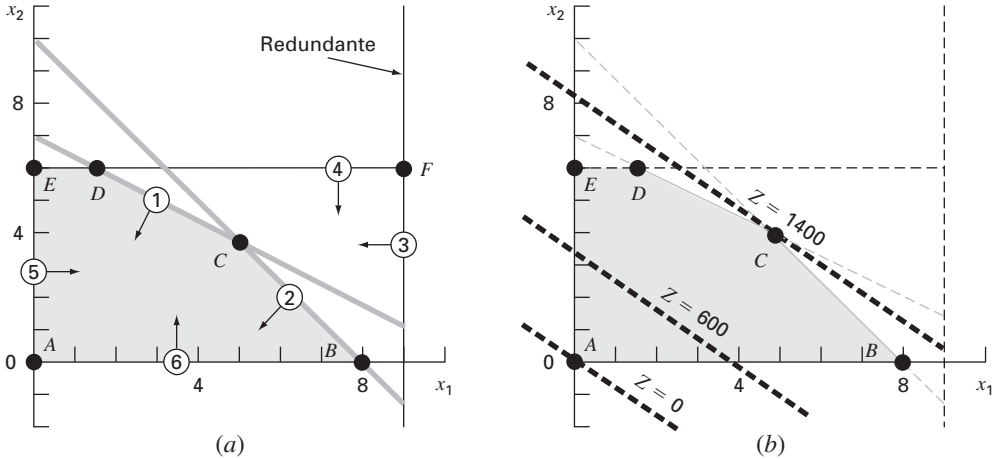


FIGURA 15.1
Solução gráfica de um problema de programação linear. (*a*) As restrições definem o espaço das soluções viáveis. (*b*) A função objetivo pode ser aumentada até atingir o valor máximo que obedeça a todas as restrições. Graficamente, a função se move para cima até tocar o espaço viável em um único ponto ótimo.

Como mostrado na Figura 15.1*b*, isso representa a reta tracejada que passa pela origem. Agora, já que o interesse é maximizar Z, pode-se aumentá-lo, por exemplo, para 600, e a função objetivo é

$$x_2 = \frac{600}{175} - \frac{150}{175}x_1$$

Assim, o aumento do valor da função objetivo move a reta para longe da origem. Uma vez que essa reta ainda cai no espaço de solução, o resultado ainda é viável. Por essa mesma razão, entretanto, ainda existe a possibilidade de melhora. Então, Z pode continuar crescendo até que mais aumento leve a função objetivo para além da região viável. Como mostrado na Figura 15.1*b*, o valor máximo de Z corresponde a aproximadamente 1.400. Nesse ponto, x_1 e x_2 são iguais a aproximadamente 4,9 e 3,9, respectivamente. Assim, a solução gráfica diz que, se forem produzidas essas quantidades de regular e premium, haverá um lucro máximo de cerca de 1.400.

Além de determinar os valores ótimos, a abordagem gráfica fornece mais informações intuitivas sobre o problema. Isso pode ser apreciado substituindo-se as respostas de volta nas equações das restrições,

$$7(4{,}9) + 11(3{,}9) \cong 77$$

$$10(4{,}9) + 8(3{,}9) \cong 80$$

$$4{,}9 \leq 9$$

$$3{,}9 \leq 6$$

Conseqüentemente, como fica claro também a partir da figura, produzir a quantidade ótima de cada produto leva direto a um ponto no qual as restrições de recurso (1) e tempo (2) são satisfeitas, mas no limite. Tais restrições são ditas *limitantes*. Ademais, como também é evidente graficamente, nenhuma das restrições de armazenamento [(3) e (4)] age como uma limitação. Tais restrições são chamadas *não-limitantes*. Isso leva à conclusão

prática de que, nesse caso, pode-se aumentar os lucros aumentando as fontes de recurso (o gás bruto) ou o tempo de produção. Além disso, tudo indica que um aumento no armazenamento não teria impacto no lucro.

O resultado obtido no exemplo anterior é um dos quatro possíveis que podem ser obtidos em um problema de programação linear. São eles:

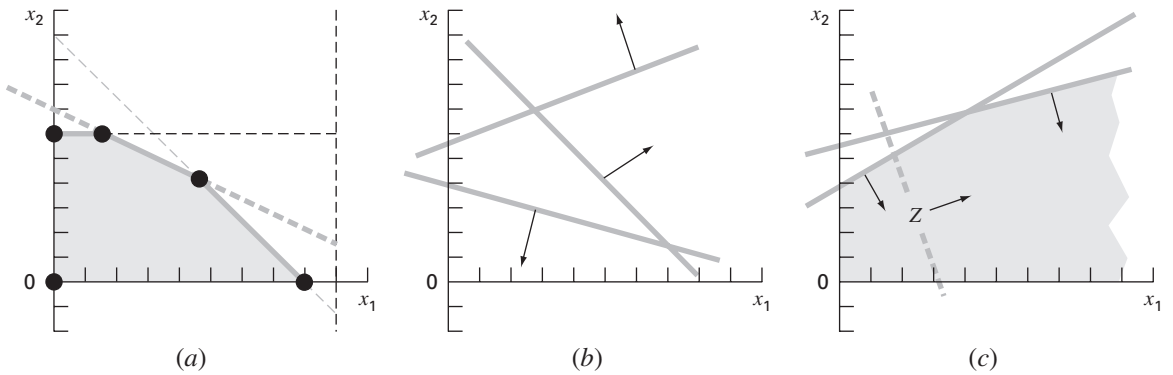
1. *Solução única*. Como no exemplo, o máximo da função objetivo intercepta a região viável em um único ponto.
2. *Soluções alternativas*. Suponha que a função objetivo do exemplo tenha coeficientes de modo que seja precisamente paralela a uma das restrições. No problema, uma maneira de isso ocorrer seria se o lucro fosse modificado para $ 140/ton e $ 220/ton. Então, em vez de um único ponto, o problema teria um número infinito de pontos ótimos, correspondentes a um segmento de reta (Figura 15.2*a*).
3. *Nenhuma solução viável*. Como na Figura 15.2*b*, é possível que o problema seja enunciado de modo que não existam soluções viáveis. Isso pode ocorrer por causa do fato de se estar lidando com um problema insolúvel ou de erros na descrição do problema. O último caso pode ocorrer se o problema está sobre-restrito, ao ponto de nenhuma solução poder satisfazer todas as restrições.
4. *Problemas ilimitados*. Como na Figura 15.2*c*, isso usualmente significa que o problema é sub-restrito e, portanto, está aberto. Como no item anterior, nenhuma solução viável, esse caso pode resultar de erros cometidos durante a especificação do problema.

Suponha agora que o problema envolva uma solução única. A abordagem gráfica pode sugerir uma estratégia enumerativa para procurar o máximo. Olhando a Figura 15.1, fica claro que o ótimo sempre ocorre em um dos vértices, onde duas restrições se encontram. Tal ponto é conhecido formalmente como *ponto extremo*. Assim, de um número infinito de possibilidades no espaço de decisão, concentrar-se nos pontos extremos claramente reduz as opções possíveis.

Além disso, pode-se reconhecer que nem todo ponto extremo é viável, isto é, satisfaz todas as restrições. Por exemplo, observe que o ponto *F* na Figura 15.1*a* é um ponto extremo, mas não é viável. Limitaremo-nos aos *pontos extremos viáveis* para reduzir ainda mais o espaço de busca.

Finalmente, uma vez que os pontos extremos sejam identificados, aquele que fornece o melhor valor para a função objetivo representa a solução ótima. A busca por essa solução ótima pode ser feita exaustivamente (e de forma ineficiente) calculando-se o

FIGURA 15.2
Além de uma solução única para o ponto ótimo (por exemplo, a Figura 15.1*b*), há três outros possíveis resultados para um problema de programação linear: (*a*) pontos ótimos alternativos, (*b*) nenhuma solução viável, e (*c*) um resultado ilimitado.



valor da função objetivo em cada ponto extremo viável. A seção seguinte discute o método simplex, que oferece uma estratégia mais atraente, ao mapear um caminho seletivo por uma seqüência de pontos extremos viáveis para chegar ao ponto ótimo de uma maneira extremamente eficiente.

15.1.3 O Método Simplex

O método simplex é baseado na suposição de que uma solução ótima é um ponto extremo. Assim, a abordagem deve ser capaz de discernir se, durante a solução do problema, um ponto extremo ocorre. Para fazer isso, as equações de restrição são reformuladas como igualdades introduzindo-se as chamadas variáveis de folga.

Variáveis de Folga. Como o nome indica, uma *variável de folga* (ou *de desvio*) mede quanto de recurso restrito está disponível, isto é, quanta "folga" de recurso está disponível. Por exemplo, lembre-se da restrição de recurso usada nos Exemplos 15.1 e 15.2,

$$7x_1 + 11x_2 \leq 77$$

Pode-se definir a variável de folga S_1 como a quantidade de gás bruto que não é usada para um nível de produção particular (x_1, x_2). Se essa quantidade for adicionada do lado esquerdo da restrição, torna a relação exata,

$$7x_1 + 11x_2 + S_1 = 77$$

Agora, perceba o que a variável de folga diz. Se for positiva, significa que se tem alguma "folga" na restrição. Ou seja, há alguma sobra de recurso que não está sendo completamente utilizada. Se for negativa, isso diz que a restrição foi excedida. Finalmente, se for zero, satisfez-se exatamente a restrição — todos os recursos disponíveis foram usados. Como esta é exatamente a condição na qual as retas de restrição se interceptam, a variável de folga fornece um meio para se detectar esses pontos extremos.

Uma variável de folga diferente é usada para cada equação de restrição, resultando no que é chamado de *versão completamente aumentada*,

$$\text{Maximize } Z = 150x_1 + 175x_2$$

sujeito a

$$7x_1 + 11x_2 + S_1 = 77 \tag{15.4a}$$

$$10x_1 + 8x_2 + S_2 = 80 \tag{15.4b}$$

$$x_1 + S_3 = 9 \tag{15.4c}$$

$$x_2 + S_4 = 6 \tag{15.4d}$$

$$x_1,\ x_2,\ S_1,\ S_2,\ S_3,\ S_4 \geq 0$$

Observe que escrevemos as quatro equações com igualdade de modo que as incógnitas estejam alinhadas em colunas. Fizemos isso para enfatizar que estamos agora lidando com um sistema de equações algébricas lineares (lembre-se da Parte Três). Na seção seguinte, será mostrado como essas equações podem ser usadas para determinar os pontos extremos algebricamente.

Soluções Algébricas. Em contraste com a Parte Três, em que havia n equações com n incógnitas, o sistema exemplo [Equações (15.4)] é *subespecificado ou subdeterminado*, isto é, tem mais incógnitas que equações. Em termos gerais, existem n *variáveis estruturais* (ou seja, as incógnitas originais), m *variáveis de folga* ou *de sobra* (uma por restrição) e $n + m$ variáveis no total (estruturais mais de folga). No problema de produção de gás, há 2 variáveis estruturais, 4 variáveis de folga e 6 variáveis no total. Assim, o problema envolve a resolução de 4 equações com 6 variáveis.

A diferença entre o número de incógnitas e o número de equações (igual a 2 no problema) está diretamente relacionada a como é possível distinguir um ponto extremo

viável. Especificamente, todo ponto viável tem 2 variáveis entre as 6 iguais a zero. Por exemplo, os cinco vértices da área *ABCDE* têm os seguintes valores iguais a zero:

Ponto Extremo	Variáveis Iguais a Zero
A	x_1, x_2
B	x_2, S_2
C	S_1, S_2
D	S_1, S_4
E	x_1, S_4

Essa observação leva à conclusão que os pontos extremos podem ser determinados de forma padrão fazendo duas das variáveis iguais a zero. No exemplo, isso reduz o problema à forma solúvel de 4 equações com 4 incógnitas. Por exemplo, para o ponto *E*, fazer $x_1 = S_4 = 0$ reduz o sistema à forma padrão

$$\begin{aligned} 11x_2 + S_1 &= 77 \\ 8x_2 + S_2 &= 80 \\ + S_3 &= 9 \\ x_2 &= 6 \end{aligned}$$

o qual pode ser resolvido por $x_2 = 6$, $S_1 = 11$, $S_2 = 32$, e $S_3 = 9$. Conjuntamente com $x_1 = S_4 = 0$, esses valores definem o ponto *E*.

Generalizando, uma solução básica para *m* equações lineares com *n* incógnitas é desenvolvida fazendo-se $n - m$ variáveis iguais a zero e resolvendo-se as *m* equações para as *m* incógnitas remanescentes. As variáveis iguais a zero são formalmente chamadas de *variáveis não-básicas*, enquanto que as *m* variáveis remanescentes são chamadas de *variáveis básicas*. Se todas as variáveis básicas forem não-negativas, o resultado é chamado *solução básica viável*. O ponto ótimo será uma dessas soluções.

Assim, uma abordagem direta para calcular o ponto ótimo seria calcular todas as soluções básicas, determinar quais delas são viáveis e, entre estas, qual tem o maior valor de *Z*. Há duas razões pelas quais tal abordagem não é interessante.

Primeiro, mesmo para problemas de tamanho moderado, a abordagem pode envolver a solução de um grande número de equações. Para *m* equações com *n* incógnitas, isso resulta na solução de

$$C_m^n = \frac{n!}{m!(n-m)!}$$

sistemas de equações simultâneas. Por exemplo, se existirem 10 equações ($m = 10$) com 16 incógnitas ($n = 16$), haverá 8.008 [$= 16!/(10!\ 6!)$] sistemas de equações 10×10 para resolver!

Segundo, uma quantidade significativa dessas soluções pode não ser viável. Por exemplo, no problema corrente, entre $C_6^4 = 15$ pontos extremos, apenas 5 são viáveis. Claramente, se for possível evitar resolver todos esses sistemas desnecessários, um algoritmo mais eficiente será desenvolvido. Tal abordagem é descrita a seguir.

Implementação do Método Simplex. O método simplex evita as ineficiências delineadas na seção anterior. Ele faz isso começando com uma solução básica viável. A seguir, move-se por uma seqüência de outras soluções básicas viáveis que melhoram sucessivamente o valor da função objetivo. Eventualmente, o valor ótimo é atingido e o método é encerrado.

Essa abordagem será ilustrada com o problema do processamento de gás dos Exemplos 15.1 e 15.2. O primeiro passo é começar em uma solução básica viável (um ponto extremo do espaço viável). Para casos como o nosso, um ponto de partida óbvio seria o ponto A; isto é, $x_1 = x_2 = 0$. O sistema original com 6 equações e 4 incógnitas se torna

$$S_1 = 77$$
$$S_2 = 80$$
$$S_3 = 9$$
$$S_4 = 6$$

Assim, os valores iniciais para as variáveis básicas são dados automaticamente como sendo iguais ao lado direito das equações de restrição.

Antes de prosseguir para o próximo passo, a informação inicial pode agora ser resumida em uma forma tabular conveniente chamada *tableau*. Como mostrado a seguir, um *tableau* fornece um resumo da informação principal que constitui um problema de programação linear.

Básicas	Z	x_1	x_2	S_1	S_2	S_3	S_4	Solução	Intersecção
Z	1	−150	−175	0	0	0	0	0	
S_1	0	7	11	1	0	0	0	77	11
S_2	0	10	8	0	1	0	0	80	8
S_3	0	1	0	0	0	1	0	9	9
S_4	0	0	1	0	0	0	1	6	∞

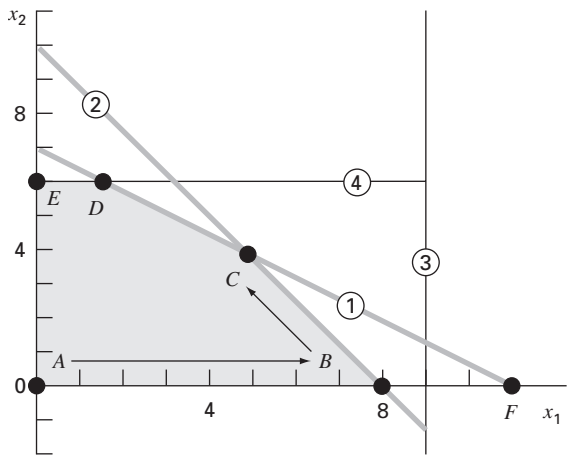
Observe que, no que diz respeito ao *tableau*, a função objetivo é expressa como

$$Z - 150x_1 - 175x_2 - 0S_1 - 0S_2 - 0S_3 - 0S_4 = 0 \tag{15.5}$$

O próximo passo envolve o movimento para uma nova solução básica viável que leve a uma melhoria na função objetivo. Consegue-se isso aumentando uma variável não-básica corrente (nesse ponto, x_1 ou x_2) acima do zero, de modo que Z aumente. Lembre-se de que, no exemplo presente, os pontos extremos devem ter 2 valores iguais a zero. Portanto, uma das variáveis básicas (S_1, S_2, S_3, ou S_4) deve ser igualada a zero.

Para resumir esse passo importante: uma das variáveis não-básicas correntes deve tornar-se básica (não-nula). Essa variável é chamada *variável de entrada*. Nesse processo, uma das variáveis básicas correntes deve tornar-se não-básica (nula). Essa variável é chamada *variável de saída*.

FIGURA 15.3
Representação gráfica de como o método simplex se move sucessivamente por soluções básicas viáveis até chegar a um ponto ótimo, de maneira eficiente.



Agora, será desenvolvida uma abordagem matemática para a escolha das variáveis de entrada e saída. Por causa da convenção que é usada para escever a função objetivo [Equação (15.5)], a variável de entrada pode ser qualquer variável na função objetivo que tenha um coeficiente negativo (porque isso faria Z crescer). A variável com maior valor negativo é convencionalmente escolhida porque usualmente leva a um maior incremento em Z. Nesse caso, x_2 seria a variável de entrada, já que seu coeficiente, –175, é menor que o coeficiente de x_1, –150.

Nesse ponto, a solução gráfica pode ser consultada para se conseguir intuição sobre os próximos passos. Como na Figura 15.3, começamos no ponto inicial A. Com base no seu coeficiente, x_2 deve ser escolhido como entrada. Entretanto, para que esse exemplo seja breve, escolheremos x_1, já que se pode ver do gráfico que isso levará mais rapidamente ao máximo.

Em seguida, deve-se escolher a variável de saída de um grupo de variáveis básicas correntes: S_1, S_2, S_3 ou S_4. Graficamente, vê-se que existem duas possibilidades. Mover-se para o ponto B fará com que S_2 seja zero, enquanto que mover-se para o ponto F fará S_1 igual a zero. Entretanto, o gráfico também mostra claramente que ir para F não é possível, pois este está fora do espaço das soluções viáveis. Portanto, a decisão é mover-se de A para B.

Como o mesmo resultado pode ser deduzido matematicamente? Uma maneira é calcular os valores nos quais as retas de restrição interceptam o eixo ou a reta correspondente à variável de entrada (no caso, o eixo x_1). Pode-se calcular esse valor como a razão entre o lado direito da restrição (a coluna "Solução" do *tableau*) e o coeficiente correspondente de x_1. Por exemplo, para a variável de folga da primeira restrição S_1, o resultado é

$$\text{Intersecção} = \frac{77}{7} = 11$$

As intersecções restantes podem ser calculadas e listadas como na última coluna do *tableau*. Uma vez que 8 é o menor inteiro positivo, isso significa que a segunda reta de restrição será atingida primeiro quando x_1 crescer. Assim, S_2 deve ser a variável de saída.

Nesse ponto, o movimento terá sido para B ($x_2 = S_2 = 0$), e a nova solução básica se torna

$$\begin{aligned} 7x_1 + S_1 &= 77 \\ 10x_1 &= 80 \\ x_1 + S_3 &= 9 \\ S_4 &= 6 \end{aligned}$$

A solução desse sistema de equações define efetivamente os valores das variáveis básicas no ponto B: $x_1 = 8$, $S_1 = 21$, $S_3 = 1$ e $S_4 = 6$.

O *tableau* pode ser usado para fazer os mesmos cálculos o método de Gauss-Jordan. Lembre-se de que a estratégia básica por trás do método de Gauss-Jordan envolve a conversão do elemento pivô para 1 e a eliminação dos coeficientes na mesma coluna acima e abaixo do elemento pivô (ver Seção 9.7).

Nesse exemplo, a linha do pivô é S_2 (a variável de saída) e o elemento pivô é 10 (o coeficiente da variável de entrada x_1). Dividindo a linha por 10 e substituindo S_2 por x_1, obtém-se

Básicas	Z	x_1	x_2	S_1	S_2	S_3	S_4	Solução	Intersecção
Z	1	−150	−175	0	0	0	0	0	
S_1	0	7	11	1	0	0	0	77	
x_1	0	1	0,8	0	0,1	0	0	8	
S_3	0	1	0	0	0	1	0	9	
S_4	0	0	1	0	0	0	1	6	

Em seguida, os coeficientes de x_1 nas outras linhas podem ser eliminados. Por exemplo, para a linha da função objetivo, a linha do pivô é multiplicada por -150 e o resultado é subtraído da primeira linha, resultando

Z	x_1	x_2	S_1	S_2	S_3	S_4	Solução
1	−150	−175	0	0	0	0	0
−0	−(−150)	−(−120)	−0	−(−15)	0	0	−(−1.200)
1	0	−55	0	15	0	0	1.200

Operações semelhantes podem ser feitas nas linhas restantes para dar o novo *tableau*,

Básicas	Z	x_1	x_2	S_1	S_2	S_3	S_4	Solução	Intersecção
Z	1	0	−55	0	15	0	0	1.200	
S_1	0	0	5,4	1	−0,7	0	0	21	3,889
x_1	0	1	0,8	0	0,1	0	0	8	10
S_3	0	0	−0,8	0	−0,1	1	0	1	−1,25
S_4	0	0	1	0	0	0	1	6	6

Assim, esse novo *tableau* sintetiza toda a informação para o ponto B. Isso inclui o fato de que o movimento aumentou a função objetivo para $Z = 1.200$.

O *tableau* pode então ser usado para mapear o próximo e, nesse caso, o último passo. Apenas mais uma variável, x_2, tem um valor negativo na função-objetivo e, portanto, é escolhida como variável de entrada. Segundo os valores da intersecção (agora calculados como a coluna solução sobre os coeficientes na coluna x_2), a primeira restrição tem o menor valor positivo e, portanto, S_1 é escolhida como a variável de saída. Assim, o método simplex leva do ponto B ao ponto C na Figura 15.3. Finalmente, a eliminação de Gauss-Jordan pode ser implementada para resolver as equações simultâneas. O resultado é o *tableau* final,

Básicas	Z	x_1	x_2	S_1	S_2	S_3	S_4	Solução
Z	1	0	0	10,1852	7,8704	0	0	1.413,889
x_2	0	0	1	0,1852	−0,1296	0	0	3,889
x_1	0	1	0	−0,1481	0,2037	0	0	4,889
S_3	0	0	0	0,1481	−0,2037	1	0	4,111
S_4	0	0	0	−0,1852	0,1296	0	1	2,111

Sabe-se que o resultado é final porque não existem coeficientes negativos remanescentes na linha da função objetivo. A solução final é tabulada como $x_1 = 3{,}889$ e $x_2 = 4{,}889$, o que fornece um máximo da função objetivo de $Z = 1.413{,}889$. Além disso, uma vez que S_3 e S_4 ainda estão entre as variáveis básicas, sabe-se que a solução é limitada pelas primeira e segunda restrições.

15.2 OTIMIZAÇÃO COM RESTRIÇÕES NÃO-LINEARES

Há várias abordagens para tratar de problemas de otimização na presença de restrições, e elas podem, geralmente, ser divididas nas abordagens indiretas e diretas (Rao, 1996). Uma abordagem indireta típica usa as chamadas *funções de penalização*, que envolvem a colocação de uma expressão adicional para fazer que a função objetivo se torne menos ótima quando a solução se aproxima de uma restrição. Assim, a solução será desencorajada de violar as restrições. Embora esses métodos sejam úteis para alguns problemas, eles podem tornar-se trabalhosos quando o problema envolve muitas restrições.

O método de busca do *gradiente reduzido generalizado* (GRG) é um dos mais populares entre os métodos diretos (para detalhes, ver Fylstra et al., 1998; Lasdon et al., 1978; Lasdon e Smith, 1992). Esse é, de fato, o método não-linear usado no Excel Solver.

Ele primeiro "reduz" o problema a um problema de otimização sem restrições, o que é feito resolvendo-se um conjunto de equações não-lineares para as variáveis básicas em termos das variáveis não-básicas. Assim, o problema sem restrições é resolvido usando-se abordagens parecidas com aquelas descritas no Capítulo 14. Inicialmente, uma direção de busca é escolhida ao longo da qual se tenha um incremento na função objetivo. A escolha-padrão é uma *abordagem quasi-Newton* (BFGS), que, como descrito no Capítulo 14, requer o armazenamento de uma aproximação para a matriz hessiana. Essa abordagem tem um bom desempenho na maior parte dos casos. A abordagem do *gradiente conjugado* também está disponível no Excel como uma alternativa para problemas de grande porte. O Excel Solver tem a boa característica de automaticamente mudar para o método do gradiente conjugado, dependendo do espaço disponível. Uma vez que a direção de busca é estabelecida, é realizada uma busca unidimensional ao longo dessa direção, com a utilização de uma abordagem de tamanho de passo variável.

15.3 OTIMIZAÇÃO COM PACOTES

Bibliotecas e pacotes de software têm grandes recursos para otimização. Nesta seção, será feita uma introdução a alguns dos mais utilizados.

15.3.1 EXCEL PARA PROGRAMAÇÃO LINEAR

Há uma grande variedade de pacotes de software expressamente projetados para implementar programação linear. Entretanto, por causa dessa grande disponibilidade, nos concentramos na planilha Excel. Isso envolve o uso da opção Solver previamente empregada no Capítulo 7 para a localização das raízes.

A forma como o Solver é usado na programação linear é parecida com as nossas aplicações prévias, pelo fato de os dados serem introduzidos em células da planilha. A estratégia básica é chegar a uma única célula que deve ser otimizada como uma função das variações de outras células da planilha. O exemplo seguinte ilustra como isso pode ser feito para o problema do processamento de gás.

EXEMPLO 15.3 Usando o Solver do Excel para um Problema de Programação Linear

Enunciado do Problema. Use o Excel para resolver o problema do processamento de gás examinado neste capítulo.

Solução. Uma planilha Excel preparada para calcular os valores pertinentes para o problema do processamento de gás é mostrada na Figura 15.4. As células não-sombreadas

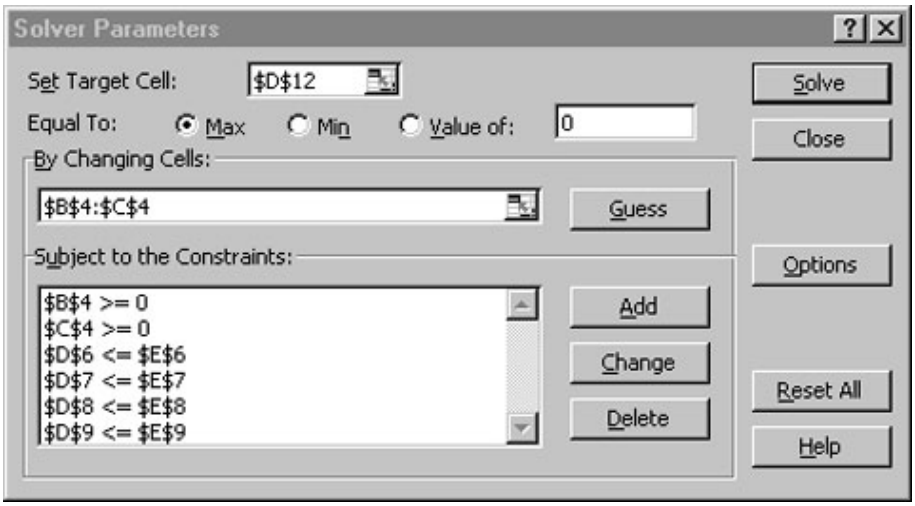
FIGURA 15.4
Planilha Excel preparada para o uso do Solver na programação linear.

	A	B	C	D	E
1	**Problema do Processamento de Gás**				
2					
3		Regular	Premium	Total	Disponível
4	Produzido	0	0		
5					
6	Gás bruto	7	11	**0**	77
7	Tempo	10	8	**0**	80
8	Armazenamento Regular			**0**	9
9	Armazenamento Premium			**0**	6
10					
11	Lucro por Unidade	150	175		
12	Lucro	**0**	**0**	**0**	

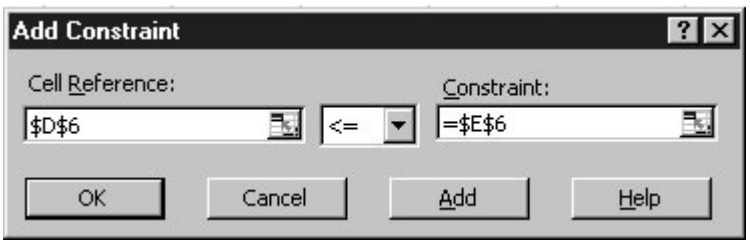
=B6*B4+C6*C4
=B7*B4+C7*C4
=B4
=C4
=B4*B11
=C4*C11
=B12+C12

são aquelas que contêm dados numéricos e de rotulação. As células sombreadas envolvem quantidades que são calculadas com base em outras células. Observe que a célula a ser maximizada é a D12, que representa o lucro total. As células a serem variadas são B4:C4, as quais armazenam as quantidades de gás regular e premium produzidas.

Uma vez criada a planilha, o **Solver** é selecionado no menu Tools. Nesse ponto, uma janela de diálogo será mostrada, perguntando pela informação pertinente. As células pertinentes da janela de diálogo do Solver são preenchidas como



As restrições são acrescentadas uma a uma selecionando-se o botão "Add". Essa ação abrirá uma janela de diálogo como a mostrada a seguir:



A restrição de que o gás bruto total (célula D6) deve ser inferior ou igual ao suprimento disponível (E6) pode ser inserida como já mostrado. Após acrescentar cada restrição, o botão "Add" pode ser selecionado. Quando todas as quatro restrições forem inseridas, o botão OK é selecionado para retornar à janela de diálogo do Solver.

Agora, antes de executar, o botão de opções do Solver deve ser selecionado (Options) e o campo indicado por "Assume linear model" deve ser marcado. Isso fará que o Excel empregue uma versão do algoritmo simplex (em vez do solver não-linear mais geral que é comumente usado), o que irá acelerar a sua aplicação.

Depois de selecionar essa opção, retorne para o menu do Solver. Quando o botão OK for selecionado, uma janela de diálogo se abre com a mensagem sobre o sucesso da operação. No caso presente, o Solver obtém a solução correta (Figura 15.5).

	A	B	C	D	E
1	**Problema de Processamento de Gás**				
2					
3		Regular	Premium	Total	Disponível
4	Produzido	**4.888889**	**3.888889**		
5					
6	Gás Bruto	7	11	77	77
7	Tempo	10	8	80	80
8	Armazenamento Regular			4.888889	9
9	Armazenamento Premium			3.888889	6
10					
11	Lucro por Unidade	150	175		
12	Lucro	733.3333	680.5556	**1413.889**	

FIGURA 15.5
Planilha Excel mostrando a solução do problema de programação linear.

Além de obter a solução, o Solver também fornece alguns relatórios resumidos úteis. Vamos explorá-los na aplicação em engenharia descrita na Seção 16.2.

15.3.2 Excel para Otimização Não-linear

A forma como o Solver é usado para a otimização não-linear é parecida com as aplicações anteriores, pelo fato de os dados serem introduzidos nas células da planilha. Mais uma vez, a estratégia básica é chegar a uma única célula que deve ser otimizada como função da variação de outras células da planilha. O exemplo seguinte ilustra como isso pode ser feito para o problema do pára-quedista discutido na introdução a esta parte do livro (lembre-se do Exemplo PT4.1).

EXEMPLO 15.4 Usando o Solver Excel para a Otimização Não-linear com Restrições

Enunciado do Problema. Lembre-se, do Exemplo PT4.1, que foi desenvolvida uma otimização não-linear com restrições para minimizar o custo de uma queda de pára-quedas em um campo de refugiados.

Parâmetro	Símbolo	Valor	Unidade
Massa total	M_t	2.000	kg
Aceleração da gravidade	g	9,8	m/s^2
Coeficiente de custo (constante)	c_0	200	$
Coeficiente de custo (comprimento)	c_1	56	$/m
Coeficiente de custo (área)	c_2	0,1	$/m^2
Velocidade de impacto crítica	v_c	20	m/s
Efeito área no arrasto	k_c	3	$kg/(s \cdot m^2)$
Altura inicial da queda	z_0	500	m

Substituindo esses valores nas Equações (PT4.11) a (PT4.19), tem-se

$$\text{Minimize } C = n(200 + 56\ell + 0{,}1A^2)$$

sujeito a

$$v \leq 20$$

$$n \geq 1$$

onde n é um inteiro e todas as outras variáveis são reais. Além disso, as seguintes quantidades são definidas por

$$A = 2\pi r^2$$
$$\ell = \sqrt{2}r$$
$$c = 3A$$
$$m = \frac{M_t}{n} \tag{15.6}$$
$$t = \text{raiz}\left[500 - \frac{9{,}8m}{c}t + \frac{9{,}8m^2}{c^2}\left(1 - e^{-(c/m)t}\right)\right] \tag{15.7}$$
$$v = \frac{9{,}8m}{c}\left(1 - e^{-(c/m)t}\right)$$

Use Excel para resolver esse problema, determinando as variáveis de projeto r e n que minimizem o custo C.

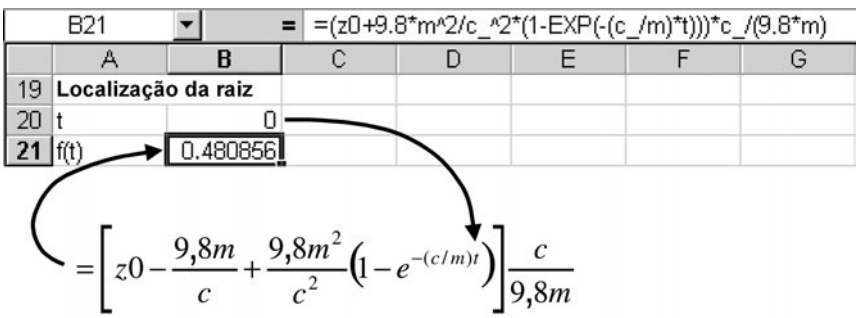
Solução. Antes da implementação desse problema em Excel, deve-se primeiro tratar do problema de determinar a raiz na formulação da Equação (15.7). Uma forma poderia ser o desenvolvimento de uma macro para implementar um método de localização da raiz, como a bissecção ou o método da secante. (Observe que ilustraremos como isso é feito no próximo capítulo, na Seção 16.3).

Por enquanto, é possível uma abordagem mais simples desenvolvendo a seguinte solução pela iteração de ponto fixo da Equação (15.7),

$$t_{i+1} = \left[500 + \frac{9{,}8m^2}{c^2}\left(1 - e^{-(c/m)t_i}\right)\right]\frac{c}{9{,}8m} \tag{15.8}$$

Assim, t pode ser ajustado até que a Equação (15.8) seja satisfeita. Pode ser mostrado que, para o intervalo dos parâmetros usado nesse problema, essa fórmula sempre converge.

Agora, como essa equação pode ser resolvida em uma planilha? Como mostrado a seguir, duas células podem ser preparadas para armazenar o valor de t e o lado direito da Equação (15.8) [isto é, $f(t)$].



Você pode digitar a Equação (15.8) na célula B21 de modo que esta obtenha o valor de tempo da célula B20 e os outros valores de parâmetro de células em outros locais da planilha (ver a seguir como preparar a planilha como um todo). Então, vá até a célula B20 e direcione seu valor para a célula B21.

Uma vez que você entre com todas essas formulações, imediatamente verá uma mensagem de erro: "Cannot resolve circular references", porque B20 depende de B21 e vice-versa. Agora, selecione Tools/Options no menu principal e selecione **calculation**. A partir da janela de diálogo do "calculation", marque "iteration" e pressione "OK". Imediatamente a planilha irá iterar essas células e o resultado virá como

	A	B
19	**Localização da raiz**	
20	t	10,2551
21	f(t)	10,25595

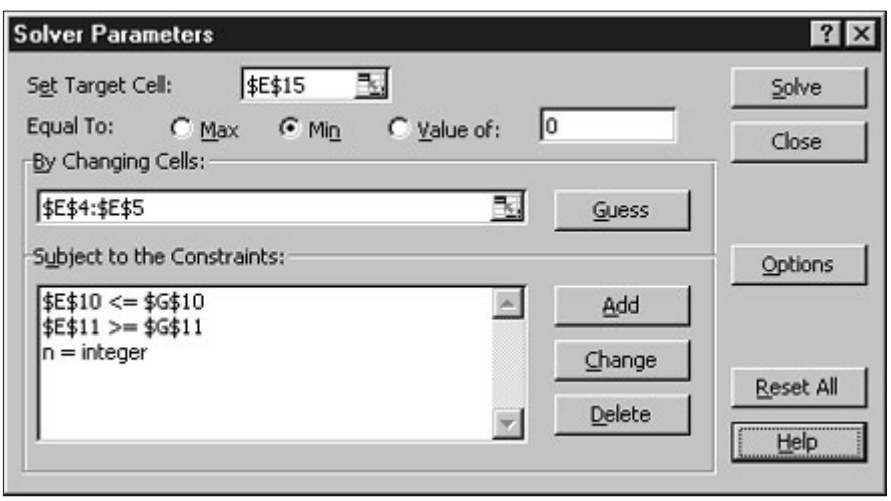
	A	B	C	D	E	F	G
1	**Problema de Otimização do Pára-quedas**						
2							
3	**Parâmetros:**			**Variáveis de projeto:**			
4	Mt	2000		r	1		
5	g	9.8		n	1		
6	cost1	200					
7	cost2	56		**Restrições:**			
8	cost3	0.1					
9	vc	20		**Variável**		**Tipo**	**Limite**
10	kc	3		v	**95.8786**	<=	20
11	z0	500		n	**1**	>=	1
12							
13	**Valores calculados:**			**Função objetivo:**			
14	A	**6.283185**					
15	l	**1.414214**		Custo	**283.1438**		
16	c	**18.84956**					
17	m	**2000**					
18							
19	**Localização da raiz:**						
20	t	**10.26439**					
21	f(t)	**10.26439**					

FIGURA 15.6
Planilha Excel preparada para otimização não-linear do problema do pára-quedas.

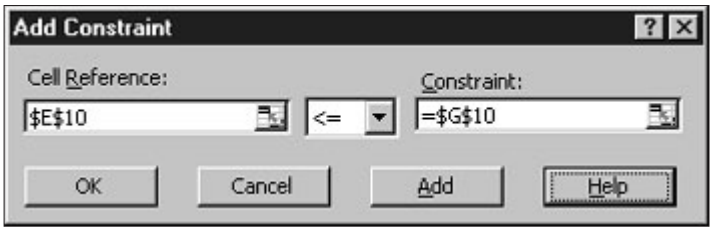
Assim, as células convergirão para a raiz. Se você quiser tornar o processo mais preciso, simplesmente pressione a tecla F9 para fazê-lo iterar mais algumas vezes (o padrão é de 100 iterações, o que pode ser modificado se você desejar).

Uma planilha Excel para calcular os valores pertinentes pode ser preparada como mostrado na Figura 15.6. As células não-sombreadas contêm valores numéricos e de rotulação. As células sombreadas envolvem quantidades que são calculadas com base em outras células. Por exemplo, a massa em B17 é calculada com a Equação (15.6), com base nos valores para M_t (B4) e n (E5). Observe também que algumas células são redundantes. Por exemplo, a célula E11 aponta para a célula E5. A informação é repetida na célula E11 de modo que a estrutura das restrições seja evidente na planilha. Finalmente, perceba que a célula a ser minimizada é a célula E15, a qual contém o custo total. As células a serem variadas são E4:E5, as quais armazenam o raio e o número de pára-quedas.

Uma vez que a planilha seja criada, escolha a opção Solver do menu Tools. Nesse ponto, uma janela de diálogo aparecerá, perguntando por informações pertinentes. As células pertinentes da caixa de diálogo do Solver devem ser preenchidas como



As restrições devem ser acrescentadas uma a uma selecionando-se o botão "Add". Isso abrirá uma caixa de diálogo como a mostrada a seguir



Como mostrado, a restrição de que a velocidade de impacto (célula E10) deva ser inferior ou igual à velocidade requerida (G10) pode ser inserida como mostrado na figura anterior. Depois de adicionar cada restrição, o botão "Add" deve ser selecionado. Observe que a seta para baixo lhe permite escolher entre diversos tipos de restrição (<=, >=, =, e inteiro). Portanto, pode-se forçar o número de pára-quedas (E5) a ser um inteiro.

Quando todas as três restrições tiverem sido introduzidas, o botão "OK" deve ser pressionado para retornar à janela de diálogo do Solver. Depois de selecionar essa opção, retorne para o menu do Solver. Quando o botão "OK" for selecionado, uma janela de diálogo aparecerá anunciando o sucesso da operação. Para o caso presente, o Solver obtém as soluções corretas, como na Figura 15.7.

	A	B	C	D	E	F	G
1	**Problema de Otimização do Pára-quedas**						
2							
3	**Parâmetros:**			**Variáveis de projeto:**			
4	Mt	2000		r	**2.943652**		
5	g	9.8		n	**6**		
6	cost1	200					
7	cost2	56		**Restrições:**			
8	cost3	0.1					
9	vc	20		**Variável**		**Tipo**	**Limite**
10	kc	3		v	**20**	<=	20
11	z0	500		n	**6**	>=	1
12							
13	**Valores calculados:**			**Função objetivo:**			
14	A	54.44435					
15	I	4.162953		Custo	**4377.264**		
16	c	163.333					
17	m	333.3333					
18							
19	**Localização da raiz:**						
20	t	27.04077					
21	f(t)	27.04077					

FIGURA 15.7
Planilha Excel mostrando a solução para a otimização não-linear do problema do pára-quedas.

Assim, foi determinado que o custo mínimo de $4.377,26 ocorrerá se a carga for dividida em seis partes com um pára-quedas de raio de 2,944 m. Além de obter a solução, o Solver também emitirá alguns relatórios resumidos, os quais serão explorados na aplicação de engenharia descrita na Seção 16.2.

15.3.3 MATLAB

Como resumido na Tabela 15.1, o software MATLAB tem uma variedade de funções internas para implementar otimização. Os dois exemplos seguintes ilustram como elas são usadas.

TABELA 15.1 Funções do MATLAB para implementar otimização.

Função	Descrição
fminbnd	Minimiza uma função de uma variável com restrições limitantes
fminsearch	Minimiza uma função de muitas variáveis

EXEMPLO 15.5 Usando MATLAB para Otimização Unidimensional

Enunciado do Problema. Use a função *fminbnd* do MATLAB para achar o máximo de

$$f(x) = 2 \operatorname{sen} x - \frac{x^2}{2}$$

no intervalo $x_l = 0$ e $x_u = 4$. Lembre-se de que, no Capítulo 13, foram usados vários métodos para resolver esse problema e determinar que $x = 1{,}7757$ e $f(x) = 1{,}4276$.

Solução. Primeiro, deve-se criar um M-file para armazenar a função.

```
function f=fx(x)
f = -(2*sen(x)-x^2/10)
```

Introduzimos o oposto da função, pois estamos interessados na maximização. Então, invocaremos a função fminbnd com

```
>> x=fminbnd('fx',0,4)
```

O resultado é

```
f =
   -1.7757

x =
    1.4275
```

Observe que argumentos adicionais podem ser incluídos. Um adendo bastante útil seria a inclusão de opções de otimização como a tolerância a erros ou número máximo de iterações. Isso é feito com a função optimset, a qual foi usada previamente no Exemplo 7.6 e tem o formato geral,

```
optimset('param1',value1,'param2',value2,...)
```

onde $param_i$ é um parâmetro especificando o tipo da opção e $value_i$ é o valor atribuído a essa opção. Por exemplo, se você quiser estipular uma tolerância de 1×10^{-2},

```
optimset('TolX',1e-2)
```

Assim, a solução do problema presente até uma tolerância de 1×10^{-2} pode ser gerada com

```
>> fminbnd('fx',0,4,optimset('TolX',1e-2))
```

com o resultado

```
f =
   -1.7757

ans =
    1.4270
```

Um conjunto completo de parâmetros pode ser encontrado invocando-se Help do seguinte modo:

```
>> Help optimset
```

EXEMPLO 15.6 Usando MATLAB para Otimização Multidimensional

Enunciado do Problema. Use a função do MATLAB *fminsearch* para achar o máximo de

$$f(x, y) = 2xy + 2x - x^2 - 2y^2$$

Use as aproximações iniciais $x = -1$ e $y = 1$. Lembre-se de que, no Capítulo 14, foram usados vários métodos para resolver esse problema e determinar que $x = 2$ e $y = 1$ com $f(x, y) = -2$.

Solução. Primeiro, deve-se criar um M-file para armazenar a função:

```
function f=fxy(x)
f = -(2*x(1)*x(2)+2*x(1)-x(1)^2-2*x(2)^2)
```

Entramos com o oposto da função porque estamos interessados na maximização. Então, invocaremos fminsearch com

```
>> x=fminsearch('fxy',[-1,1])
```

O resultado é

```
f =
   -2.0000

x =
    1.9999    1.0000
```

Como no caso da fminbnd, argumentos podem ser incluídos de modo a especificar parâmetros adicionais do processo de otimização. Por exemplo, a função optimset pode ser usada para limitar o número máximo de iterações

```
x=fminsearch('fxy',[-1,1],optimset('MaxIter',2))
```

com o resultado

```
f =
    7.0025

Exiting: Maximum number of iterations has been exceeded
         - increase MaxIter option.

         Current function value: 7.000000

x =
   -1    1
```

Então, como foi estipulado um limite muito rigoroso para o número de iterações, a otimização terminará bem antes de o máximo ser atingido.

TABELA 15.2 Rotinas IMSL para otimização.

Categoria	Rotina	Capacidade
Minimização sem restrições		
Função de uma variável		
	UVMIF	Usa apenas valores da função
	UVMID	Usa valores da função e da primeira derivada
	UVMGS	Função não-lisa
Função de várias variáveis		
	UMINF	Usa gradiente com diferenças finitas
	UMING	Usa o gradiente analítico
	UMIDH	Usa a hessiana com diferenças finitas
	UMIAH	Usa a hessiana analítica
	UMCGF	Usa gradiente conjugado, com gradiente com diferenças finitas
	UMCGG	Usa gradiente conjugado, com gradiente analítico
	UMPOL	Função não-lisa
Mínimos quadrados não-lineares	UNLSF	Usa jacobiano com diferenças finitas
	UNLSJ	Usa jacobiano analítico
Minimização com limites simples		
	BCONF	Usa gradiente com diferenças finitas
	BCONG	Usa gradiente analítico
	BCODH	Usa hessiana com diferenças finitas
	BCOAH	Usa hessiana analítica
	BCPOL	Função não-lisa
	BCLSF	Mínimos quadrados não-lineares, com jacobiano com diferenças finitas
	BCLSJ	Mínimos quadrados não-lineares, com jacobiano analítico
Minimização com restrições lineares		
	DLPRS	Programação linear densa
	QPROG	Programação quadrática
	LCONF	Função objetivo geral, com gradiente com diferenças finitas
	LCONG	Função objetivo geral, com gradiente analítico
Minimização com restrições não-lineares		
	NCONF	Usa gradiente com diferenças finitas
	NCONG	Usa gradiente analítico
Rotinas de serviço		
	CDGRD	Gradiente com diferenças centradas
	FDGRD	Gradiente com diferenças progressivas
	FDHES	Hessiana com diferenças progressivas
	GDHES	Hessiana com diferenças progressivas e gradiente analítico
	FDJAC	Jacobiano com diferenças progressivas
	CHGRD	Gradiente com teste fornecido pelo usuário
	CHHES	Hessiana com teste fornecido pelo usuário
	CHJAC	Jacobiano com teste fornecido pelo usuário
	GGUES	Geração de pontos de partida

15.3.4 IMSL

O IMSL tem várias sub-rotinas em Fortran para otimização (Tabela 15.2). Na presente discussão, será focada a rotina UVMID. Essa rotina localiza o ponto mínimo de uma função lisa de uma única variável usando cálculos da função e da primeira derivada.

UVMID é implementada pelo seguinte comando CALL:

```
CALL UVMID (F, G, XGUESS, ERREL, GTOL, MAXFN, A, B, X, FX, GX)
```

onde

F = FUNCTION fornecida pelo usuário para calcular o valor da função a ser minimizada. A forma é F(X), onde X é o ponto no qual a função é calculada. (Entrada) X não deve ser modificado por F. F é o valor calculado da função no ponto X. (Saída)

G = FUNCTION fornecida pelo usuário para calcular a derivada da função, onde G é o valor calculado da derivada da função no ponto X. (Saída)

F e G devem ser declaradas EXTERNAL na chamada do programa.

XGUESS = Uma aproximação inicial do ponto de mínimo de F. (Entrada)

ERREL = Acurácia relativa pedida para o valor final de X. (Entrada)

GTOL = Tolerância na derivada usada para decidir se o ponto em questão é um mínimo. (Entrada)

MAXFN = Número máximo de cálculos de funções permitido. (Entrada)

A = Extremidade inferior de um intervalo no qual o máximo está localizado. (Entrada)

B = Extremidade superior de um intervalo no qual o máximo está localizado. (Entrada)

FX = Valor da função em X. (Saída)

GX = Valor da derivada em X. (Saída)

EXEMPLO 15.7 Usando IMSL para localizar um Único Ponto Ótimo

Enunciado do Problema. Use UVMID para determinar o máximo da função unidimensional tratada no Capítulo 13 (ver Exemplos 13.1 a 13.3)

$$f(x) = 2 \operatorname{sen} x - \frac{x^2}{10}$$

Solução. Um exemplo de um programa principal e função em Fortran 90, utilizando UVMIF, para resolver esse problema pode ser escrito como

```
PROGRAM Oned

USE mimsl

IMPLICIT NONE
INTEGER::maxfn=50
REAL::xguess=0.,errel=1.E-6,gtol=1.E-6,a=-2.,b=2.
REAL::x,f,g,fx,gx
EXTERNAL f,g
CALL UVMID(f,g,xguess,errrel,gtol,maxfn,a,b,x,fx,gx)
PRINT *,x,fx,gx
END PROGRAM

FUNCTION f(x)
IMPLICIT NONE
REAL::x,f
f=-(2.*SIN(X) - x**2/10.)
END FUNCTION
```

```
FUNCTION g(x)
IMPLICIT NONE
REAL::x,g
g=-(2.*COS(x) - 2.*x/10.)
END FUNCTION
```

Observe que, como a rotina foi preparada para minimização, entrou-se com o oposto da função. Um exemplo de execução é

```
1.427334    -1.775726    -4.739729E-04
```

PROBLEMAS

15.1 Uma companhia faz dois tipos de produto, A e B. Esses produtos são fabricados durante 40 horas de trabalho por semana e, então, despachados no final da semana. Eles vão requerer, respectivamente, 20 e 5 kg de matéria-prima para cada quilo de produto, e a companhia dispõe de 9.500 kg de matéria-prima por semana. Apenas um produto pode ser fabricado por vez, com o tempo de produção para cada um de 0,04 e 0,12 hora, respectivamente. A fábrica pode armazenar apenas 550 kg de produto total por semana. Finalmente, a companhia tem lucro de $ 45 e $ 20 em cada unidade de A e B, respectivamente. Cada unidade de produto é equivalente a um quilo.

(a) Escreva um problema de programação linear para maximizar o lucro.
(b) Resolva o problema de programação linear graficamente.
(c) Resolva o problema de programação linear pelo método simplex.
(d) Resolva o problema com um pacote de software.
(e) Avalie qual das seguintes opções aumentará mais os lucros: aumentar a quantidade de matéria-prima, a capacidade de armazenamento, ou aumentar o tempo de produção.

15.2 Suponha que, para o Exemplo 15.1, a indústria de processamento de gás decida produzir um terceiro tipo de produto com as seguintes características,

	Supreme
Gás bruto	15 m^3/ton
Tempo de produção	12 hr/ton
Armazenamento	5 tons
Lucro	$250/ton

Além disso, suponha que uma nova fonte de gás bruto tenha sido descoberta, de modo que a quantidade total disponível é duplicada para 154 m^3/semana.

(a) Escreva o problema de programação linear para maximizar o lucro.
(b) Resolva o problema de programação linear pelo método simplex.
(c) Resolva o problema com um pacote de software.
(d) Avalie qual das seguintes opções aumentará mais os lucros: aumentar a quantidade de matéria-prima, a capacidade de armazenamento, ou aumentar o tempo de produção.

15.3 Considere o problema de programação linear:

$$\text{Maximize } f(x, y) = 1{,}75x + 1{,}25y$$

sujeito a

$$1{,}2x + 2{,}25y \leq 14$$
$$x + 1{,}1y \leq 8$$
$$2{,}5x + y \leq 9$$
$$x \geq 0$$
$$y \geq 0$$

Obtenha a solução:
(a) Graficamente.
(b) Usando o método simplex.
(c) Usando um pacote ou biblioteca de software apropriado (por exemplo, Excel, MATLAB ou IMSL).

15.4 Considere o problema de programação linear:

$$\text{Maximize } f(x, y) = 6x + 8y$$

sujeito a

$$5x + 2y \leq 40$$
$$6x + 6y \leq 60$$
$$2x + 4y \leq 32$$
$$x \geq 0$$
$$y \geq 0$$

Obtenha a solução:
(a) Graficamente.
(b) Usando o método simplex.
(c) Usando um pacote ou biblioteca de software apropriado (por exemplo, Excel, MATLAB ou IMSL).

15.5 Use um pacote ou biblioteca de software (por exemplo, Excel, MATLAB ou IMSL) para resolver o seguinte problema de otimização não-linear com restrições:

$$\text{Maximize } f(x, y) = 1{,}2x + 2y - y^3$$

sujeito a

$$2x + y \leq 2$$
$$x \geq 0$$
$$y \geq 0$$

15.6 Use um pacote ou biblioteca de software (por exemplo, Excel, MATLAB ou IMSL) para resolver o seguinte problema de otimização não-linear com restrições:

$$\text{Minimize } f(x, y) = 15x + 15y$$

sujeito a

$$x^2 + y^2 \leq 1$$
$$x + 2y \leq 2{,}1$$
$$x \geq 0$$
$$y \geq 0$$

15.7 Considere o seguinte problema de otimização não-linear com restrições:

Minimize $f(x, y) = (x - 3)^2 + (y - 3)^2$

sujeito a

$$x + 2y = 4$$

(a) Use uma abordagem gráfica para fazer uma estimativa da solução.
(b) Use um pacote ou biblioteca de software (por exemplo, o Excel) para obter uma estimativa mais acurada.

15.8 Use um pacote ou biblioteca de software para determinar o máximo de

$$f(x, y) = 2{,}25xy + 1{,}75y - 1{,}5x^2 - 2y^2$$

15.9 Use um pacote ou biblioteca de software para determinar o máximo de

$$f(x, y) = 4x + 2y + x^2 - 2x^4 + 2xy - 3y^2$$

15.10 Dada a seguinte função,

$$f(x, y) = -8x + x^2 + 12y + 4y^2 - 2xy$$

use um pacote ou biblioteca de software para determinar o mínimo:
(a) Graficamente.
(b) Numericamente.
(c) Substitua o resultado de **(b)** de volta na função para determinar o mínimo $f(x, y)$.
(d) Determine a hessiana e seu determinante e substitua o resultado do item **(b)** de volta nesse último para verificar que um mínimo foi de fato encontrado.

15.11 Você foi contratado para desenhar um poço cônico com cobertura para armazenar 50 m^3 de dejetos líquidos. Assuma o custo da escavação como \$ 100/m^3, o custo do revestimento lateral como \$ 50/m^2, e o custo da cobertura como \$ 25/m^2. Determine as dimensões do poço que minimizam o custo **(a)** se a inclinação lateral for irrestrita e **(b)** se a inclinação lateral tiver de ser menor que 45º.

15.12 Uma companhia de automóveis tem duas versões do mesmo modelo de carro para venda, um modelo compacto com duas portas e um com quatro portas em tamanho completo.

(a) Resolva graficamente quantos carros de cada modelo devem ser produzidos para maximizar o lucro e qual é esse lucro.
(b) Resolva o mesmo problema com Excel.

	Duas Portas	**Quatro Portas**	**Disponibilidade**
Lucro	\$13.500/carro	\$15.000/carro	
Tempo de Produção	15 h/carro	20 h/carro	8.000 h/ano
Armazenamento	400 carros	350 carros	
Demanda do Consumidor	700/carro	500/carros	240.000 carros.

15.13 Og é o líder de uma tribo de homens das cavernas, a Águas Calmas, surpreendentemente avançada matematicamente, embora tecnologicamente da "época roda do moinho". Ele deve decidir o número de clavas e o número de machados de pedra que devem ser produzidos para a futura batalha contra a tribo vizinha, a Pôr-do-Sol Tranqüilo. A experiência ensinou a ele que cada clava causa em média 0,45 morte e 0,65 mutilação, enquanto cada machado causa 0,70 morte e 0,35 mutilação. A produção de clavas requer 5,1 lb de pedra e 2,1 homens-hora de trabalho, enquanto um machado requer 3,2 lb de pedra e 4,3 homens-hora de trabalho. A tribo de Og tem 240 lb de pedra disponíveis para a produção de armas e um total de 200 homens-hora de trabalho disponíveis antes da data esperada da batalha (que Og tem certeza de que terminará de vez com a guerra). Og avalia que uma morte vale duas mutilações na quantificação do estrago causado ao inimigo e deseja produzir um lote de armas que maximize esse estrago.
(a) Formule esse problema como um problema de programação linear. Certifique-se de definir suas variáveis de decisão.
(b) Represente esse problema graficamente, certificando-se de distinguir os vértices viáveis dos vértices não-viáveis.
(c) Resolva o problema graficamente.
(d) Resolva o problema usando o computador.

CAPÍTULO 16

Estudo de Casos: Otimização

O propósito deste capítulo é utilizar os procedimentos numéricos discutidos do Capítulo 13 ao Capítulo 15 para resolver problemas reais da engenharia envolvendo otimização, os quais são importantes porque engenheiros são freqüentemente solicitados a oferecer a "melhor" solução para um dado problema. Métodos numéricos e computadores são geralmente necessários para desenvolver soluções ótimas, pois, em muitos casos, as questões envolvem sistemas complexos e interações.

As aplicações a seguir são típicas das que se encontram durante os cursos de final de graduação e de pós-graduação. Além disso, são representativas dos problemas que você deverá enfrentar como profissional. Tais problemas são derivados das diferentes áreas da engenharia: química/bioengenharia, civil/ambiental, elétrica e mecânica/aeroespacial.

A primeira aplicação, retirada da *engenharia química/bioengenharia*, consiste em usar otimização com restrições não-lineares para projetar um tanque cilíndrico ótimo. O Excel Solver é utilizado para desenvolver a solução.

Em seguida, será usada programação linear para abordar um problema da *engenharia civil/ambiental*: minimizar o custo do tratamento de dejetos para atingir os objetivos de qualidade de água em um rio. Nesse exemplo, serão introduzidos a noção de preços-sombra e seu uso para avaliar a sensibilidade de uma solução de programação linear.

A terceira aplicação, tomada da *engenharia elétrica*, envolve a maximização da potência que atravessa um potenciômetro em um circuito elétrico. A solução envolve otimização unidimensional sem restrições. Além de resolver o problema, será ilustrado como a macrolinguagem do Visual Basic permite acessar o algoritmo de busca da região áurea, no contexto do ambiente Excel.

Finalmente, a quarta aplicação, tirada da *engenharia mecânica/aeroespacial*, envolve achar os deslocamentos da longarina em uma *mountain bike* minimizando uma equação bidimensional para a energia potencial.

16.1 PROJETO DE CUSTO MÍNIMO DE UM TANQUE (ENGENHARIA QUÍMICA/BIOENGENHARIA)

Fundamentos. Os engenheiros químicos (bem como outros especialistas como os engenheiros mecânicos ou civis) enfrentam freqüentemente o problema geral de desenvolver contêineres para transportar líquidos. Suponha que lhe seja pedido para determinar as dimensões de um pequeno tanque cilíndrico para transportar resíduos tóxicos que será montado na carroceria de uma camionete. Seu objetivo geral será minimizar o custo do tanque. Entretanto, além do custo, você deve assegurar que o tanque armazene a quantidade requerida de líquido e que não exceda as dimensões do leito da camionete. Observe que, pelo fato de o tanque carregar resíduos tóxicos, a sua espessura é especificada por normas.

Um esquema do tanque e do leito é mostrado na Figura 16.1. Como pode ser visto, o tanque consiste em um cilindro com duas placas soldadas em cada extremidade.

O custo de um tanque envolve duas componentes: (1) custo do material, o qual é baseado no peso, e (2) custo da soldagem, baseado no comprimento da solda. Observe que o último envolve tanto o cordão interior quanto o exterior onde as placas se conectam com o cilindro. Os dados necessários para o problema estão resumidos na Tabela 16.1.

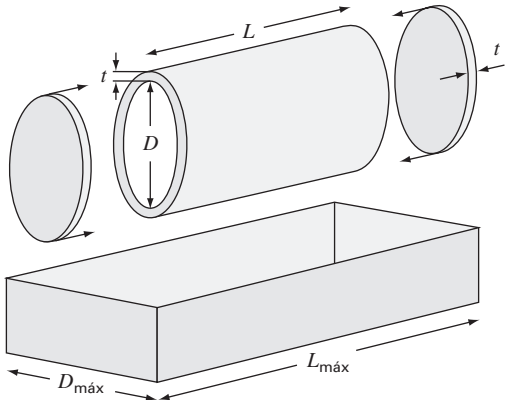


FIGURA 16.1
Parâmetros para determinar as dimensões ótimas de um tanque cilíndrico.

TABELA 16.1 Parâmetros para determinar as dimensões ótimas de um tanque cilíndrico usado para transportar resíduos tóxicos.

Parâmetro	Símbolo	Valor	Unidades
Volume requerido	V_o	0,8	m^3
Espessura	t	3	cm
Densidade	ρ	8.000	kg/m^3
Comprimento do leito	$L_{máx}$	2	m
Largura do leito	$D_{máx}$	1	m
Custo do material	c_m	4,5	\$/kg
Custo da soldagem	c_w	20	\$/m

Solução. O objetivo aqui é construir o tanque com um custo mínimo. O custo está relacionado com as variáveis de projeto (comprimento e diâmetro), já que elas afetam a massa do tanque e o comprimento dos cordões de solda. Ademais, o problema tem restrições porque o tanque deve (1) encaixar-se na carroceria da camionete e (2) carregar o volume requerido de material.

O custo consiste no material do tanque e nos custos de soldagem. Portanto, a função objetivo pode ser escrita como minimizando

$$C = c_m m + c_w \ell_w \tag{16.1}$$

onde C é o custo (\$), m é a massa (kg), ℓ_w é o comprimento da solda (m), e c_m e c_w são os fatores de custo para massa (\$/kg) e comprimento da solda (\$/m), respectivamente.

Em seguida, formula-se como a massa e o comprimento da solda estão relacionados às dimensões do tanque. Primeiro, a massa pode ser calculada como o volume de material vezes a densidade. O volume de material usado para criar as paredes laterais (isto é, o cilindro) pode ser calculado por

$$V_{\text{cilindro}} = L\pi\left[\left(\frac{D}{2} + t\right)^2 - \left(\frac{D}{2}\right)^2\right]$$

Para cada placa circular nas extremidades, tem-se

$$V_{\text{placa}} = \pi\left(\frac{D}{2} + t\right)^2 t$$

Portanto, a massa calculada é

$$m = \rho\left\{L\pi\left[\left(\frac{D}{2} + t\right)^2 - \left(\frac{D}{2}\right)^2\right] + 2\pi\left(\frac{D}{2} + t\right)^2 t\right\} \tag{16.2}$$

onde ρ é a densidade (kg/m^3).

O comprimento da solda para prender cada placa é igual à circunferência interna mais a externa do cilindro. Para as duas placas, o comprimento total da solda é

$$\ell_w = 2\left[2\pi\left(\frac{D}{2} + t\right) + 2\pi\frac{D}{2}\right] = 4\pi(D + t) \tag{16.3}$$

Dados os valores de D e L (lembre-se, a espessura t é fixada por normas), as Equações (16.1) a (16.3) fornecem um meio de calcular o custo. Perceba também que, quando as Equações (16.2) e (16.3) são substituídas na Equação (16.1), a função objetivo resultante é não-linear nas incógnitas.

Em seguida, pode-se formular as restrições. Primeiro, calcula-se qual o volume que pode ser armazenado dentro do tanque,

$$V = \frac{\pi D^2}{4}L$$

Esse volume deve ser igual ao volume desejado. Assim, uma restrição é

$$\frac{\pi D^2 L}{4} = V_o$$

onde V_o é o volume desejado (m^3).

As restrições restantes se destinam a garantir que o tanque caiba nas dimensões do leito da carroceria da camionete.

$$L \leq L_{máx}$$
$$D \leq D_{máx}$$

O problema está agora especificado. Substituindo os valores da Tabela 16.1, o problema pode ser resumido como

$$\text{Maximize } C = 4{,}5m + 20\ell_w$$

sujeito a

$$\frac{\pi D^2 L}{4} = 0{,}8$$
$$L \leq 2$$
$$D \leq 1$$

FIGURA 16.2
Planilha Excel preparada para calcular o custo de um tanque sujeito a uma exigência de volume e a restrições de tamanho.

	A	B	C	D	E	F	G
1	**Projeto ótimo de um tanque**						
2							
3	**Parâmetros**			**Variáveis de projeto**			
4	V0	0.8		D	1		
5	t	0.03		L	2		
6	rho	8000					
7	Lmax	2		**Restrições:**			
8	Dmax	1		D	**1**	<=	1
9	cm	4.5		L	**2**	<=	2
10	cw	20		Vol	**1.570796**	=	0.8
11							
12	**Valores calculados**			**Função objetivo**			
13	m	**1976.791**		C	**9154.425**		
14	lw	**12.94336**					
15							
16	Vshell	**0.19415**					
17	Vends	**0.052948**					

onde

$$m = 8000\left\{L\pi\left[\left(\frac{D}{2}+0{,}03\right)^2-\left(\frac{D}{2}\right)^2\right]+2\pi\left(\frac{D}{2}+0{,}03\right)^2 0{,}03\right\}$$

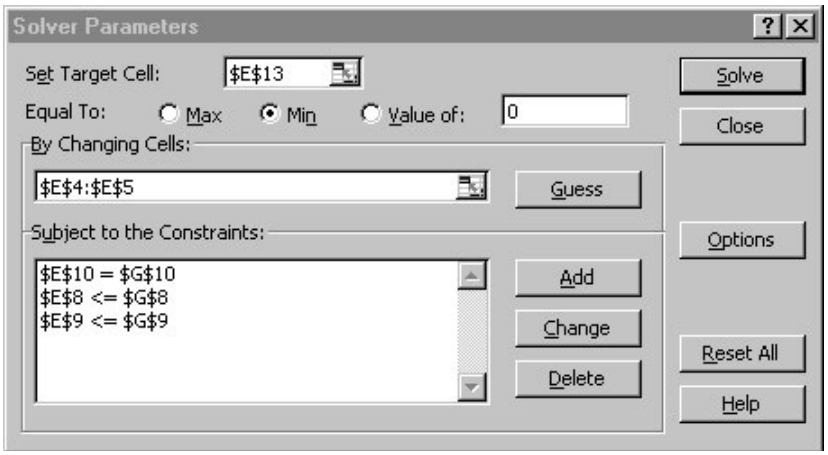
e

$$\ell_w = 4\pi(D+0{,}03)$$

O problema pode agora ser resolvido de várias maneiras. Entretanto, a abordagem mais simples para um problema dessa magnitude é usar uma ferramenta como o Excel Solver. A planilha para conseguir isso é mostrada na Figura 16.2.

Para o caso mostrado, entram-se os limites superiores para D e L. Para esse caso, o volume é maior que o requerido ($1{,}57 > 0{,}8$).

Uma vez que a planilha seja criada, a escolha Solver é feita a partir do menu de Ferramentas (Tools). Nesse ponto uma janela de diálogo aparecerá, perguntando pelas informações pertinentes. As células pertinentes da janela de diálogo do Solver devem ser preenchidas como



Quando o botão OK for selecionado, uma janela de diálogo se abrirá e mostrará um relatório de sucesso da operação. Para o caso presente, o Solver obtém a solução correta, a qual é mostrada na Figura 16.3. Observe que o diâmetro ótimo está raspando na restrição de 1 m. Portanto, se a capacidade requerida para o tanque fosse aumentada, nós atingiríamos essa restrição e nela permaneceríamos e o problema se reduziria a uma busca unidimensional pelo comprimento.

FIGURA 16.3
Resultado da minimização. O preço é reduzido de $ 9.154 para $ 5.723 por causa do menor volume usando as dimensões $D = 0{,}98$ m e $L = 1{,}05$ m.

	A	B	C	D	E	F	G
1	**Projeto ótimo de um tanque**						
2							
3	**Parâmetros**			**Variáveis de projeto:**			
4	V0	0.8		D	0.98351		
5	t	0.03		L	1.053033		
6	rho	8000					
7	Lmax	2		**Restrições:**			
8	Dmax	1		D	**0.98351**	<=	1
9	cm	4.5		L	**1.053033**	<=	2
10	cw	20		Vol	**0.799999**	=	0.8
11							
12	**Valores calculados:**			**Função objetivo:**			
13	m	**1215.206**		C	**5723.149**		
14	lw	**12.73614**					
15							
16	Vshell	**0.100587**					
17	Vends	**0.051314**					

16.2 TRATAMENTO DE CUSTO MÍNIMO PARA DEJETOS LÍQUIDOS (ENGENHARIA CIVIL/AMBIENTAL)

Fundamentos. As descargas de dejetos líquidos nas grandes cidades são freqüentemente a maior causa da poluição dos rios. A Figura 16.4 ilustra o tipo de sistema com o qual um engenheiro ambiental pode se confrontar. Várias cidades são localizadas perto dos rios e seus afluentes. Cada uma dessas cidades gera poluição em uma taxa P, medida em unidades de miligramas por dia (mg/d). A carga de poluição gerada é sujeita a um tratamento de dejetos do qual resulta uma fração de poluição removida x. Portanto, a quantidade de poluição descarregada nos rios é o excesso que não é removido pelo tratamento,

$$W_i = (1 - x_i) P_i \tag{16.4}$$

onde W_i é a descarga de dejetos da i-ésima cidade.

Quando a descarga de dejetos entra na corrente, mistura-se com a poluição das fontes a montante. Se for suposta uma mistura completa no ponto de descarga, a concentração resultante nesse ponto de descarga pode ser calculada por um simples balanço de massa,

$$c_i = \frac{W_i + Q_u c_u}{Q_i} \tag{16.5}$$

onde Q_u é o escoamento (L/d), c_u é a concentração (mg/L) no rio imediatamente a montante do ponto de descarga, e Q_i é o escoamento a jusante do ponto de descarga (L/d).

Após a concentração no ponto de mistura ter sido estabelecida, os processos de decomposição química e biológica podem remover alguma poluição enquanto esta escoa rio abaixo. No caso presente, será suposto que essa remoção pode ser representada por um simples fator de redução fracionária R.

Supondo que as nascentes (isto é, o rio acima das cidades 1 e 2) estejam livres da poluição, a poluição nos quatro nós pode ser calculada como

$$\begin{aligned} c_1 &= \frac{(1 - x_1) P_1}{Q_{13}} \\ c_2 &= \frac{(1 - x_2) P_2}{Q_{23}} \\ c_3 &= \frac{R_{13} Q_{13} c_1 + R_{23} Q_{23} c_2 + (1 - x_3) P_3}{Q_{34}} \\ c_4 &= \frac{R_{34} Q_{34} c_3 + (1 - x_4) P_4}{Q_{45}} \end{aligned} \tag{16.6}$$

A seguir, percebe-se que o tratamento de dejetos custa uma quantidade diferente, d_i ($ 1.000/mg de remoção), em cada uma das estações. Portanto, o custo total do tratamento (em uma base diária) pode ser calculado por

$$Z = d_1 P_1 x_1 + d_2 P_2 x_2 + d_3 P_3 x_3 + d_4 P_4 x_4 \tag{16.7}$$

onde Z é o custo total diário do tratamento ($ 1.000/d).

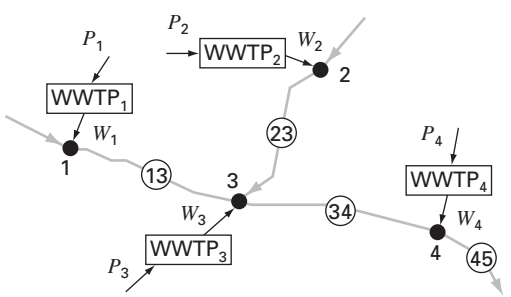


FIGURA 16.4
Quatro estações de tratamento de dejetos descarregando poluição em um sistema fluvial. Os segmentos do rio entre as cidades são indicados por números circulados.

TABELA 16.2 Parâmetros para as quatro estações de tratamento de água descarregando poluição em um sistema fluvial, com a concentração resultante (c_i) para tratamento zero. O escoamento, a remoção e os padrões para os segmentos do rio também são listados.

Cidade	P_i (mg/d)	d_i ($10$$^{-6}$/mg)	c_i (mg/L)	Segmento	Q (L/d)	R	c_s (mg/L)
1	$1{,}00 \times 10^9$	2	100	1–3	$1{,}00 \times 10^7$	0,5	20
2	$2{,}00 \times 10^9$	2	40	2–3	$5{,}00 \times 10^7$	0,35	20
3	$4{,}00 \times 10^9$	4	47,3	3–4	$1{,}10 \times 10^8$	0,6	20
4	$2{,}50 \times 10^9$	4	22,5	4–5	$2{,}50 \times 10^8$		20

A peça final no "jogo da decisão" envolve regulamentos e normas ambientais. Para proteger o uso benéfico do rio (por exemplo, navegação, pesca, banho), os regulamentos dizem que a concentração de poluentes no rio não deve exceder um padrão de qualidade de água c_s.

Os parâmetros para o sistema fluvial na Figura 16.4 estão resumidos na Tabela 16.2. Observe que existe uma diferença no custo do tratamento entre as cidades rio acima (1 e 2) e as cidades rio abaixo (3 e 4) por causa do modelo antiquado das instalações nessa parte do rio.

A concentração pode ser calculada com a Equação (16.6) e o resultado foi listado na coluna sombreada, para o caso em que nenhum tratamento de dejetos é implementado (isto é, todos os x's = 0). Observe que o padrão de 20 mg/L é violado em todos os pontos de mistura.

Use programação linear para determinar os níveis de tratamento que satisfazem o padrão de qualidade de água com custo mínimo. Além disso, avalie o impacto de tornar o padrão mais rigoroso abaixo da cidade 3. Isto é, refaça o exercício, mas com padrões para os segmentos 3-4 e 4-5 abaixados para 10 mg/L.

Solução. Todos esses fatores descritos podem ser combinados no seguinte problema de programação linear:

$$\text{Minimize } Z = d_1 P_1 x_1 + d_2 P_2 x_2 + d_3 P_3 x_3 + d_4 P_4 x_4 \tag{16.8}$$

sujeito às seguintes restrições

$$\frac{(1 - x_1) P_1}{Q_{13}} \le c_{s1}$$
$$\frac{(1 - x_2) P_2}{Q_{23}} \le c_{s2} \tag{16.9}$$

$$\frac{R_{13} Q_{13} c_1 + R_{23} Q_{23} c_2 + (1 - x_3) P_3}{Q_{34}} \le c_{s3}$$
$$\frac{R_{34} Q_{34} c_3 + (1 - x_4) P_4}{Q_{45}} \le c_{s4}$$
$$0 \le x_1, x_2, x_3, x_4 \le 1 \tag{16.10}$$

Assim, a função objetivo é minimizar o custo do tratamento [Equação (16.8)] sujeito à restrição de que o padrão de qualidade de água deve ser satisfeito em todas as partes do sistema [Equação (16.9)]. Além disso, o tratamento não pode ser negativo ou superior a 100% de remoção [Equação (16.10)].

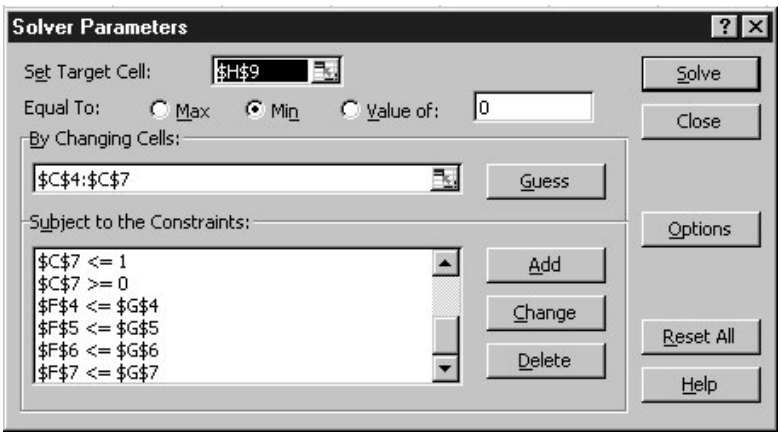
O problema pode ser resolvido usando-se vários pacotes. Para a presente aplicação, será utilizada uma planilha Excel. Como visto na Figura 16.5, os dados, assim como as concentrações calculadas, podem ser naturalmente colocados em uma planilha.

FIGURA 16.5
Planilha Excel preparada para calcular o custo do tratamento de dejetos em um sistema fluvial regulamentado. A coluna F contém o cálculo da concentração segundo a Equação (16.6). As células F4 e H4 são destacadas para mostrar as fórmulas usadas para calcular c_1 e o custo do tratamento para a cidade 1. Adicionalmente, a célula H9 é destacada para mostrar a fórmula (Equação 16.8) para calcular o custo total, que deve ser minimizado.

	A	B	C	D	E	F	G	H
1	**Tratamento de Dejetos Líquidos de Baixo Custo**							
2		**Não tratado**	**Tratamento**	**Decarga**	**Custo unt.**	**Conc.**	WQ	**Custo do**
3	**Cidade**	P	x	W	d	**rio**		**tratamento**
4	1	1.00E+09	0	1.00E+09	2.00E-06	**100.00**	20.00	**0.00**
5	2	2.00E+09	0	2.00E+09	2.00E-06	40.00	20.00	0.00
6	3	4.00E+09	0	4.00E+09	4.00E-06	47.27	20.00	0.00
7	4	2.50E+09	0	2.50E+09	4.00E-06	22.48	20.00	0.00
8		**Escoamento**	**Resíduo**					
9	**Segmento**	**do rio**	**removido**				Total	**0.00**
10	1-3	1.00E+07	0.5					
11	2-3	5.00E+07	0.35					
12	3-4	1.10E+08	0.6					
13	4-5	2.50E+08						

=D4/B10 **=$B$4*$C$4*$E$4** **=SUM(H4:H7)**

Uma vez que a planilha seja criada, a escolha do Solver é feita no menu de Ferramentas. Nesse momento, uma janela de diálogo aparecerá perguntando por informação pertinente ao problema. As células pertinentes da janela de diálogo do Solver devem ser preenchidas por



Observe que nem todas as restrições são mostradas, pois a janela de diálogo mostra apenas seis restrições por vez.

FIGURA 16.6
Resultados da minimização. O padrão de qualidade da água é atingido a um custo de $ 12.600 por dia. Observe que, a despeito do fato de que nenhum tratamento é necessário na cidade 4, a concentração no seu ponto de mistura realmente satisfaz o padrão.

	A	B	C	D	E	F	G	H
1	**Tratamento de Dejetos Líquidos de Baixo Custo**							
2		**Não tratado**	**Tratamento**	**Decarga**	**Custo unt.**	**Conc.**	WQ	**Custo do**
3	**Cidade**	P	x	W	d	**rio**		**tratamento**
4	1	1.00E+09	**0.80**	2.00E+08	2.00E-06	**20.00**	20.00	1600.00
5	2	2.00E+09	**0.50**	1.00E+09	2.00E-06	**20.00**	20.00	2000.00
6	3	4.00E+09	**0.5625**	1.75E+09	4.00E-06	**20.00**	20.00	9000.00
7	4	2.50E+09	**0.00**	2.50E+09	4.00E-06	**15.28**	20.00	0.00
8		**Escoamento**	**Resíduo**					
9	**Segmento**	**do rio**	**removido**				Total	**12600.00**
10	1-3	1.00E+07	0.5					
11	2-3	5.00E+07	0.35					
12	3-4	1.10E+08	0.6					
13	4-5	2.50E+08						

Quando o botão OK for selecionado, uma janela de diálogo aparecerá com o relatório de sucesso da operação. Para o caso presente, o Solver obtém a solução correta, que é mostrada na Figura 16.6. Antes de aceitar a solução (selecionando o botão OK da caixa de Relatório do Solver), observe que três relatórios podem ser gerados: Resposta, Sensibilidade e Limites. Selecione o Relatório de Sensibilidade e então clique no botão OK para aceitar a solução. O Solver automaticamente gerará o Relatório de Sensibilidade, que aparece como na Figura 16.7.

Agora vamos examinar a solução (Figura 16.6). Observe que o padrão será satisfeito em todos os pontos de mistura. De fato, a concentração na cidade 4 será menor que o padrão (16,28 mg/L), embora nenhum tratamento tenha sido necessário na cidade 4.

Como exercício final, pode-se baixar o padrão nos trechos 3–4 e 4–5 para 10 mg/L. Antes de fazer isso, o Relatório de Sensibilidade é examinado. Para o caso presente, a coluna-chave da Figura 16.7 é o multiplicador de Lagrange (também conhecido como preço-sombra). O *preço-sombra* é um valor que expressa a sensibilidade da função objetivo (no caso, o custo) a uma variação unitária de uma das restrições (o padrão de qualidade da água). Portanto, ele representa o custo adicional que incorrerá caso se torne o padrão mais exigente. No exemplo, é bastante revelador que o maior preço-sombra, $-\$\ 440/\Delta c_{s3}$, ocorre para uma mudança no padrão que se está considerando (isto é, a jusante da cidade 3), indicando que a mudança será dispendiosa.

Isso pode ser confirmado executando-se novamente o Solver com o novo padrão (ou seja, reduzindo o valor nas células G6 e G7 a G10). Como visto na Tabela 16.3, o

FIGURA 16.7
Relatório de Sensibilidade para planilha preparada para calocular o custo do tratamento de dejetos líquidos em um sistema fluvial.

Microsoft Excel 11.0 Sensitivity Report
Worksheet: [Sec1602.xls]Sheet1
Report Created: 2/28/2005 11:37:39 AM

Adjustable Cells

Cell	Name	Final Value	Reduced Gradient
$C$4	x	0.80	0.00
$C$5	x	0.50	0.00
$C$6	x	0.5625	0.0000
$C$7	x	0.00	10000.00

Constraints

Cell	Name	Final Value	Lagrange Multiplier
$F$4	conc	20.00	0.00
$F$5	conc	20.00	-30.00
$F$6	conc	20.00	-440.00
$F$7	conc	15.28	0.00

TABELA 16.3 Comparação de dois cenários envolvendo o impacto de diferentes valores de padrão no custo do tratamento.

Cenário 1: Todo $c_s = 20$			**Cenário 2: A Jusante $c_s = 10$**		
Cidade	***x***	***c***	**Cidade**	***x***	***c***
1	0,8	20	1	0,8	20
2	0,5	20	2	0,5	20
3	0,5625	20	3	0,8375	10
4	0	15,28	4	0,264	10
Custo = $12.600			Custo = $19.640		

resultado é que o custo do tratamento sobe de $ 12.600 por dia para $ 19.640 por dia. Além disso, reduzir o padrão das concentrações para atingir valores menores significa que a cidade 4 deve começar a tratar os seus dejetos e que a cidade 3 deve melhorar o seu tratamento. Observe também que o tratamento das cidades a montante não é afetado.

16.3 POTÊNCIA MÁXIMA TRANSFERIDA PARA UM CIRCUITO (ENGENHARIA ELÉTRICA)

Fundamentos. O circuito resistor simples da Figura 16.8 contém três resistores fixos e um resistor ajustável. Resistores ajustáveis são chamados *potenciômetros*. Os valores para os parâmetros são $V = 80$ V, $R_1 = 8\ \Omega$, $R_2 = 12\ \Omega$ e $R_3 = 10\ \Omega$. (a) Encontre o valor da resistência ajustável R_a que maximiza a potência transferida nos terminais 1 e 2. (b) Faça uma análise de sensibilidade para determinar como a potência máxima e o correspondente ajuste do potenciômetro R_a variam conforme V varia no intervalo de 45 a 105V.

Solução. Uma expressão para a potência do circuito pode ser deduzida das leis de Kirchhoff como sendo

$$P(R_a) = \frac{\left[\dfrac{VR_3R_a}{R_1(R_a + R_2 + R_3) + R_3R_a + R_3R_2}\right]^2}{R_a} \tag{16.11}$$

Substituir os valores dos parâmetros resulta no gráfico mostrado na Figura 16.9. Observe que a potência máxima transferida ocorre com uma resistência de cerca de 16 Ω.

Esse problema será resolvido de duas formas com a planilha Excel. Primeiro, aplicaremos tentativa e erro e a opção Solver. A seguir, desenvolveremos um macro em Visual Basic para fazer a análise de sensibilidade.

(*a*) Uma planilha Excel para implementar a Equação (16.11) é mostrada na Figura 16.10. Como indicado, a Equação (16.11) pode ser introduzida na célula B9. Assim, o valor de R_a (célula B8) pode variar por tentativa e erro até que uma potência máxima seja determinada. Para esse exemplo, o resultado é a potência de 30,03W e um valor ajustado no potenciômetro de $R_a = 16{,}44\ \Omega$.

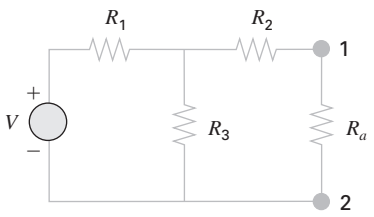


FIGURA 16.8
Um circuito resistor com um resistor ajustável ou potenciômetro.

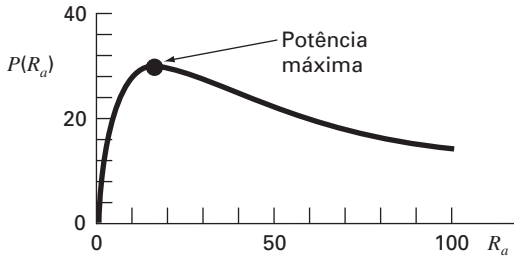


FIGURA 16.9
Um gráfico da potência transferida entre os terminais 1-2 da Figura 16.8 em função da resistência do potenciômetro R_a.

Uma abordagem ainda melhor envolve o uso da opção Solver no menu de Ferramentas da planilha. Nesse ponto, uma janela de diálogo aparecerá, perguntando por informação pertinente. As células pertinentes para a janela de diálogo do Solver devem ser preenchidas como

Set target cell:	B9	
Equal to ● max ❍ min ❍ equal to		0
By changing cells	B8	

Quando o botão OK for selecionado, uma janela de diálogo aparecerá com um relatório de sucesso da operação. Para o presente caso, o Solver obtém a mesma solução correta que é mostrada na Figura 16.10.

(b) Agora, embora a abordagem anterior seja excelente para um único cálculo, não é conveniente para casos nos quais múltiplas otimizações serão utilizadas — como na segunda parte dessa aplicação, quando o interesse é determinar como a potência máxima varia para diferentes valores escolhidos da voltagem. Naturalmente, o Solver pode ser invocado várias vezes para diferentes valores dos parâmetros, mas isso seria ineficiente. É preferível um caminho envolvendo o desenvolvimento de uma macrofunção que retornaria o ponto ótimo.

Tal função é listada na Figura 16.11. Observe quão grande é a semelhança com o pseudocódigo da busca da seção áurea apresentado previamente, na Figura 13.5. Além disso, observe que deve também ser definida uma função para calcular a potência segundo a Equação (16.11).

Uma planilha Excel utilizando essa macro para avaliar a sensibilidade da solução à voltagem é dada na Figura 16.12. É montada uma coluna de valores que gera o intervalo de valores de V (isto é, de 45 até 105 V). Uma chamada de função para a macro é escrita na célula B9, que referencia o valor adjacente de V (o 45, em A9). Além disso, os outros parâmetros no argumento da função também são incluídos. Observe que, enquanto a referência a V é relativa, a referência às aproximações superiores e inferiores e às resistências são absolutas (ou seja, incluindo o custo como dominante). Isso foi feito de modo que, quando a fórmula é copiada, as referências absolutas permanecem fixas enquanto as referências relativas correspondem à voltagem na mesma linha. Uma estratégia semelhante é usada para colocar a Equação (16.11) na célula C9.

Quando as fórmulas são copiadas para baixo, o resultado é como mostrado na Figura 16.12. A potência máxima pode ser traçada para mostrar o impacto das variações de voltagem. Como visto na Figura 16.13, a potência cresce com V.

FIGURA 16.10
Método de determinação pela planilha Excel da potência máxima em um potenciômetro, por tentativa e erro.

B9 = =(V*Res3*Ra/(Res1*(Ra+Res2+Res3)+Res3*Ra+Res3*Res2))^2/Ra

	A	B	C	D	E	F	G	H
1	**Potência Máxima Transferida**							
2								
3	V	80						
4	Res1	8						
5	Res2	12						
6	Res3	10						
7								
8	Ra	16.444444						
9	P(Ra)	**30.03003**						
10								

Os resultados para os valores correspondentes do potenciômetro (R_a) são mais interessantes. A planilha indica que para o mesmo valor de ajuste, 16,44 Ω, obtém-se a potência máxima. Tal resultado poderia ser difícil de intuir com base na inspeção casual da Equação (16.11).

FIGURA 16.11
Macro Excel escrito em Visual Basic para determinar o máximo mediante a busca da seção áurea.

```
Option Explicit

Function Golden(xlow, xhigh, R1, R2, R3, V)
Dim iter As Integer, maxit As Integer, ea As Double, es As Double
Dim fx As Double, xL As Double, xU As Double, d As Double, x1 as Double
Dim x2 As Double, f1 As Double, f2 As Double, xopt As Double
Const R As Double = (5 ^ 0.5 - 1) / 2
maxit = 50
es = 0.001
xL = xlow
xU = xhigh
iter = 1
d = R * (xU - xL)
x1 = xL + d
x2 = xU - d
f1 = f(x1, R1, R2, R3, V)
f2 = f(x2, R1, R2, R3, V)
If f1 > f2 Then
  xopt = x1
  fx = f1
Else
  xopt = x2
  fx = f2
End If
Do
  d = R * d
  If f1 > f2 Then
    xL = x2
    x2 = x1
    x1 = xL + d
    f2 = f1
    f1 = f(x1, R1, R2, R3, V)
  Else
    xU = x1
    x1 = x2
    x2 = xU - d
    f1 = f2
    f2 = f(x2, R1, R2, R3, V)
  End If
  iter = iter + 1
  If f1 > f2 Then
    xopt = x1
    fx =f1
Else
    xopt = x2
    fx = f2
  End If
  If xopt <> 0 Then ea = (1 - R) * Abs((xU - xL) / xopt) * 100
  If ea <= es Or iter >= maxit Then Exit Do
Loop
Golden = xopt
End Function

Function f(Ra, R1, R2, R3, V)
f = (V * R3 * Ra / (R1 * (Ra + R2 + R3) + R3 * Ra + R3 * R2)) ^ 2 / Ra
End Function
```

	A	B	C	D
1	**Potência Máxima Transferida**			
2				
3	R1	8		
4	R2	12		
5	R3	10		
6	Rmin	0.1		
7	Rmax	100		
8	V	Ra	P(Ra)	
9	45	**16,44444**	**9,501689**	
10	60	16.44444	16.89189	
11	75	16.44444	26.39358	
12	90	16.44444	38.00676	
13	105	16.44444	51.73142	

Chamada para a função macro em Visual Basic

= Golden($B$6,$B$7,$B$3,$B$4,$B$5,A9)

Cálculo da potência

=(A9*$B$5*B9/($B$3*(B9+$B$4+$B$5)+$B$5*B9+$B$3*$B$4))^2/B9

FIGURA 16.12
Planilha Excel para implementar uma análise de sensibilidade da potência máxima para as variações de voltagem. Essa rotina acessa o programa macro para a busca da seção áurea da Figura 16.11.

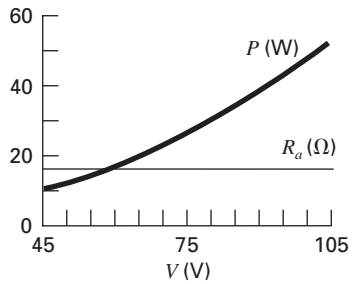


FIGURA 16.13
Resultados da análise de sensibilidade do efeito das variações de voltagem na potência máxima.

16.4 PROJETO DE UMA *MOUNTAIN BIKE* (ENGENHARIA MECÂNICA/AEROESPACIAL)

Fundamentos. Por causa do seu trabalho na indústria de construção, engenheiros civis são comumente associados ao projeto estrutural. No entanto, outras especialidades da engenharia também têm de lidar com o impacto de forças nos dispositivos que desenvolvem. Particularmente, engenheiros mecânicos e aeroespaciais devem avaliar tanto respostas estáticas quanto dinâmicas para uma grande variedade de veículos, de automóveis a veículos espaciais.

O interesse recente no ciclismo competitivo e recreativo tem obrigado os engenheiros a direcionar suas habilidades para o projeto e teste de *mountain bikes* (Figura 16.14*a*). Suponha que você foi escalado para prever o deslocamento horizontal e vertical do quadro da bicicleta em resposta a uma força. Suponha que as forças que você deve analisar possam ser simplificadas como descrito na Figura 16.4*b*. Você está interessado em testar a resposta do suporte a uma força exercida em uma direção qualquer, identificada pelo ângulo θ.

Os parâmetros para o problema são $E = 2 \times 10^{11}$ Pa, o módulo de Young, $A = 0{,}0001$ m^2, a área da seção transversal, $w = 0{,}44$ m, a largura, $\ell = 0{,}56$ m, o comprimento, e $h = 0{,}5$ m, a altura. Os deslocamentos x e y podem ser calculados determinando-se os valores que resultam na energia potencial mínima. Determine os deslocamentos para uma força de 10.000 N e um intervalo de θ's de 0º (horizontal) a 90º (vertical).

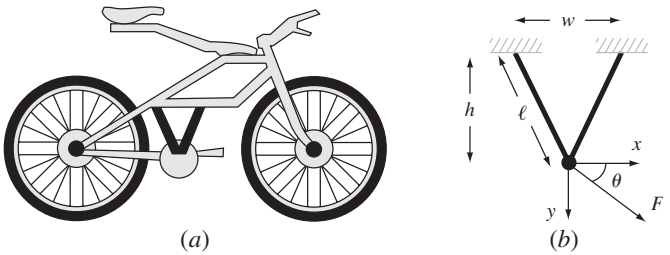


FIGURA 16.14
(*a*) Uma *mountain bike* juntamente com o (*b*) diagrama de corpo livre para uma parte da estrutura.

Solução. Esse problema pode ser abordado desenvolvendo-se a seguinte equação para a energia potencial do quadro,

$$V(x, y) = \frac{EA}{\ell}\left(\frac{w}{2\ell}\right)^2 x^2 + \frac{EA}{\ell}\left(\frac{h}{\ell}\right)^2 y^2 - Fx\cos\theta - Fy\ \text{sen}\ \theta \tag{16.12}$$

A solução para um ângulo particular é imediata. Para $\theta = 30^\circ$, os valores dados dos parâmetros podem ser substituídos na Equação (16.12) para obter

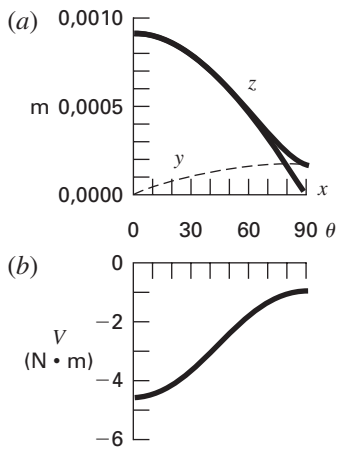
$$V(x, y) = 5{,}512{,}026x^2 + 28{,}471{,}210y^2 - 5000x - 8660y$$

O mínimo para essa função pode ser determinado de várias maneiras. Por exemplo, usando o Excel Solver, a energia potencial mínima é $-3{,}62$ com deflexões de $x = 0{,}000786$ e $y = 0{,}0000878$ m.

Naturalmente, o Excel Solver pode ser implementado repetidamente para diferentes valores de θ, de modo que será possível avaliar como a solução muda com o ângulo. Alternativamente, pode-se escrever uma macro da mesma forma que na Seção 16.3, de tal modo que múltiplas otimizações possam ser implementadas simultaneamente. É claro que, para esse caso, um algoritmo de busca multidimensional deve ser implementado. Uma terceira abordagem para o problema seria usar uma linguagem de programação como o Fortran 90 juntamente com uma biblioteca de software de métodos numéricos como o IMSL.

Qualquer que seja a abordagem, os resultados serão como os mostrados na Figura 16.15. Como era de esperar (Figura 16.15*a*), a deflexão x tem seu máximo quando a carga está

FIGURA 16.15
(*a*) Impacto de diferentes ângulos na deflexão (observe que z é a resultante das componentes x e y), e (*b*) energia potencial de uma parte do quadro de uma *mountain bike* sujeita a uma força constante.



apontando para a direção x ($\theta = 0^\circ$) e a deflexão y tem o seu máximo quando a carga aponta para a direção y ($\theta = 90^\circ$). Entretanto, observe que a deflexão na direção x é muito mais pronunciada que aquela na direção y. Isso também pode ser visto na Figura 16.15*b*, em que a energia potencial é maior para os ângulos menores. Ambos os resultados se devem à geometria do suporte. Se w se tornasse maior, as deflexões seriam mais uniformes.

PROBLEMAS

Engenharia Química/Bioengenharia

16.1 Projete um contêiner cilíndrico ótimo (Figura P16.1) aberto em uma das extremidades e que tem uma espessura desprezível. O contêiner deve conter 0,5 m^3. Faça o desenvolvimento de modo que a área da base mais a da lateral seja minimizada.

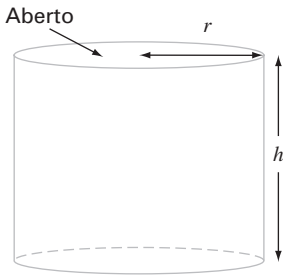


Figura P16.1
Um contêiner cilíndrico sem tampa.

16.2 **(a)** Projete um contêiner cônico ótimo (Figura P16.2) que tenha cobertura e paredes de espessura desprezível. O contêiner deve conter 0,5 m^3. Faça o desenvolvimento de modo que a área da base mais a da lateral seja minimizada. **(b)** Repita **(a)** para um contêiner cônico sem a cobertura.

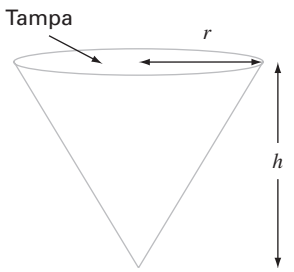


Figura P16.2
Um contêiner cônico com tampa.

16.3 A taxa de crescimento de uma cultura específica que produz um antibiótico é função da concentração de alimento c,

$$g = \frac{2c}{4 + 0{,}8c + c^2 + 0{,}2c^3}$$

Como mostrado na Figura P16.3, o crescimento tende a zero em concentrações muito baixas em decorrência da limitação de alimento, e também tende a zero em concentrações muito altas por causa do efeito de toxinas (toxicidade). Encontre o valor de c no qual o crescimento é máximo.

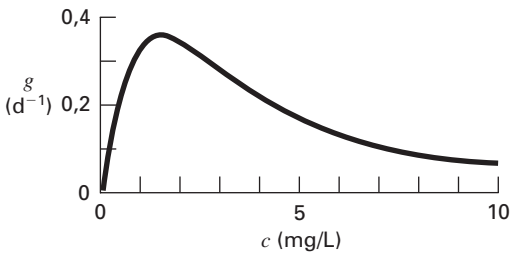


Figura P16.3
Taxa de crescimento específico de uma cultura que produz antibiótico *versus* a concentração de alimento.

16.4 Uma instalação química faz três produtos semanalmente. Cada um dos três produtos requer uma certa quantidade de matéria química bruta e diferentes tempos de produção e produz lucros diferentes. A informação pertinente se encontra na Tabela P16.4. Observe que há espaço suficiente no armazém da empresa para armazenar um total de 450 kg/semana.

(a) Escreva um problema de programação linear para maximizar o lucro.
(b) Resolva o problema de programação linear pelo método simplex.
(c) Resolva o problema com um pacote de software.
(d) Avalie qual das seguintes opções aumentará mais os lucros: aumentar a matéria química bruta, o tempo de produção ou o armazenamento.

16.5 Recentemente, engenheiros químicos têm-se envolvido em uma área chamada de *minimização de dejetos*. Isso envolve a operação de uma indústria química de modo que o impacto ambiental seja minimizado. Suponha que uma refinaria desenvolva um produto *Z*1 feito de duas substâncias *X* e *Y*. A produção de 1 tonelada métrica de produto envolve 1 ton de *X* e 2,5 ton de *Y* e produz 1 ton de dejeto líquido *W*. Os engenheiros propuseram três formas alternativas de lidar com o dejeto:

TABELA P16.4

	Produto 1	Produto 2	Produto 3	Recursos Disponíveis
Matéria química bruta	6 kg/kg	4 kg/kg	12 kg/kg	2.500 kg
Tempo de produção	0,05 h/kg	0,1 h/kg	0,2 h/kg	55 h/semana
Lucro	$30/kg	$30/kg	$35/kg	

- Produzir uma tonelada cúbica de um produto secundário Z2 acrescentando uma tonelada cúbica adicional de X a cada tonelada cúbica de W.
- Produzir uma tonelada cúbica de outro produto secundário Z3 acrescentando uma tonelada cúbica adicional de Y para cada tonelada cúbica de W.
- Tratar os dejetos de modo a poder descartá-los na natureza.

Os produtos fornecem lucro de \$ 2.000, –\$ 75 e \$ 250/ton para Z1, Z2 e Z3, respectivamente. Observe que a produção de Z2 ocasiona, de fato, um prejuízo. O processo de tratamento custa \$ 300/ton. Além disso, a companhia tem um limite de aquisição de até 7.500 e 12.500 toneladas cúbicas de X e Y, respectivamente, durante o período de produção. Determine quanto de produto e dejeto pode ser criado de modo a maximizar o lucro.

16.6 Uma mistura de benzeno e tolueno deve ser separada em um tanque. A que temperatura deve operar o tanque para atingir a máxima pureza do tolueno na fase líquida (maximizando x_T)? A pressão no tanque é de 800 mm de Hg. As unidades para a equação de Antoine são mm de Hg e ºC para pressão e temperatura, respectivamente.

$$x_B P_{\text{sat}B} + x_T P_{\text{sat}T} = P$$

$$\log_{10}(P_{\text{sat}B}) = 6{,}905 - \frac{1211}{T + 221}$$

$$\log_{10}(P_{\text{sat}T}) = 6{,}953 - \frac{1344}{T + 219}$$

16.7 Um composto A será convertido em B em um tanque de mistura reator. O produto B e o produto A não processado são purificados em uma unidade de separação. O produto A não processado é reciclado para o reator. Um engenheiro de processo descobriu que o custo inicial do sistema é função da taxa de conversão x_A. Encontre a taxa de conversão que resultará no mais baixo custo do sistema. C é uma constante de proporcionalidade.

$$\text{Custo} = C\left[\left(\frac{1}{(1-x_A)^2}\right)^{0{,}6} + 6\left(\frac{1}{x_A}\right)^{0{,}6}\right]$$

16.8 No problema 16.7, apenas um reator é usado. Se dois reatores forem usados em série, a equação que governa o sistema muda. Encontre a taxa de conversão para ambos os reatores x_{A1} e x_{A2} de modo que o custo total do sistema seja minimizado.

$$\text{Custo} = C\left[\left(\frac{x_{A1}}{x_{A2}(1-x_{A1})^2}\right)^{0{,}6} + \left(\frac{1-\left(\frac{x_{A1}}{x_{A2}}\right)}{(1-x_{A2})^2}\right)^{0{,}6} + 6\left(\frac{1}{x_{A2}}\right)^{0{,}6}\right]$$

16.9 Para a reação

$$2A + B \Leftrightarrow C$$

o equilíbrio pode ser expresso por:

$$K = \frac{[C]}{[A]^2[B]} = \frac{[C]}{[A_0 - 2C]^2[B_0 - C]}$$

Se $K = 2\ M^{-1}$, a concentração inicial de A (A_0) pode ser variada. A concentração inicial de B é fixada pelo processo, $B_0 = 100$. A custa \$ 1/M e C é vendido por \$ 10/M. Qual deveria ser a concentração inicial ótima de A para que os lucros fossem maximizados?

16.10 Uma indústria química precisa de 10^6 L/dia de solução. Três fontes estão disponíveis a diferentes preços e taxas de fornecimento. Cada fonte também tem uma concentração diferente de uma impureza que deve ser mantida abaixo de um nível mínimo para se prevenir interferência com a indústria química. Os dados para as três fontes estão resumidos na tabela a seguir. Determine a quantidade de cada fonte para se atingir a quantidade necessária com custo mínimo.

	Fonte 1	Fonte 2	Fonte 3	Necessidade
Custo (\$/L)	0,50	1,00	1,20	minimizar
Fornecimento (10^5 L/dia)	20	10	5	$\geq$10
Concentração (mg/L)	135	100	75	$\leq$100

16.11 Você deve projetar um canal triangular aberto para servir de esgoto para uma indústria química e que deságua em um tanque de estabilização (Figura P16.11). A velocidade média aumenta com o raio hidráulico $R_h = A/p$, onde A é a área da seção transversal e p é igual ao perímetro molhado. Uma vez que a taxa de escoamento máxima corresponde à máxima velocidade, o projeto ótimo corresponde à minimização do perímetro molhado. Determine as dimensões para minimizar o perímetro molhado para uma dada área da seção transversal. As dimensões relativas são universais?

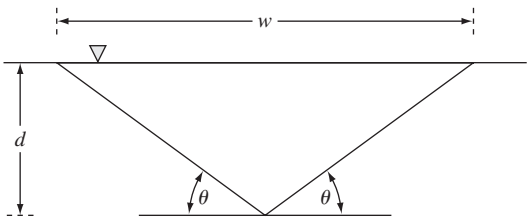


Figura P16.11

16.12 Como engenheiro agrícola, você deve projetar um canal trapezoidal aberto para transportar água de irrigação (Figura P16.12). Determine as dimensões ótimas para minimizar o perímetro molhado para uma área da seção transversal de 50 m^2. As dimensões relativas são universais?

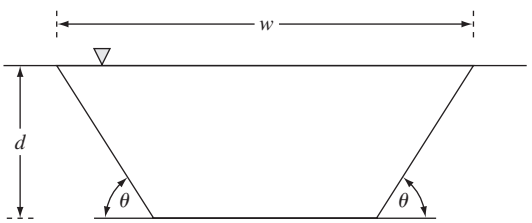


Figura P16.12

16.13 Encontre as dimensões ótimas para um tanque cilíndrico aquecido conter 10 m^3 de fluido. As extremidades e as laterais custam \$ 200/m^2 e \$ 100/m^2, respectivamente. Além disso, um revestimento é aplicado em toda a área do tanque a um custo de \$ 50/m^2.

Engenharia Civil/Ambiental

16.14 Um modelo de elementos finitos de uma viga em balanço sujeita a carga e momentos (Figura P16.14) é dada otimizando

$$f(x, y) = 5x^2 - 5xy + 2{,}5y^2 - x - 1{,}5y$$

onde x é o deslocamento da extremidade e y é o momento da extremidade. Encontre os valores de x e y que minimizam $f(x, y)$.

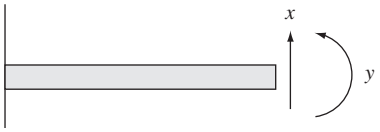


Figura P16.14
Uma viga em balanço.

16.15 Suponha que lhe é pedido para projetar uma coluna para suportar uma carga compressiva P, como mostrado na Figura P16.15*a*. A coluna tem a seção transversal na forma de um tubo com paredes finas, como mostrado na Figura P16.15*b*.

As variáveis de projeto são o diâmetro médio d do tubo e a espessura da parede t. O custo do tubo é calculado como sendo

$$\text{Custo} = f(t, d) = c_1 W + c_2 d$$

onde $c_1 = 4$ e $c_2 = 2$ são fatores de custo e W é o peso do tubo,

$$W = \pi\, dt\, H\rho$$

onde $\rho = 0{,}0025$ kg/cm^3 é a densidade do material do tubo. A coluna deve suportar a tensão da carga compressiva sem romper. Portanto,

Tensão efetiva (σ) $\leq$ resistência máxima à tensão de compressão $= \sigma_y = 550$ kg/cm^2
Tensão efetiva $\leq$ tensão de ruptura

A tensão efetiva é dada por

$$\sigma = \frac{P}{A} = \frac{P}{\pi dt}$$

A tensão de ruptura é dada por

$$\sigma_b = \frac{\pi EI}{H^2\, dt}$$

onde E é o módulo de elasticidade e I é o momento de inércia da área da seção transversal. O cálculo pode ser usado para mostrar que

$$I = \frac{\pi}{8}\, dt\,(d^2 + t^2)$$

Finalmente, os diâmetros dos tubos disponíveis estão entre d_1 e d_2 e a espessura, entre t_1 e t_2. Desenvolva e resolva esse problema determinando os valores de d e t que minimizam o custo. Observe que $H = 275$ cm, $P = 2.000$ kg, $E = 900.000$ kg/cm^2, $d_1 = 1$ cm, $d_2 = 10$ cm, $t_1 = 0{,}1$ cm e $t_2 = 1$ cm.

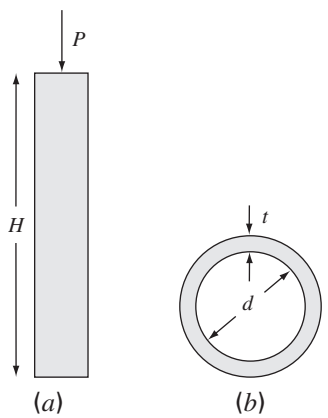


Figura P16.15
(*a*) Uma coluna suportando uma carga compressiva P.
(*b*) A seção transversal da coluna na forma de um tubo com paredes finas.

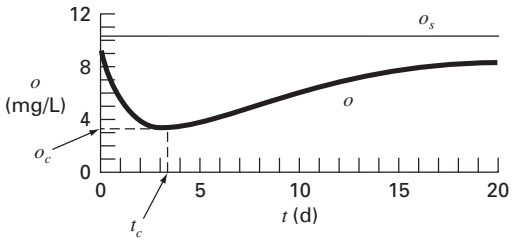


Figura P16.16
Queda do oxigênio latente abaixo de um ponto de descarga de esgoto em um rio.

16.16 O *modelo de Streeter-Phelps* pode ser usado para calcular a concentração de oxigênio dissolvido em um rio abaixo do ponto de descarga de água de esgoto (Figura P16.16),

$$o = o_s - \frac{k_d L_o}{k_d + k_s - k_a}\left(e^{-k_a t} - e^{-(k_d + k_s)t}\right) - \frac{S_b}{k_a}\left(1 - e^{-k_a t}\right) \tag{P16.16}$$

onde o é a concentração de oxigênio dissolvido [mg/L], o_s é a concentração de saturação do oxigênio [mg/L], t é o tempo de percurso [d], L_o é a concentração da demanda bioquímica de oxigênio (BOD) no ponto de mistura [mg/L], k_d é a taxa de decomposição de BOD [d^{-1}], k_s é a taxa de depósito do BOD [d^{-1}], k_a é a taxa de rearejamento [d^{-1}] e S_b é a demanda de oxigênio sedimentado [mg/L/d].

Como indicado na Figura P16.16, a Equação (P16.16) produz uma queda do oxigênio que atinge um nível crítico mínimo o_c para um tempo de percurso t_c abaixo do ponto de descarga. Esse ponto é chamado crítico porque representa a posição na qual o biossistema que mais depende de oxigênio (como os peixes) seria mais pressionado. Determine o tempo crítico de percurso e a concentração crítica, dados os seguintes valores:

$o_s = 10$ mg/L $\quad k_d = 0{,}2$ d^{-1} $\quad k_a = 0{,}8$ d^{-1}
$k_s = 0{,}06$ d^{-1} $\quad L_o = 50$ mg/L $\quad S_b = 1$ mg/L/d

16.17 A distribuição bidimensional da concentração de poluentes em um canal pode ser descrita por

$$c(x, y) = 7{,}7 + 0{,}15x + 0{,}22y - 0{,}05x^2 - 0{,}016y^2 - 0{,}007xy$$

Determine a localização exata do pico de concentração dados a função e o conhecimento de que o pico fica entre os limites $-10 \leq x \leq 10$ e $0 \leq y \leq 20$.

16.18 O escoamento Q [m^3/s] em um canal aberto pode ser previsto pela equação de Manning (lembre-se da Seção 8.2)

$$Q = \frac{1}{n} A_c R^{2/3} S^{1/2}$$

onde n é o coeficiente de rugosidade de Manning (um número adimensional usado para parametrizar o atrito no canal), A_c é a área da seção transversal do canal (m^2), S é a inclinação do canal (adimensional, metros em declive por metro de comprimento), e R é o raio hidráulico (m), o qual está relacionado a parâmetros mais fundamentais por $R = A_c/P$, onde P é o perímetro molhado (m). Como o nome indica, o perímetro molhado é o comprimento das laterais do canal e do fundo que está sob a água. Por exemplo, para um canal retangular, ele é definido como $P = B + 2H$, onde H é a profundidade (m). Suponha que você está usando essa fórmula para projetar um canal com revestimento (observe que fazendeiros revestem canais para minimizar perdas por vazamento).

(a) Dados os parâmetros $n = 0{,}03$, $S = 0{,}0004$ e $Q = 1$ m^3/s, determine os valores de B e H para minimizar o perímetro molhado. Observe que tais cálculos minimizariam os custos se os custos do revestimento fossem muito maiores do que o custo de escavação.

(b) Repita a parte (**a**) para incluir o custo de escavação. Para fazê-lo, minimize a seguinte função custo,

$$C = c_1 A_c + c_2 P$$

onde $c_1 = \$100/\text{m}^2$ é o fator de custo para a escavação e $c_2 = \$50/\text{m}$ é o fator de custo de revestimento.

(c) Discuta as implicações dos seus resultados.

16.19 Uma viga cilíndrica é carregada com uma carga de compressão $P = 3.000$ kN. Para prevenir a ruptura da viga, essa carga deve ser menor que a carga crítica,

$$P_c = \frac{\pi^2 EI}{L^2}$$

onde $E = 200 \times 10^9$ N/m^2 é o módulo de Young, $I = \pi r^4/4$ (o momento de inércia da área de uma viga cilíndrica de raio r), e L é o comprimento da viga. Se o volume V da viga não puder exceder 0.075 m^3, encontre o maior comprimento L que pode ser utilizado e o raio correspondente.

16.20 O rio Splash tem uma vazão 2×10^6 m^3/d, da qual 70% pode ser desviada para dois canais que escoam através do município de Splish. Esses canais são usados para transporte, irrigação e geração de energia elétrica, com os dois últimos sendo fonte de receita. O uso para transporte exige uma vazão desviada mínima de $0{,}3 \times 10^6$ m^3/d para o Canal 1 e $0{,}2 \times 10^6$ m^3/d para o Canal 2. Por razões políticas, foi decidido que a diferença absoluta entre as vazões não pode exceder 40% da vazão total desviada para os canais. O Comitê de Gerenciamento de Água do município de Splish também limitou os custos de manutenção do sistema de canais a não mais que $\$1{,}8 \times 10^6$ por ano. A estimativa de custo de manutenção anual é baseada na vazão diária. O custo anual do Canal 1 é estimado multiplicando- se \$1,1 por m^3/d de escoamento; enquanto para o Canal 2 o fator multiplicativo é \$1,4 por m^3/d. A receita anual da produção de energia elétrica é também estimada com base na vazão diária. Para o Canal 1, ela é de \$ 4,0 por m^3/d, enquanto que para o Canal 2 é de \$ 3,0 por m^3/d . A receita anual proveniente da irrigação também é estimada com base na vazão diária, mas as vazões devem ser primeiro corrigidas pela perda de água nos canais antes da entrega para a irrigação. Essa perda é de 30% no Canal 1 e de 20% no Canal 2. Em ambos os canais, a receita é \$ 3,2 por m^3/d. Determine as vazões nos canais que maximizem o lucro.

16.21 Determine a área da seção transversal da viga que resulta em um peso mínimo para a treliça estudada na Seção 12.2 (Figura 12.4). O rompimento crítico e a força elástica máxima de compressão e componentes de tensão são 10 e 20 ksi, respectivamente. A treliça deve ser construída com aço (densidade igual a 3,5 lb/pés-pol^2). Observe que o comprimento do membro horizontal (2) é 50 pés. Além disso, lembre-se de que a tensão em cada membro é igual à força dividida pela área da seção transversal. Estruture esse problema como um problema de programação linear. Obtenha a solução graficamente e com o Excel Solver.

Engenharia Elétrica

16.22 Uma carga total Q está uniformemente distribuída em um condutor na forma de um anel de raio a. Uma carga q é colocada a uma distância x do centro do anel (Figura P16.22). A força exercida na carga pelo anel é dada por

$$F = \frac{1}{4\pi e_0} \frac{qQx}{(x^2 + a^2)^{3/2}}$$

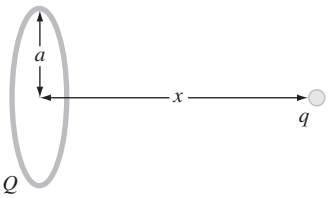


Figura P16.22

onde $e_0 = 8{,}85 \times 10^{-12}$ C^2/(N m^2), $q = Q = 2 \times 10^{-5}$ C e $a = 0{,}9$ m. Determine a distância x onde a força é máxima.

16.23 Um sistema consiste em duas usinas de força que devem liberar carga para uma rede de transmissão. Os custos da geração de energia das usinas 1 e 2 são dados por

$$F_1 = 2p_1 + 2$$
$$F_2 = 10p_2$$

onde p_1 e p_2 são a energia produzida por cada usina. As perdas de energia decorrentes da transmissão L são dadas por

$$L_1 = 0{,}2p_1 + 0{,}1p_2$$
$$L_2 = 0{,}2p_1 + 0{,}5p_2$$

A demanda total de energia é 30 e p_1 não pode exceder 42. Determine a geração de energia necessária para atender à demanda ao mesmo tempo em que minimiza o custo, usando uma rotina de otimização como aquelas encontradas, por exemplo, no Excel, no software MATLAB e no IMSL.

16.24 O torque transmitido a um motor de indução é função do escorregamento entre a rotação do campo estator e a velocidade do rotor s, com o escorregamento definido como

$$s = \frac{n - n_R}{n}$$

onde n é o número de revoluções por segundo do estator em rotação e n_R é a velocidade do rotor. As leis de Kirchhoff podem ser usadas para mostrar que o torque (expresso em uma forma adimensional) e o escorregamento estão relacionados por

$$T = \frac{15(s - s^2)}{(1 - s)(4s^2 - 3s + 4)}$$

A Figura P16.24 mostra essa função. Use um método numérico para determinar o escorregamento no qual ocorre o torque máximo.

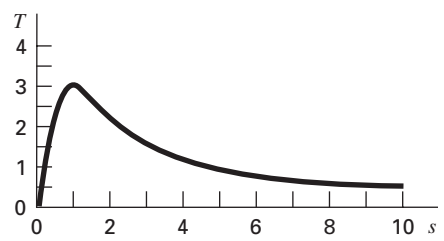


Figura P16.24
Torque transmitido a um indutor como função do escorregamento.

16.25

(a) Um fabricante de produtos computacionais produz *scanners* e impressoras. Os insumos necessários para a produção desses dispositivos e os lucros correspondentes são

Dispositivo	Capital ($/unid.)	Trabalho (h/unid)	Lucro ($/unid.)
Scanner	300	20	500
Impressora	400	10	400

Se existirem valores de $ 127.000 de capital e 4.270 horas de trabalho disponíveis a cada dia, quanto de cada dispositivo deve ser produzido por dia para maximizar o lucro?

(b) Repita o problema, mas agora suponha que o lucro para cada impressora vendida P_p depende do número de impressoras produzidas X_p, como em

$$P_p = 400 - X_p$$

16.26 Um fabricante fornece microchips especializados. Durante os próximos 3 meses suas vendas, seus custos e tempo disponível são

	Mês 1	Mês 2	Mês 3
Chips necessários	1.000	2.500	2.200
Custo de tempo normal ($/chip)	100	100	120
Custo da hora extra ($/chip)	110	120	130
Tempo normal de operação (h)	2.400	2.400	2.400
Hora extra (h)	720	720	720

Não existem chips em estoque no início do primeiro mês. Leva 1,5 hora para produzir um chip e custa $ 5 para armazenar um chip de um mês para outro. Determine o cronograma de produção que atenda aos requisitos sem exceder as limitações no tempo de produção mensal e minimize o custo. Observe que nenhum chip deve estar em estoque no final dos 3 meses.

Engenharia Mecânica/Aeroespacial

16.27 O arrasto total de um aerofólio pode ser estimado por

$$D = \underbrace{0{,}01\sigma V^2}_{\text{atrito}} + \underbrace{\frac{0{,}95}{\sigma}\left(\frac{W}{V}\right)^2}_{\text{sustentação}}$$

onde D é o arrasto, σ é a razão da densidade do ar entre a altitude de vôo e o nível do mar, W é o peso e V é a velocidade. Como visto na Figura P16.27, os dois fatores que contribuem para o arrasto são afetados de forma diferente por um aumento de velocidade. Enquanto o arrasto por atrito aumenta com a velocidade, o arrasto decorrente da sustentação decresce. A combinação desses dois fatores leva a um arrasto mínimo.

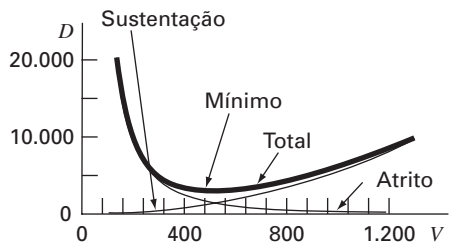


Figura P16.27
Gráfico do arrasto *versus* a velocidade para um aerofólio.

(a) Se $\sigma = 0{,}6$ e $W = 16.000$, determine o arrasto mínimo e a velocidade na qual ele ocorre.

(b) Além disso, desenvolva uma análise de sensibilidade para determinar como esse ponto ótimo varia em resposta a uma variação $W = 12.000$ a 20.000 com $\sigma = 0{,}6$.

16.28 Rolos de moinho estão sujeitos à fadiga causada pela grande força de contato F (Figura P16.28).

Pode-se mostrar que o problema de determinar a localização da tensão máxima ao longo do eixo x é equivalente a maximizar a função

$$f(x) = \frac{0{,}4}{\sqrt{1+x^2}} - \sqrt{1+x^2}\left(1 - \frac{0{,}4}{1+x^2}\right) + x$$

Encontre o x que maximiza $f(x)$.

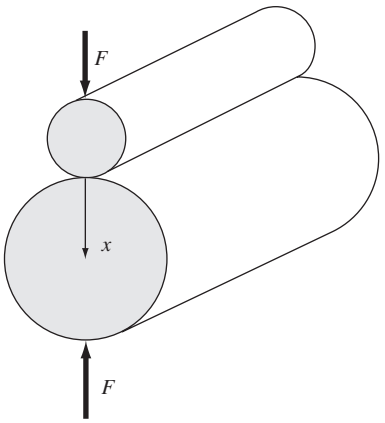


Figura P16.28
Rolos de moinho.

16.29 Uma companhia aeroespacial está desenvolvendo um novo aditivo de combustível para linhas aéreas comerciais. O aditivo é composto de três ingredientes: X, Y e Z. Para o pico de performance, a quantidade total de aditivo deve ser pelo menos 6 mL/L de combustível. Por razões de segurança, a soma dos ingredientes altamente inflamáveis X e Y não pode exceder 2,5 mL/L. Além disso, a quantidade do ingrediente X deve sempre ser igual ou maior à de Y e a quantidade de Z deve ser maior que a metade de Y. Se o custo por mL para os ingredientes X, Y e Z for 0,05, 0,025 e 0,15, respectivamente, determine o custo mínimo da mistura para cada litro de combustível.

16.30 Uma empresa de manufatura produz cinco tipos de peças de automóvel. Cada uma é primeiro fabricada e depois é feito o acabamento. As horas de trabalho necessárias e o lucro para cada peça são

	Peça			
	A	B	C	D
Tempo de fabricação (h/100 unids.)	2,5	1,5	2,75	2
Tempo de acabamento (h/100 unids.)	3,5	3	3	2
Lucro ($/100 unids.)	375	275	475	325

A capacidade de fabricação e acabamento para o próximo mês é de 640 e 960 horas, respectivamente. Determine quanto de cada peça deve ser produzido de modo a maximizar o lucro.



EPÍLOGO: PARTE QUATRO

Os epílogos das outras partes deste livro contêm uma discussão e um resumo tabular dos prós e contras dos vários métodos, bem como as fórmulas e relações importantes. A maioria dos métodos desta parte é bastante complicada e, conseqüentemente, eles não podem ser resumidos em fórmulas simples e resumos tabulares. Portanto, desviamo-nos um pouco para fornecer a narrativa a seguir discutindo os prós e contras e as referências adicionais dos métodos discutidos.

PT4.4 PRÓS E CONTRAS

O Capítulo 13 trata da busca do ponto ótimo de uma função de uma única variável sem restrições. O método de busca da seção áurea é um método intervalar que exige que se conheça um intervalo contendo um único ponto ótimo. Tem a vantagem de minimizar o número de cálculos da função e sempre converge. A interpolação quadrática também funciona bem quando implementada como um método intervalar, embora possa ser programada como um método aberto. Entretanto, em tais casos, pode divergir. Ambas, a busca da seção áurea e a interpolação quadrática, dispensam o cálculo das derivadas. Assim, os dois métodos são apropriados quando o intervalo pode ser prontamente definido e o cálculo da função é custoso.

O método de Newton é um método aberto que não requer que o ótimo esteja delimitado. Pode ser implementado de forma fechada quando a primeira e a segunda derivadas puderem ser determinadas analiticamente. Ele também pode ser implementado de forma similar ao método da secante, com representação em diferenças finitas para as derivadas. Embora o método de Newton convirja rapidamente próximo ao ponto ótimo, é freqüentemente divergente para uma estimativa ruim. A convergência também depende da natureza da função.

O Capítulo 14 cobre dois tipos de métodos gerais para resolver problemas de otimização multidimensionais, sem restrições. Métodos diretos, como busca aleatória e a busca unidimensional, não precisam do cálculo das derivadas da função e são, freqüentemente, ineficientes. Entretanto, eles também fornecem uma ferramenta para encontrar o ponto ótimo global em vez de um ponto ótimo local. Métodos de busca de padrão, como o método de Powell, podem ser muito eficientes e também não exigem o cálculo de derivadas.

Os métodos gradiente usam a primeira e, às vezes, a segunda derivada para encontrar o ponto ótimo. O método do aclive máximo ou declive fornece uma abordagem confiável, mas, algumas vezes, lenta. Em contraste, o método de Newton sempre converge rapidamente quando está nas vizinhanças de um ponto ótimo, porém, às vezes, diverge. O método de Marquardt usa o método do declive máximo em uma posição inicial longe da região do ponto ótimo e muda para o método de Newton próximo do ponto ótimo, em uma tentativa de aproveitar as vantagens de cada método.

O método de Newton pode ser computacionalmente custoso porque exige o cálculo tanto do vetor gradiente quanto da matriz hessiana. Abordagens quasi-Newton tentam contornar esses problemas usando aproximações que reduzam o número de cálculos de matrizes (particularmente o cálculo, o armazenamento e a inversão da hessiana).

As pesquisas atuais continuam a busca e a exploração das características e dos respectivos pontos fortes dos vários métodos híbridos e combinados. Alguns exemplos são o método do gradiente conjugado de Fletcher-Reevers e o método quasi-Newton de Davidson-Fletcher-Powell.

O Capítulo 15 foi dedicado à otimização com restrições. Para problemas lineares, a programação linear com base no método simplex fornece um meio eficiente para obter soluções. Abordagens como o método GRG estão disponíveis para resolver problemas não-lineares com restrições.

Pacotes de software e bibliotecas incluem uma grande variedade de recursos para otimização. A mais genérica é a biblioteca IMSL, que contém muitas sub-rotinas para implementar a maioria dos algoritmos padrão para otimização. Quando este livro foi publicado, o Excel tinha os recursos de otimização mais úteis na forma da sua ferramenta Solver, porque essa ferramenta é projetada para implementar as formas mais gerais de otimização — otimização não-linear com restrições — e, portanto, pode ser usada para resolver problemas em todas as áreas cobertas nessa parte do livro.

PT4.5 REFERÊNCIAS ADICIONAIS

Para problemas em uma dimensão, o método de Brent é um híbrido que tenta levar em conta a natureza da função para garantir tanto uma convergência lenta e segura para uma estimativa inicial pobre, quanto uma convergência rápida próxima do ponto ótimo. Veja Press et al. (1992) para detalhes. Para problemas multidimensionais, informação adicional pode ser encontrada em Dennis e Schnabel (1996), Fletcher (1980, 1981), Gill et al. (1981) e Luenberger (1984).

PARTE CINCO



AJUSTE DE CURVAS

PT5.1 MOTIVAÇÃO

Em geral, os dados são fornecidos em um conjunto discreto de valores entre um contínuo de possibilidades. Entretanto, pode ser necessário fazer estimativas em pontos que estão entre os valores discretos. Esta parte do livro descreve técnicas para ajustar curvas a tais dados, com o objetivo de obter estimativas intermediárias. Além disso, você pode precisar de uma versão simplificada de uma função complicada. Uma forma de fazer isso é calcular valores da função em alguns valores discretos no intervalo de interesse. Então, uma função mais simples pode ser desenvolvida para ajustar esses dados. Ambas as aplicações são conhecidas como *ajuste de curvas.*

Há duas abordagens gerais para o ajuste de curvas, as quais são distinguidas entre si com base na quantidade de erro associada com os dados. Primeiro, quando os dados exibirem um grau significativo de erro ou "ruído", a estratégia será encontrar uma única curva que represente a tendência geral dos dados. Como cada ponto individual pode estar incorreto, não será feito nenhum esforço para passar a curva por todos os pontos. Em vez disso, a curva é escolhida para seguir o padrão dos pontos considerados como um grupo. Uma abordagem dessa natureza é chamada de regressão por mínimos quadrados (Figura PT5.1*a*).

Segundo, quando se souber que os dados são muito precisos, a abordagem básica é ajustar uma curva ou uma série de curvas que passam diretamente por cada um dos pontos. Tais dados usualmente se originam de tabelas. Exemplos são os valores da densidade da água ou da capacidade térmica dos gases como função da temperatura. A estimativa de valores entre pontos discretos bem conhecidos é chamada *interpolação* (Figura PT5.1*b* e *c*).

PT5.1.1 Métodos Não-computacionais para o Ajuste de Curvas

O método mais simples para ajustar uma curva aos dados é marcar os pontos e então esboçar uma curva que visualmente esteja de acordo com os dados. Embora essa seja uma opção válida quando for necessária uma estimativa rápida, os resultados dependem do ponto de vista subjetivo da pessoa esboçando a curva.

Por exemplo, a Figura PT5.1 mostra esboços deduzidos do mesmo conjunto de dados por três engenheiros. O primeiro não tentou conectar os pontos, e, em vez disso, caracterizou a tendência geral para cima dos dados por uma reta (Figura PT5.1*a*). O segundo engenheiro usou segmentos de reta ou interpolação linear para ligar os pontos (Figura PT5.1*b*). Essa é uma prática muito comum em engenharia. Se os valores forem verdadeiramente próximos de serem lineares ou se estiverem espaçados próximos uns dos outros, tal aproximação fornece estimativas que são adequadas para muitos cálculos em engenharia. Entretanto, se a relação envolvida for altamente curvilínea ou se os dados estiverem espaçados muito longe uns dos outros, erros significativos podem ser introduzidos por tal interpolação linear. O terceiro engenheiro usou curvas para tentar capturar as oscilações sugeridas pelos dados (Figura PT5.1*c*). Um quarto ou um quinto engenheiros provavelmente desenvolveriam ajustes alternativos. Obviamente, o intuito aqui é desenvolver métodos sistemáticos e objetivos com o propósito de encontrar tais curvas.

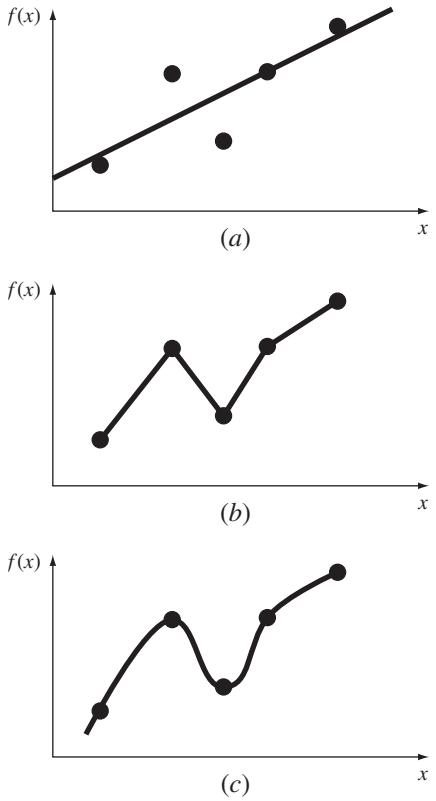


FIGURA PT5.1
Três tentativas de ajustar a "melhor" curva aos cinco pontos dados. (*a*) Regressão por mínimos quadrados, (*b*) interpolação linear e (*c*) interpolação curvilínea.

PT5.1.2 Ajuste de Curvas e a Prática da Engenharia

Sua primeira exposição ao ajuste de curvas pode ter sido determinar valores intermediários a partir de dados tabulados — por exemplo, a partir de tabelas de juros em engenharia econômica ou de tabelas de evaporação em termodinâmica. Pelo resto de sua carreira, você terá ocasiões freqüentes de fazer estimativas de valores intermediários a partir de tais tabelas.

Embora muitas das propriedades de engenharia amplamente usadas sejam tabuladas, há muitas outras que não estão disponíveis nessa forma conveniente. Casos especiais e novos contextos de problemas freqüentemente exigem que você meça seus próprios dados e desenvolva sua relação de previsão. Em geral, são encontrados dois tipos de aplicações nas quais dados experimentais são ajustados: análise de tendência e teste de hipótese.

A análise de tendência representa o processo de usar o padrão dos dados para fazer previsões. Nos casos nos quais os dados foram medidos com alta precisão, você poderia usar interpolação polinomial. Os dados imprecisos são, em geral, analisados com regressão por mínimos quadrados.

A *análise de tendência* pode ser usada para prever valores da variável dependente. Isso pode envolver extrapolação além dos limites dos dados observados ou interpolação dentro do intervalo dos dados. Todos os campos da engenharia comumente envolvem problemas desse tipo.

A segunda aplicação, em engenharia, do ajuste de curvas a dados experimentais é no *teste de hipótese*. Aqui, um modelo matemático já existente é comparado com os dados medidos. Se os coeficientes do modelo forem desconhecidos, pode ser necessário deter-

minar valores que melhor ajustem os dados observados. Por outro lado, se estimativas para os coeficientes dos modelos já estiverem disponíveis, pode ser apropriado comparar os valores previstos pelo modelo com os valores observados para testar a adequação do modelo. Em geral, modelos alternativos são comparados e o "melhor" é escolhido com base nas observações.

Além dessas aplicações em engenharia, o ajuste de curvas é importante em outros métodos numéricos, como integração e solução aproximada de equações diferenciais. Finalmente, as técnicas de ajuste de curvas podem ser usadas para encontrar funções simples que aproximem funções complicadas.

PT5.2 FUNDAMENTOS MATEMÁTICOS

Os fundamentos matemáticos necessários para interpolação são encontrados no material sobre expansões em séries de Taylor e diferenças divididas finitas introduzidas no Capítulo 4. A regressão por mínimos quadrados exige informação adicional do campo da estatística. Se você estiver familiarizado com os conceitos de média, desvio-padrão, soma dos quadrados residual, distribuição normal e intervalos de confiança, sinta-se livre para pular as próximas páginas e proseguir diretamente para PT5.3. Se você não estiver familiarizado com esses conceitos ou precisar de uma revisão, o objetivo do material a seguir é uma breve introdução a tais tópicos.

PT5.2.1 Estatísticas Simples

Suponha que, no decorrer de um estudo em engenharia, tenham sido feitas diversas medidas de uma quantidade particular. Por exemplo, a Tabela PT5.1 contém 24 leituras do coeficiente de expansão térmica de um aço estrutural. Tomados como valores de face, os dados fornecem uma quantidade limitada de informação — isto é, os valores variam de um mínimo de 6,395 a um máximo de 6,775. Informação adicional pode ser obtida resumindo-se os dados em uma ou mais estatísticas bem escolhidas que transportem tanta informação quanto possível sobre as características específicas desse conjunto de dados. Tais estatísticas descritivas são, em geral, escolhidas para representar (1) a posição do centro da distribuição de dados e (2) o grau de dispersão do conjunto de dados.

A estatística de posição mais comum é a média aritmética. A *média aritmética* ($\bar{y}$) de uma amostra é definida como a soma dos pontos individuais dados (y_i) dividida pelo número de pontos (n), ou

$$\bar{y} = \frac{\Sigma y_i}{n} \tag{PT5.1}$$

onde a somatória (e todas as somatórias posteriores nesta introdução) vai de $i = 1$ até n.

A medida de dispersão mais comum para uma amostra é o *desvio-padrão* (s_y) em torno da média,

$$s_y = \sqrt{\frac{S_t}{n-1}} \tag{PT5.2}$$

TABELA PT5.1 Medidas do coeficiente de expansão térmica do aço estrutural [$\times 10^{-6}$ pol/(pol · °F)].

6,495	6,595	6,615	6,635	6,485	6,555
6,665	6,505	6,435	6,625	6,715	6,655
6,755	6,625	6,715	6,575	6,655	6,605
6,565	6,515	6,555	6,395	6,775	6,685

onde S_t é a soma total dos quadrados dos resíduos entre os pontos dados e a média, ou

$$S_t = \Sigma(y_i - \bar{y})^2 \tag{PT5.3}$$

Portanto, se as medidas individuais estiverem amplamente dispersas em torno da média, S_t (e, conseqüentemente, s_y) será grande. Se elas estiverem bem agrupadas, o desvio-padrão será pequeno. A dispersão também pode ser representada pelo quadrado do desvio-padrão, o qual é chamado de *variância*:

$$s_y^2 = \frac{\Sigma(y_i - \bar{y})^2}{n - 1} \tag{PT5.4}$$

Observe que o denominador em ambas as Equações (PT5.2) e (PT5.4) é $n - 1$. A quantidade $n - 1$ é chamada de graus de liberdade. Portanto, dizemos que S_t e s_y são baseados em $n - 1$ *graus de liberdade*. Essa nomenclatura vem do fato de a soma das quantidades nas quais S_t é baseada (isto é, $\bar{y} - y_1, \bar{y} - y_2, \ldots, \bar{y} - y_n$) ser zero. Conseqüentemente, se $\bar{y}$ for conhecido e $n - 1$ valores forem especificados, o valor remanescente está fixado. Portanto, diz-se que apenas $n - 1$ dos valores são livremente determinados. Outra justificativa para dividir por $n - 1$ é o fato de que não existe tal coisa como a dispersão de um único ponto dado. Para os casos em que $n = 1$, as Equações (PT5.2) e (PT5.4) fornecem resultados infinitos, sem significado.

Deve ser observado que uma fórmula alternativa, mais conveniente, está disponível para calcular o desvio-padrão,

$$s_y^2 = \frac{\Sigma y_i^2 - (\Sigma y_i)^2 / n}{n - 1}$$

Essa versão não exige o cálculo prévio de $\bar{y}$ e fornece um resultado idêntico ao da Equação (PT5.4).

Uma estatística final que tem utilidade na quantificação da dispersão dos dados é o *coeficiente de variação* (*c.v.*). Essa estatística é a razão do desvio-padrão pela média. Como tal, o *c.v.* fornece uma medida normalizada da dispersão. Ele é, em geral, multiplicado por 100, de modo que possa ser expresso como uma porcentagem:

$$\text{c.v.} = \frac{s_y}{\bar{y}} 100\% \tag{PT5.5}$$

Observe que o coeficiente de variação é parecido em espírito com o erro relativo porcentual (ε_t) discutido na Seção 3.3, ou seja, é a razão entre uma medida de erro (s_y) e uma estimativa do valor verdadeiro ($\bar{y}$).

EXEMPLO PT5.1 Estatísticas Simples de uma Amostra

Enunciado do Problema. Calcule a média, a variância, o desvio-padrão e o coeficiente de variação para os dados da Tabela PT5.1.

Solução. Os dados são somados (Tabela PT5.2) e os resultados são usados para calcular [Equação (PT5.1)]

$$\bar{y} = \frac{158{,}4}{24} = 6{,}6$$

TABELA PT5.2 Cálculos das estatísicas para as leituras do coeficiente de expansão térmica. As freqüências e as extremidades foram desenvolvidos para construir o histograma na Figura PT5.2.

				Intervalo	
i	y_i	$(y_i - \bar{y})^2$	Freqüência	Extremidade Inferior	Extremidade Superior
1	6,395	0,042025	1	6,36	6,40
2	6,435	0,027225	1	6,40	6,44
3	6,485	0,013225			
4	6,495	0,011025			
5	6,505	0,009025	4	6,48	6,52
6	6,515	0,007225			
7	6,555	0,002025			
8	6,555	0,002025	2	6,52	6,56
9	6,565	0,001225			
10	6,575	0,000625			
11	6,595	0,000025	3	6,56	6,60
12	6,605	0,000025			
13	6,615	0,000225			
14	6,625	0,000625			
15	6,625	0,000625	5	6,60	6,64
16	6,635	0,001225			
17	6,655	0,003025			
18	6,655	0,003025			
19	6,665	0,004225	3	6,64	6,68
20	6,685	0,007225			
21	6,715	0,013225			
22	6,715	0,013225	3	6,68	6,72
23	6,755	0,024025	1	6,72	6,76
24	6,775	0,030625	1	6,76	6,80
Σ	158,4	0,217000			

Como na Tabela PT5.2, a soma dos quadrados dos resíduos é 0,217000, o qual pode ser usado para calcular o desvio-padrão [Equação (PT5.2)]:

$$s_y = \sqrt{\frac{0{,}217000}{24-1}} = 0{,}097133$$

a variância [Equação (PT5.4)]:

$$s_y^2 = 0{,}009435$$

e o coeficiente de variação [Equação (PT5.5)]:

$$\text{c.v.} = \frac{0{,}097133}{6{,}6}100\% = 1{,}47\%$$

PT5.2.2 A Distribuição Normal

Uma outra característica relacionada com a discussão presente é a *distribuição de dados* — isto é, a forma como os dados estão dispersos em torno da média. Um histograma fornece uma representação visual simples da distribuição. Como visto na Tabela PT5.2, o histograma é construído classificando-se as medidas em intervalos. As unidades de medidas são marcadas na abscissa e as freqüências de ocorrência em cada intervalo são marcadas na ordenada. Portanto, cinco das medidas estão entre 6,60 e 6,64. Como na Figura PT5.2, o histograma sugere que em sua maioria os dados estão agrupados próximo do valor médio de 6,6.

Se tivéssemos um conjunto muito grande de dados, o histograma em geral poderia ser aproximado por uma curva lisa. A curva simétrica, em forma de sino, superposta na Figura PT5.2 é uma de tais formas características — a *distribuição normal*. Se fossem dadas medidas adicionais suficientes, o histograma para esse caso particular poderia eventualmente se aproximar da distribuição normal.

Os conceitos de média, desvio-padrão, soma dos quadrados dos resíduos e distribuição normal têm todos grande relevância na prática da engenharia. Um exemplo muito simples é a sua utilização para determinar a confiança que pode ser associada a uma medida particular. Se uma quantidade for normalmente distribuída, o intervalo definido por $\bar{y} - s_y$ a $\bar{y} + s_y$ irá conter aproximadamente 68% do total das medidas. Analogamente, o intervalo definido por $\bar{y} - 2s_y$ a $\bar{y} + 2s_y$ irá conter aproximadamente 95%.

Por exemplo, para os dados na Tabela PT5.1 ($\bar{y} = 6{,}6$ e $s_y = 0{,}097133$), pode-se fazer a afirmação de que aproximadamente 95% das leituras deveriam estar entre 6,405734 e 6,794266. Se alguém diz que mediu um valor de 7,35, pode-se suspeitar que a medida esteja errada. A seção seguinte discorre sobre tais cálculos.

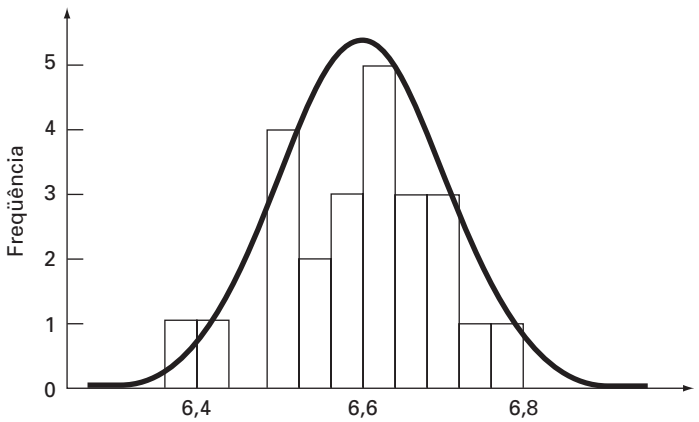
PT5.2.3 Estimativas de Intervalos de Confiança

Como deveria ter ficado claro das seções anteriores, um dos objetivos primários da estatística é fazer estimativas das propriedades de uma *população* com base em uma *amostra* limitada escolhida da população. Claramente, é impossível medir o coeficiente de dilatação térmica para todos os pedaços de aço estrutural que tenham sido produzidos. Conseqüentemente, como visto nas Tabelas PT5.1 e PT5.2, pode-se fazer aleatoriamente algumas medidas e, com base na amostra, tentar caracterizar as propriedades da população inteira.

Como nós vamos "inferir" propriedades da população desconhecida a partir de uma amostra limitada, o esforço é chamado de *inferência estatística*. Pelo fato de os resultados, em geral, serem expressos como estimativas sobre os parâmetros da população, o processo também é conhecido como *estimação*.

Já foi mostrado como fazer uma estimativa da tendência central (média da amostra, $\bar{y}$) e da dispersão (desvio-padrão e variância da amostra) de uma amostra limitada. Agora, será descrito brevemente como se pode associar afirmações probabilísticas à qualidade dessas estimativas. Em particular, discutiremos como se pode definir um intervalo de confiança em torno da estimativa da média. Foi escolhido esse tópico particular por causa de sua relevância direta nos modelos de regressão descritos no Capítulo 17.

FIGURA PT5.2
Um histograma usado para descrever uma distribuição de dados. Conforme o número de pontos dados aumenta, o histograma poderia aproximar-se de uma curva lisa, em forma de sino, chamada distribuição normal.



Observe que, na discussão a seguir, os símbolos $\bar{y}$ e s_y se referem à média e ao desvio-padrão da amostra, respectivamente. Os símbolos μ e σ se referem à média e ao desvio-padrão da população, respectivamente. As primeiras são, algumas vezes, chamadas de média e desvio-padrão "estimados", enquanto que as últimas são às vezes chamadas de média e desvio-padrão "verdadeiros".

Um *intervalo estimador* dá o intervalo de valores no qual é esperado que o parâmetro esteja com uma dada probabilidade. Tais intervalos são descritos como sendo unilaterais ou bilaterais. Como o nome indica, *um intervalo unilateral* expressa a confiança de que o parâmetro estimado seja menor que ou maior que o valor verdadeiro. Em contraste, o *intervalo bilateral* trata da afirmação mais geral de que a estimativa esteja de acordo com o valor verdadeiro, sem nenhuma consideração com o sinal da discrepância. Vamos nos concentrar no intervalo bilateral por ser ele mais geral.

Um intervalo bilateral pode ser descrito pela expressão

$$P\{L \leq \mu \leq U\} = 1 - \alpha$$

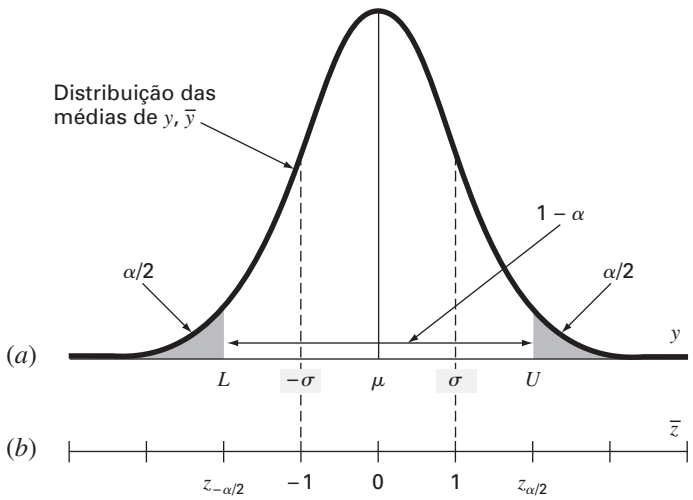
a qual é lida: "a probabilidade de que a média verdadeira de y, μ, caia dentro dos limites de L a U é $1 - \alpha$". A quantidade α é chamada de *nível de significância*. Assim, o problema de definir um intervalo de confiança se reduz a fazer estimativas de L e de U. Embora não seja absolutamente necessário, é costumeiro considerar o intervalo bilateral com a probabilidade α distribuída uniformemente como $\alpha/2$ em cada cauda da distribuição, como na Figura PT5.3.

Se a variância verdadeira da distribuição de y, σ^2, for conhecida (o que, em geral, não é o caso), a teoria estatística afirma que a média da amostra $\bar{y}$ vem de uma distribuição normal com média μ e variância σ^2/n (Quadro PT5.1). No caso ilustrado na Figura PT5.3, na realidade não se conhece μ. Portanto, não se sabe onde a curva normal está exatamente localizada em relação a $\bar{y}$. Para contornar esse dilema, calcula-se uma nova quantidade, a *estimativa normal padrão*

$$\bar{z} = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \tag{PT5.6}$$

a qual representa a distância normalizada entre $\bar{y}$ e μ. De acordo com a teoria estatística, essa quantidade deveria ser normalmente distribuída, com média 0 e variância 1.

FIGURA PT5.3
Um intervalo de confiança bilateral. A escala da abscissa em (*a*) é escrita em unidades naturais da variável aleatória *y*. A versão normalizada da abscissa em (*b*) tem a média na origem e faz uma mudança de escala de modo que o desvio-padrão corresponda a um valor unitário.



Quadro PT5.1 Um Pouco de Estatística

Em sua maioria, os engenheiros fazem diversas disciplinas para se tornarem proficientes em estatística. Como você pode ainda não ter feito tais cursos, gostaríamos de mencionar algumas idéias que poderiam tornar a presente seção mais coerente.

Como já afirmado, o "jogo" da inferência estatística supõe que a variável aleatória da qual você está tomando amostras, y, tem uma média (μ) e uma variância (σ^2) verdadeiras. Além disso, na discussão presente, supõe-se também que ela tem uma distribuição particular: a distribuição normal. A variância dessa distribuição normal tem um valor finito que especifica a "largura" da distribuição normal. Se a variância for grande, a distribuição é larga. Reciprocamente, se a variância for pequena, a distribuição é estreita. Logo, a variância verdadeira quantifica a incerteza intrínseca da variável aleatória.

No jogo da estatística, toma-se um número limitado de medidas dessa quantidade, chamada de uma amostra. A partir dessa amostra, pode-se calcular uma média ($\bar{y}$) e uma variância (s_y^2) estimadas. Quanto mais medidas forem tomadas, melhor será a aproximação do valor real pela estimativa. Isto é, quando $n \to \infty$, $\bar{y} \to \mu$ e $s_y^2 \to \sigma^2$.

Suponha que tomemos n amostras e calculemos uma média estimada $\bar{y}_1$. Depois, tomamos outras n amostras e calculamos uma outra, $\bar{y}_2$. Podemos continuar a repetir esse processo até termos gerado uma amostra de médias: $\bar{y}_1, \bar{y}_2, \bar{y}_3, \ldots, \bar{y}_m$, onde m é grande. Podemos então fazer um histograma dessas médias e determinar uma "distribuição das médias", bem como uma "média das médias". Agora, aparece a seguinte questão: essa nova distribuição de médias e sua estatística se comportam de uma forma previsível?

Existe um teorema extremamente importante, conhecido como *Teorema do Limite Central*, que responde diretamente a essa questão. Ele pode ser enunciado como

Sejam $y_1, y_2, \ldots, y_n$ uma amostra aleatória de tamanho n de uma distribuição de média μ e variância σ^2. Então, para n grande, $\bar{y}$ é aproximadamente normal, com média μ e variância σ^2/n. Além disso, para n grande, a variável aleatória $(\bar{y} - \mu)/(\sigma/\sqrt{n})$ é aproximadamente normal padrão.

Portanto, o teorema afirma o resultado notável que a distribuição das médias será sempre normalmente distribuída, independentemente da distribuição subjacente da variável aleatória! Ele também fornece o resultado esperado de que, dada uma amostra suficientemente grande, a média das médias deveria convergir para a média verdadeira da população, μ.

Além disso, esse teorema diz que, conforme o tamanho das amostras se torna maior, a variância das médias deve aproximar-se de zero. Isso faz sentido, porque se n for pequeno, nossa estimativa individual da média deverá ser sofrível e a variância das médias será grande. Conforme n aumenta, nossas estimativas da média melhorarão e, portanto, sua dispersão deveria diminuir. O Teorema do Limite Central define de forma simples exatamente como essa diminuição se relaciona tanto com a variância verdadeira quanto com o tamanho da amostra, ou seja, como σ^2/n.

Finalmente, o teorema afirma o importante resultado dado como Equação (PT5.6). Como mostrado nesta seção, esse resultado é a base para a construção de intervalos de confiança para a média.

Além disso, a probabilidade de que $\bar{z}$ caia dentro da região não sombreada da Figura PT5.3 deveria ser $1 - \alpha$. Portanto, pode-se fazer a afirmação que

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} < -z_{\alpha/2} \qquad \text{ou} \qquad \frac{\bar{y} - \mu}{\sigma/\sqrt{n}} > z_{\alpha/2}$$

com probabilidade α.

A quantidade $z_{\alpha/2}$ é uma variável aleatória normal padrão. Essa é a distância medida ao longo do eixo normalizado acima e abaixo da média, que delimita a probabilidade $1 - \alpha$ (Figura PT5.3*b*). Os valores de $z_{\alpha/2}$ estão tabulados em livros de estatística (por exemplo, Milton e Arnold, 1995). Eles também podem ser calculados usando-se funções em pacotes de software e em bibliotecas como o Excel e o IMSL. Como um exemplo, para $\alpha = 0{,}05$ (em outras palavras, definindo um intervalo delimitando 95%), $z_{\alpha/2}$ é igual a cerca de 1,96. Isso significa que um intervalo em torno da média de largura $\pm 1{,}96$ vez o desvio-padrão delimitará aproximadamente 95% da distribuição.

Esses resultados podem ser reorganizados para fornecer

$$L \le \mu \le U$$

com probabilidade $1 - \alpha$, onde

$$L = \bar{y} - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \qquad U = \bar{y} + \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \tag{PT5.7}$$

Agora, embora o precedente forneça uma estimativa para L e para U, ela é baseada no conhecimento da variância verdadeira σ. No nosso caso, conhecemos apenas a variância estimada s_y. Uma alternativa simples seria deduzir uma versão da Equação (PT5.6) baseada em s_y,

$$t = \frac{\bar{y} - \mu}{s_y/\sqrt{n}} \tag{PT5.8}$$

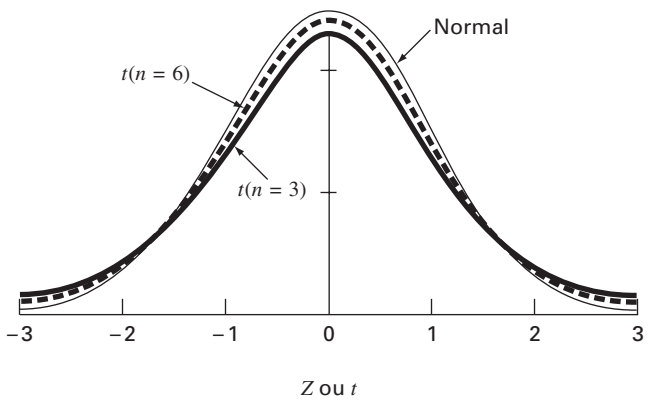
Mesmo quando são tomadas amostras de uma distribuição normal, essa fração não será normalmente distribuída, particularmente quando n for pequeno. Foi descoberto por W. S. Gossett que a variável aleatória definida pela Equação (PT5.8) segue a chamada *distribuição t de Student* ou, simplesmente, *distribuição t.* Nesse caso,

$$L = \bar{y} - \frac{s_y}{\sqrt{n}} t_{\alpha/2,n-1} \qquad U = \bar{y} + \frac{s_y}{\sqrt{n}} t_{\alpha/2,n-1} \tag{PT5.9}$$

onde $t_{\alpha/2,\,n-1}$ é a variável aleatória padrão para a distribuição t para uma probabilidade de $\alpha/2$. Como era o caso para $z_{\alpha/2}$, os valores estão tabulados em livros de estatística e também podem ser calculados usando-se pacotes de software e bibliotecas. Por exemplo, se $\alpha = 0{,}05$ e $n = 20$, $t_{\alpha/2,n-1} = 2{,}086$.

A distribuição t pode ser considerada como uma modificação da distribuição normal que leva em conta o fato de se ter uma estimativa imperfeita do desvio-padrão. Quando n for pequeno, ela tende a ser mais achatada do que a normal (ver Figura PT5.4). Portanto, para um número pequeno de medidas, ela fornece intervalos de confiança mais largos e, assim, mais conservadores. Conforme n se torna grande, a distribuição t converge para a normal.

FIGURA PT5.4
Comparação da distribuição normal com a distribuição t, para $n = 3$ e $n = 6$. Observe como a distribuição t é em geral mais achatada.



EXEMPLO PT5.2

Intervalo de Confiança da Média

Enunciado do Problema. Determine a média e o intervalo de confiança de 95% para os dados da Tabela PT5.1. Faça três estimativas baseadas (a) nas primeiras 8, (b) nas primeiras 16 e (c) em todas as 24 medidas.

Solução. (a) A média e o desvio-padrão para os primeiros 8 pontos são

$$\bar{y} = \frac{52{,}72}{8} = 6{,}59 \qquad s_y = \sqrt{\frac{347{,}4814 - (52{,}72)^2/8}{8-1}} = 0{,}089921$$

A estatística t apropriada pode ser calculada como

$$t_{0{,}05/2,8-1} = t_{0{,}025,7} = 2{,}364623$$

a qual pode ser usada para calcular o intervalo

$$L = 6{,}59 - \frac{0{,}089921}{\sqrt{8}} 2{,}364623 = 6{,}5148$$

$$U = 6{,}59 + \frac{0{,}089921}{\sqrt{8}} 2{,}364623 = 6{,}6652$$

ou

$$6{,}5148 \leq \mu \leq 6{,}6652$$

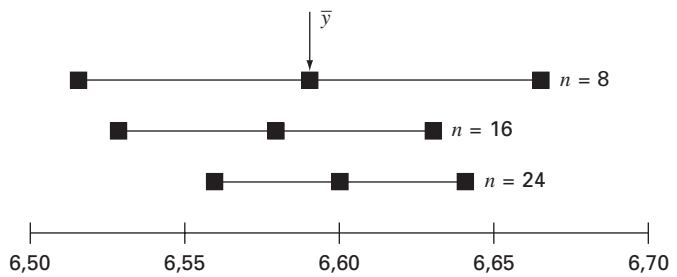


FIGURA PT5.5
Estimativas da média e intervalos de confiança de 95% para tamanhos diferentes da amostra.

Portanto, com base nas primeiras oito medidas, conclui-se que existe uma probabilidade de 95% de a média verdadeira cair no intervalo de 6,5148 a 6,6652.

Os outros dois casos para (*b*) 16 pontos e (*c*) 24 pontos podem ser calculados de maneira análoga e os resultados estão tabulados junto com o caso (*a*) em

n	$\bar{y}$	s_y	$t_{\alpha/2,n-1}$	L	U
8	6,5900	0,089921	2,364623	6,5148	6,6652
16	6,5794	0,095845	2,131451	6,5283	6,6304
24	6,6000	0,097133	2,068655	6,5590	6,6410

Esses resultados, que também estão resumidos na Figura PT5.5, indicam o resultado esperado que o intervalo de confiança se torna mais estreito conforme n cresce. Então, quanto mais medidas tomarmos, mais refinada será nossa estimativa do valor verdadeiro.

Esse é apenas um exemplo simples de como a estatística pode ser usada para fazer julgamentos com relação a dados incertos. Tais conceitos também terão relevância direta na discussão de modelos de regressão. Você pode consultar qualquer livro de estatística básica (por exemplo, Milton e Arnold, 1995) para obter informações adicionais sobre o assunto.

PT5.3 ORIENTAÇÃO

Antes de prosseguir para os métodos numéricos para ajuste de curvas, alguma orientação pode ser útil. O propósito do que vem a seguir é dar uma visão geral do material discutido na Parte Cinco. Além disso, foram formulados alguns objetivos para lhe ajudar a focalizar seus esforços quando estudar o material.

PT5.3.1 Escopo e Visão Geral

A Figura PT5.6 fornece uma visão geral gráfica do material que será coberto na Parte Cinco. O Capítulo 17 será dedicado à *regressão por mínimos quadrados*. Você aprenderá inicialmente como ajustar uma reta a um conjunto de dados que apresentem incertezas. Essa técnica é chamada de *regressão linear*. Além de se discutir como calcular a inclinação e a intersecção dessa reta com o eixo y, serão apresentados também métodos quantitativos e visuais para avaliar a validade dos resultados.

Além de ajustar uma reta, será também apresentada uma técnica geral para ajustar um "melhor" polinômio. Assim, você vai aprender a determinar um polinômio de segundo, terceiro ou de graus mais altos que ajuste de forma ótima dados incertos. A regressão linear é um subconjunto dessa abordagem mais geral, que é chamada *regressão polinomial*.

O próximo tópico coberto no Capítulo 17 é a *regressão linear múltipla*. Ela se aplica aos casos em que a variável independente y é uma função linear de duas ou mais variáveis independentes $x_1, x_2, \ldots, x_m$. Essa abordagem tem utilidade especial no tratamento de dados experimentais quando a variável de interesse depende de diversos fatores diferentes.

Depois da regressão múltipla, será ilustrado como a regressão polinomial e múltipla são ambas subconjuntos de um *modelo linear geral de mínimos quadrados*. Entre outras coisas, isso vai permitir introduzir uma representação matricial concisa da regressão e discutir suas propriedades estatísticas gerais.

Finalmente, as últimas seções do Capítulo 17 são dedicadas à *regressão não-linear*. Essa abordagem se destina ao cálculo de ajuste por mínimos quadrados de uma equação não-linear aos dados.

No Capítulo 18, será descrita uma técnica alternativa de ajuste de curvas, chamada interpolação. Como discutido anteriormente, a interpolação é usada para fazer estimativas de valores intermediários entre pontos dados precisamente. No Capítulo 18, serão determinados polinômios com esse propósito. Serão introduzidos os conceitos básicos de interpolação polinomial usando retas e parábolas para ligar pontos. A seguir, será deduzido um procedimento geral para ajustar um polinômio de grau n. Serão apresentadas duas formas para expressar esses polinômios na forma de uma equação. A primeira, chamada *polinômio interpolador de Newton*, é preferível quando o grau apropriado do polinômio for desconhecido. A segunda, chamada de *polinômio interpolador de Lagrange*, é preferível quando o grau adequado for conhecido antecipadamente.

A última seção do Capítulo 18 apresenta uma técnica alternativa para o ajuste de dados precisos. Essa técnica, chamada *interpolação por splines*, ajusta polinômios aos dados, mas por partes. Dessa forma, ela é particularmente adequada para o ajuste de dados que sejam lisos no geral, mas que apresentem mudanças locais abruptas.

O Capítulo 19 trata da abordagem do ajuste de curvas por transformadas de Fourier, na qual funções periódicas são ajustadas aos dados. A ênfase nessa seção será na *transformada rápida de Fourier*. No final desse capítulo, também incluiremos uma visão geral de diversos pacotes e bibliotecas de software que podem ser usados no ajuste de curvas. Eles são o Excel, o software MATLAB e o IMSL.

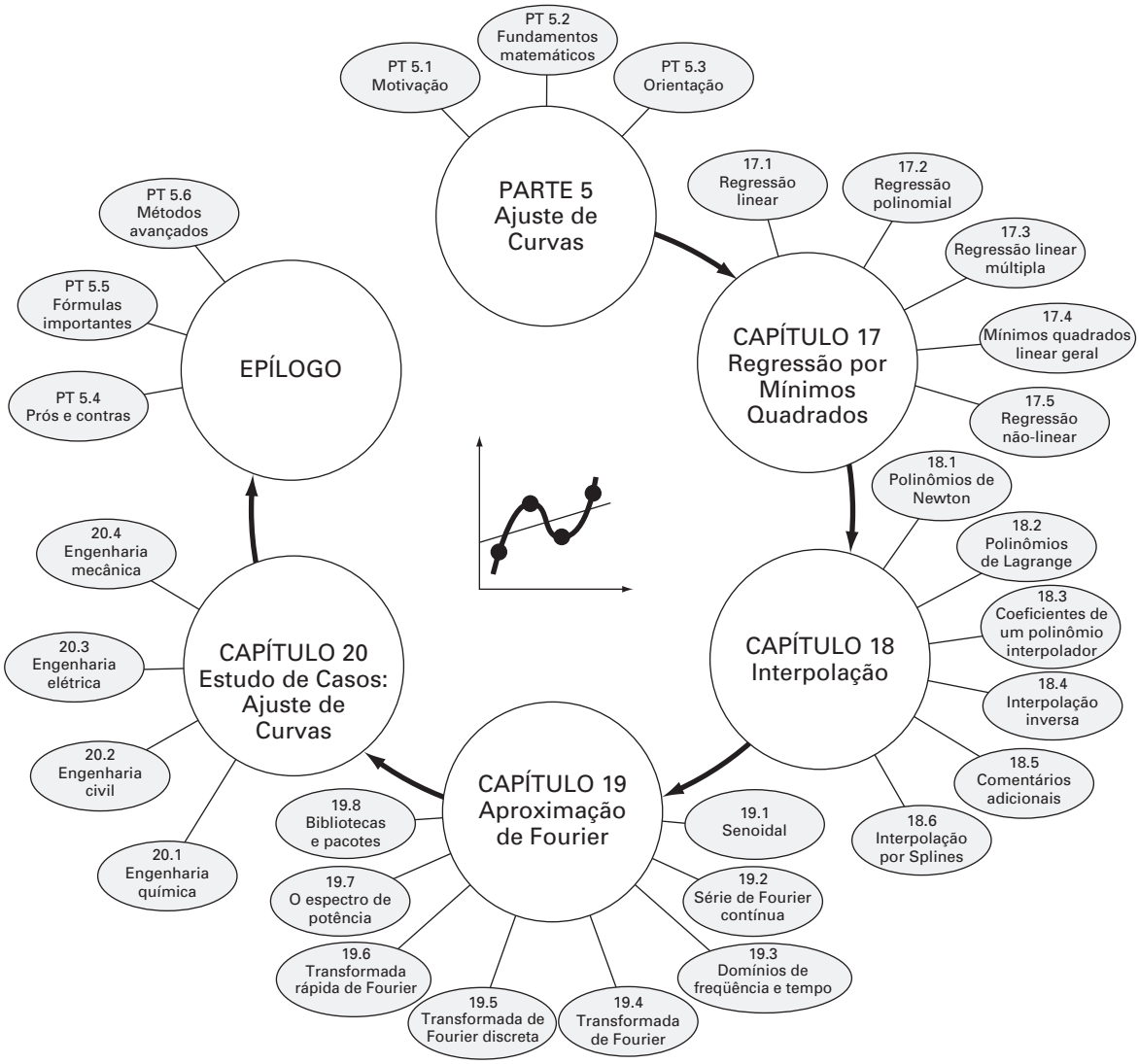


FIGURA PT5.6
Esquema da organização do material da Parte Cinco: Ajuste de Curvas.

O Capítulo 20 é dedicado a aplicações em engenharia que ilustrem a utilidade dos métodos numéricos nos contextos de problemas de engenharia. Os exemplos foram tirados das quatro áreas de especialidade principais de engenharia: química, civil, elétrica e mecânica. Além disso, algumas das aplicações ilustram como os pacotes de software podem ser usados para resolver problemas de engenharia.

Finalmente, será incluído um epílogo no final da Parte Cinco. Ele contém um resumo das fórmulas e dos conceitos importantes relacionados ao ajuste de curvas, bem como uma discussão dos prós e contras das técnicas e sugestões para estudos futuros.

PT5.3.2 Metas e Objetivos

Objetivos de Estudo. Depois de completar a Parte Cinco, você deve ter aumentado muito sua capacidade de ajustar curvas aos dados. Em geral, você deve ter dominado as técnicas, aprendido a avaliar a confiabilidade de suas respostas e deve ser capaz de escolher o método (ou métodos) preferível para qualquer problema particular. Além dessas metas gerais, os conceitos específicos na Tabela PT5.3 devem ser assimilados e dominados.

Objetivos Computacionais. Foram fornecidos a você algoritmos computacionais simples para implementar as técnicas discutidas na Parte Cinco. É possível que você também tenha acesso a pacotes e bibliotecasde software. Todos têm utilidade como ferramentas de aprendizado.

São fornecidos pseudocódigos para a maioria dos métodos na Parte Cinco. Essa informação vai lhe permitir ampliar sua biblioteca de software para incluir técnicas além da regressão polinomial. Por exemplo, você pode considerar útil, do ponto de vista profissional, ter um software para implementar a regressão linear múltipla, o polinômio interpolador de Newton, a interpolação por splines cúbicos e a transformada rápida de Fourier.

Além disso, uma de suas metas mais importantes deveria ser dominar diversos dos pacotes de software de propósito geral que estão amplamente disponíveis. Em particular, você deveria tornar-se adepto a utilizar essas ferramentas para implementar métodos numéricos para a resolução de problemas de engenharia.

TABELA PT5.3 Objetivos específicos de estudo para a Parte Cinco.

1. Entender a diferença fundamental entre regressão e interpolação e perceber por que confundir as duas pode levar a problemas sérios.
2. Entender a dedução da regressão linear por mínimos quadrados e ser capaz de avaliar a confiabilidade dos ajustes usando técnicas gráficas e quantitativas.
3. Saber como linearizar os dados por uma transformação.
4. Entender as situações nas quais a regressão polinomial, múltipla e não-linear são apropriadas.
5. Ser capaz de reconhecer modelos lineares gerais, entender a formulação matricial geral dos mínimos quadrados linear e saber como calcular intervalos de confiança para os parâmetros.
6. Entender que existe um e somente um polinômio de grau menor ou igual a n e que passa exatamente por $n + 1$ pontos.
7. Saber como deduzir o polinômio interpolador de Newton do primeiro grau.
8. Entender a analogia entre o polinômio de Newton e a expansão em série de Taylor e como isso se relaciona com os erros de truncamento.
9. Reconhecer que as equações de Newton e Lagrange são simplesmente formulações diferentes do mesmo polinômio interpolador e entender suas respectivas vantagens e desvantagens.
10. Perceber que resultados mais acurados são geralmente obtidos se os dados usados nas interpolações forem centrados e próximos ao ponto desconhecido.
11. Perceber que os pontos não precisam ser igualmente espaçados nem estar em alguma ordem particular, tanto para o polinômio de Newton quanto para o de Lagrange.
12. Saber por que as fórmulas de interpolação igualmente espaçadas têm utilidade.
13. Reconhecer os defeitos e riscos ligados com a extrapolação.
14. Entender por que as funções spline têm utilidade para dados com regiões localizadas de mudanças abruptas.
15. Reconhecer como a série de Fourier é usada para ajustar dados por funções periódicas.
16. Entender a diferença entre os domínios de freqüência e de tempo.

CAPÍTULO 17

Regressão por Mínimos Quadrados

Quando um erro substancial estiver associado aos dados, a interpolação polinomial é inapropriada e pode produzir resultados insatisfatórios quando usada para prever valores intermediários. Dados experimentais, em geral, são desse tipo. Por exemplo, a Figura 17.1*a* mostra sete pontos de dados obtidos experimentalmente exibindo variações significativas. A inspeção visual dos dados sugere uma possível relação entre y e x. Isto é, a tendência geral indica que valores mais altos de y estão associados com valores mais altos de x. Agora, se um polinômio interpolador de grau seis for ajustado a esses dados (Figura 17.1*b*), ele irá passar exatamente por todos os pontos. Entretanto, por causa da variabilidade dos dados, a curva vai oscilar muito no intervalo entre os pontos. Em particular, os valores interpolados em $x = 1{,}5$ e $x = 6{,}5$ parecem estar bem além do intervalo sugerido pelos dados.

Uma estratégia mais adequada para tais casos seria determinar uma função aproximadora que ajustasse a forma ou tendência geral dos dados sem necessariamente passar pelos pontos individuais. A Figura 17.1*c* ilustra como uma reta pode ser usada para caracterizar a tendência geral dos dados sem passar por nenhum dos pontos particulares.

Uma forma de determinar a reta na Figura 17.1*c* é inspecionar visualmente os pontos marcados e então esboçar a "melhor" reta pelos pontos. Embora tal abordagem "a olho" pareça atrativa do ponto de vista do bom senso e seja válida para cálculos "informais", ela é deficiente porque é arbitrária. Ou seja, a menos que os pontos definam perfeitamente uma reta (em tal caso, a interpolação seria apropriada), analistas diferentes desenhariam retas diferentes.

Para remover essa subjetividade, deve ser desenvolvido algum critério para estabelecer uma base para o ajuste. Uma forma de fazê-lo é determinar a curva que minimize a discrepância entre os dados e os pontos da curva. Uma técnica para conseguir esse objetivo, chamada *regressão por mínimos quadrados*, será discutida no presente capítulo.

17.1 REGRESSÃO LINEAR

O exemplo mais simples de aproximação por mínimos quadrados é ajustar uma reta a um conjunto de pares de observação: (x_1, y_1), (x_2, y_2), . . . , (x_n, y_n). A expressão matemática do ajuste por uma reta é

$$y = a_0 + a_1 x + e \tag{17.1}$$

onde a_0 e a_1 são coeficientes representando a intersecção com o eixo y e a inclinação, respectivamente, e e é o erro ou resíduo entre o modelo e a observação, o qual pode ser representado, depois de se reorganizar a Equação (17.1), por

$$e = y - a_0 - a_1 x$$

Portanto, o erro ou resíduo é a discrepância entre o valor verdadeiro de y e o valor aproximado, $a_0 + a_1 x$, previsto pela equação linear.

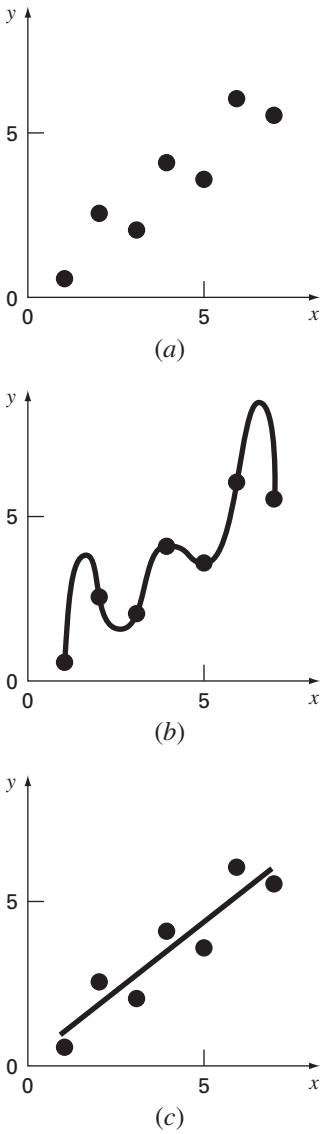


FIGURA 17.1
(*a*) Dados exibindo erros significativos. (*b*) Ajuste polinomial oscilando além do intervalo dos dados. (*c*) Resultado mais satisfatório usando ajuste por mínimos quadrados.

17.1.1 Critério para um "Melhor" Ajuste

Uma estratégia para ajustar uma "melhor" reta pelos dados seria minimizar o valor absoluto da soma dos erros residuais para todos os dados disponíveis, como em

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) \tag{17.2}$$

onde n é o número total de pontos. Entretanto, esse é um critério inadequado, como ilustrado pela Figura 17.2*a*, a qual descreve o ajuste de uma reta a dois pontos. Obviamente, o melhor ajuste é a reta ligando os pontos. Entretanto, qualquer reta passando pelo ponto médio do segmento que liga os pontos (exceto uma reta perfeitamente vertical) resulta em um valor mínimo da Equação (17.2) igual a zero, pois os erros se cancelam.

Portanto, um outro critério lógico poderia ser minimizar a soma dos valores absolutos das discrepâncias, como em

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - a_0 - a_1 x_i|$$

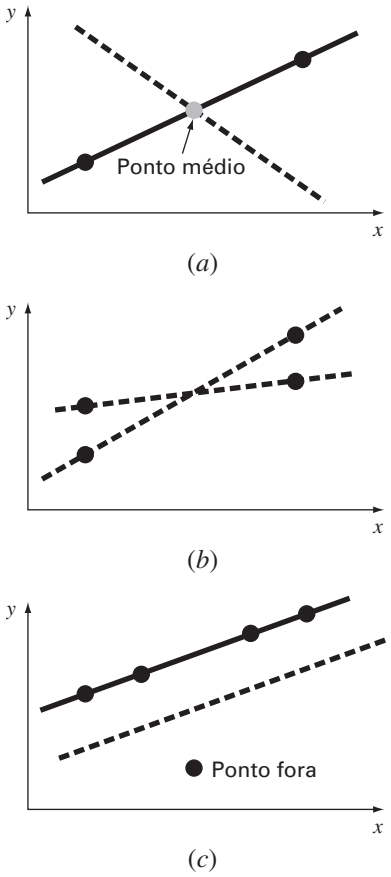


FIGURA 17.2
Exemplos de alguns critérios para "melhor ajuste" que são inadequados para a regressão: (*a*) minimizar a soma dos resíduos, (*b*) minimizar a soma dos valores absolutos dos resíduos, e (*c*) minimizar o erro máximo por qualquer ponto individual.

A Figura 17.2*b* ilustra por que esse critério também é inadequado. Para os quatro pontos mostrados, qualquer reta caindo dentro das retas tracejadas minimizaria a soma dos valores absolutos. Portanto, esse critério também não forneceria um melhor ajuste único.

Uma terceira estratégia para ajustar a melhor reta seria o critério *minimax*. Nessa técnica, é escolhida a reta que minimize a distância máxima que um ponto individual tenha da reta. Como descrito na Figura 17.2*c*, tal estratégia não é adequada para a regressão porque ela permite uma influência indevida a um "ponto discrepante", isto é, a um único ponto com um erro grande. Deve ser observado que o princípio minimax às vezes é adequado para ajustar uma função simples a uma função complicada (Carnahan, Luther e Wilkes, 1969).

Uma estratégia que supera as deficiências das abordagens anteriores é minimizar a soma dos quadrados dos resíduos entre o y medido e o y calculado com o modelo linear

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,\text{medido}} - y_{i,\text{modelo}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \tag{17.3}$$

Esse critério tem diversas vantagens, incluindo o fato de que ele fornece uma única reta para um dado conjunto de dados. Antes de se discutir essas propriedades, será apresentada uma técnica para determinar os valores de a_0 e a_1 que minimizam a Equação (17.3).

17.1.2 Ajuste por Mínimos Quadrados por uma Reta

Para determinar os valores de a_0 e a_1, a Equação (17.3) é derivada com relação a cada coeficiente:

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum [(y_i - a_0 - a_1 x_i)x_i]$$

Observe que simplificamos os símbolos de somatória; a menos que haja menção em contrário, todas as somatórias irão de $i = 1$ a n. Igualando essas derivadas a zero será obtido um S_r mínimo. Se isso for feito, as equações podem ser expressas como

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2$$

Agora, percebendo que $\Sigma a_0 = n a_0$, pode-se expressar essas equações como um conjunto de duas equações lineares simultâneas em duas variáveis (a_0 e a_1):

$$n a_0 + \left(\sum x_i\right) a_1 = \sum y_i \tag{17.4}$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i \tag{17.5}$$

Essas são as chamadas equações normais. Elas podem ser resolvidas simultaneamente

$$a_1 = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n\Sigma x_i^2 - (\Sigma x_i)^2} \tag{17.6}$$

Esse resultado pode, então, ser usado junto com a Equação (17.4) para determinar

$$a_0 = \bar{y} - a_1 \bar{x} \tag{17.7}$$

onde $\bar{y}$ e $\bar{x}$ são as médias de y e x, respectivamente.

EXEMPLO 17.1 Regressão Linear

Enunciado do Problema. Ajuste uma reta aos valores de x e y nas primeiras duas colunas da Tabela 17.1.

Solução. As seguintes quantidades podem ser calculadas:

$$n = 7 \qquad \sum x_i y_i = 119{,}5 \qquad \sum x_i^2 = 140$$

$$\sum x_i = 28 \qquad \bar{x} = \frac{28}{7} = 4$$

$$\sum y_i = 24 \qquad \bar{y} = \frac{24}{7} = 3{,}428571$$

Usando as Equações (17.6) e (17.7),

$$a_1 = \frac{7(119{,}5) - 28(24)}{7(140) - (28)^2} = 0{,}8392857$$

$$a_0 = 3{,}428571 - 0{,}8392857(4) = 0{,}07142857$$

Portanto, o ajuste por mínimos quadrados é

$$y = 0{,}07142857 + 0{,}8392857x$$

TABELA 17.1 Cálculos para uma análise de erro do ajuste linear.

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1x_i)^2$
1	0,5	8,5765	0,1687
2	2,5	0,8622	0,5625
3	2,0	2,0408	0,3473
4	4,0	0,3265	0,3265
5	3,5	0,0051	0,5896
6	6,0	6,6122	0,7972
7	5,5	4,2908	0,1993
Σ	24,0	22,7143	2,9911

A reta, junto com os dados, é mostrada na Figura 17.1*c*.

17.1.3 Quantificação do Erro da Regressão Linear

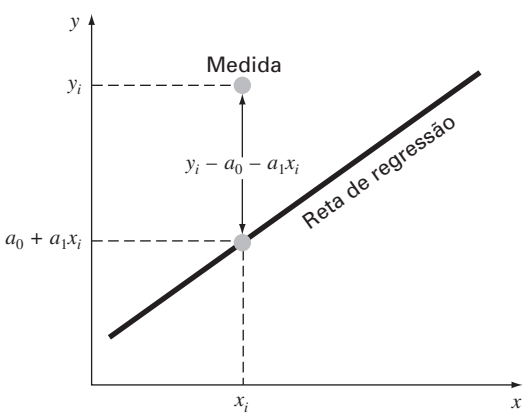
Qualquer outra reta que não a calculada no Exemplo 17.1 resulta em uma soma maior dos quadrados dos resíduos. Portanto, a reta é única e, em termos do critério escolhido por nós, é a "melhor" reta pelos pontos. Diversas propriedades adicionais desse ajuste podem ser elucidadas examinando-se com mais detalhe a forma como os resíduos foram calculados. Lembre-se de que a soma dos quadrados é definida como [Equação (17.3)]

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1x_i)^2 \tag{17.8}$$

Observe a similaridade entre as Equações (PT5.3) e (17.8). No primeiro caso, o quadrado do resíduo representava o quadrado da discrepância entre os dados e uma única estimativa da medida da tendência central — a média. Na Equação (17.8), o quadrado do resíduo representa o quadrado da distância vertical entre os dados e uma outra medida da tendência central — a reta (Figura 17.3).

A analogia pode ser estendida ainda mais nos casos em que (1) a dispersão dos pontos em torno da reta tem valor absoluto parecido ao longo de todo o intervalo dos dados e (2) a distribuição desses pontos em torno da reta é normal. Pode ser demonstrado que, se esses critérios forem satisfeitos, a regressão por mínimos quadrados fornecerá as melhores estimativas (ou seja, as mais prováveis) de a_0 e a_1 (Draper e Smith, 1981). Isso é chamado de *princípio da probabilidade máxima* em estatística. Além disso, se tais

FIGURA 17.3
O resíduo na regressão linear representa a distância vertical entre os pontos dados e a reta.



critérios forem satisfeitos, um "desvio padrão" para a reta de regressão pode ser determinado por [compare com a Equação (PT5.2)]

$$s_{y/x} = \sqrt{\frac{S_r}{n-2}} \tag{17.9}$$

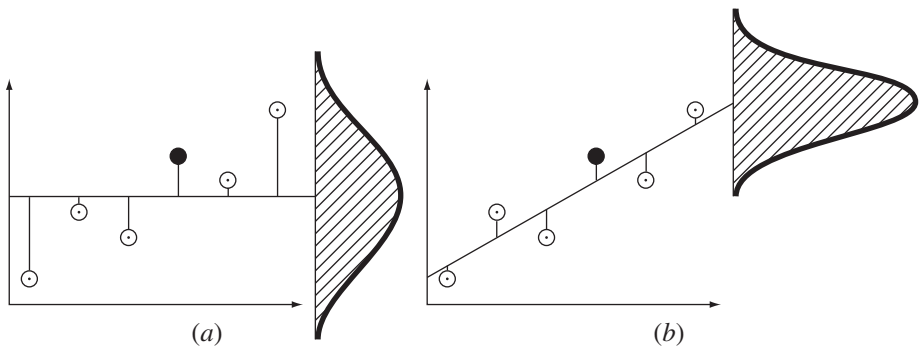
onde $s_{y/x}$ é chamado *erro padrão da estimativa*. O subscrito "y/x" indica que o erro é para um valor previsto de y correspondente a um valor particular de x. Além disso, observe que agora estamos dividindo por $n - 2$ porque duas estimativas provenientes dos dados — a_0 e a_1 — foram usadas para calcular S_r; portanto, perdemos dois graus de liberdade. Do mesmo modo como na discussão de desvio padrão em PT5.2.1, uma outra justificativa para dividir por $n - 2$ é que não existe nenhuma "dispersão de dados" em torno de uma reta ligando dois pontos. Portanto, nos casos nos quais $n = 2$, a Equação (17.9) fornece um resultado infinito, sem sentido.

Exatamente como no caso do desvio padrão, o erro padrão da estimativa quantifica a dispersão dos dados. Entretanto, $s_{y/x}$ quantifica a dispersão *em torno da reta de regressão,* como mostrado na Figura 17.4*b*, em contraste com o desvio padrão original s_y que quantificava a dispersão *em torno da média* (Figura 17.4*a*).

Esses conceitos podem ser usados para quantificar "quão bom" é o ajuste. Isso é particularmente útil para comparar diversas regressões (Figura 17.5). Para fazer isso, voltamos aos dados originais e determinamos *a soma total dos quadrados* em torno da média da variável dependente (no caso, y). Como no caso da Equação (PT5.3), essa quantidade é denotada por S_t, que é o módulo do erro residual associado com a variável dependente antes da regressão. Depois de fazer a regressão, pode-se calcular S_r, a soma dos quadrados dos resíduos em torno da reta de regressão. Isso caracteriza o erro residual que permanece depois da regressão. Portanto, às vezes ele é chamado de soma dos quadrados inexplicável. A diferença entre as duas quantidades, $S_t - S_r$, quantifica a melhora ou a redução de erro decorrente da descrição dos dados em termos de uma reta, em vez de um valor médio. Como o módulo dessa quantidade depende da escala, a diferença é normalizada por S_t para fornecer

$$r^2 = \frac{S_t - S_r}{S_t} \tag{17.10}$$

FIGURA 17.4
Dados de regressão mostrando (*a*) a dispersão dos dados em torno da média da variável dependente e (*b*) a dispersão dos dados em torno da reta de melhor ajuste. A redução na dispersão ao ir de (*a*) para (*b*), como indicada pelas curvas em forma de sino à direita, representa a melhora decorrente da regressão linear.



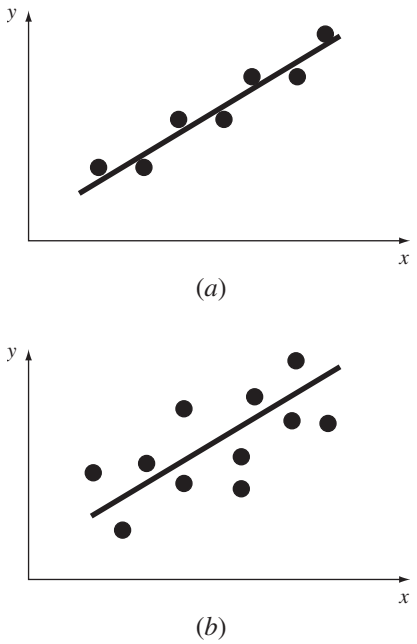


FIGURA 17.5
Exemplos de regressões lineares com erros residuais (*a*) pequenos e (*b*) grandes.

onde r^2 é chamado de *coeficiente de determinação* e r é o *coeficiente de correlação* $(= \sqrt{r^2})$. Para um ajuste perfeito, $S_r = 0$ e $r = r^2 = 1$, significando que a reta explica 100% da variação dos dados. Para $r = r^2 = 0$, $S_r = S_t$ e o ajuste não representa nenhuma melhora. Uma formulação alternativa para r que é mais conveniente para implementação computacional é

$$r = \frac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{\sqrt{n\Sigma x_i^2 - (\Sigma x_i)^2}\sqrt{n\Sigma y_i^2 - (\Sigma y_i)^2}} \tag{17.11}$$

EXEMPLO 17.2

Estimativa de Erros para um Ajuste Linear por Mínimos Quadrados

Enunciado do Problema. Calcule o desvio padrão total, o erro padrão da estimativa e o coeficiente de correlação para os dados do Exemplo 17.1.

Solução. As somatórias são feitas e apresentadas na Tabela 17.1. O desvio padrão é [Equação (PT5.2)]

$$s_y = \sqrt{\frac{22{,}7143}{7-1}} = 1{,}9457$$

e o erro padrão da estimativa é [Equação (17.9)]

$$s_{y/x} = \sqrt{\frac{2{,}9911}{7-2}} = 0{,}7735$$

Portanto, como $s_{y/x} < s_y$, o modelo de regressão linear tem mérito. A extensão da melhora é quantificada por [Equação (17.10)]

$$r^2 = \frac{22{,}7143 - 2{,}9911}{22{,}7143} = 0{,}868$$

ou

$$r = \sqrt{0{,}868} = 0{,}932$$

Esses resultados indicam que 86,8% da incerteza original foi explicada pelo modelo linear.

Antes de prosseguir para o programa computacional para a regressão linear, é preciso fazer um alerta. Embora o coeficiente de correlação forneça uma medida cômoda de quão bom é o ajuste, você deveria tomar cuidado para não associar a ele mais significado do que o devido. Apenas o fato de *r* estar "próximo" de 1 não significa que o ajuste seja necessariamente "bom". Por exemplo, é possível obter um valor relativamente alto de *r* quando a relação subjacente entre *y* e *x* não for nem mesmo linear. Draper e Smith (1981) forneceram diretrizes e material adicional relativos à avaliação dos resultados da regressão linear. Além disso, no mínimo, você deveria sempre inspecionar um gráfico dos dados junto com sua curva de regressão. Como descrito na próxima seção, os pacotes de software incluem tais recursos.

17.1.4 Programa Computacional para Regressão Linear

É relativamente trivial desenvolver um pseudocódigo para a regressão linear (Figura 17.6). Como já mencionado, uma opção para traçar os gráficos é crítica para o uso e a interpretação efetivos da regressão. Tais recursos estão incluídos nos pacotes mais usados, como o software MATLAB e o Excel. Se a sua linguagem de programação tiver recursos gráficos, é recomendável que você expanda seu programa para incluir um gráfico de *y* em função de *x*, mostrando tanto os dados quanto a reta de regressão. A inclusão desse recurso vai aumentar em muito a utilidade do programa nos contextos de resolução de problemas.

FIGURA 17.6
Algoritmo para a regressão linear.

```
SUB Regress(x, y, n, a1, a0, syx, r2)

  somax = 0: somaxy = 0: st = 0
  somay = 0: somax2 = 0: sr = 0
  DOFOR i = 1, n
    somax = somax + xi
    somay = somay + yi
    somaxy = somaxy + xi*yi
    somax2 = somax2 + xi*xi
  END DO
  xm = somax/n
  ym = somay/n
  a1 = (n*somaxy − somax*somay)/(n*somax2 − somax*somax)
  a0 = ym − a1*xm
  DOFOR i = 1, n
    st = st + (yi − ym)^2
    sr = sr + (yi − a1*xi − a0)^2
  END DO
  syx = (sr/(n − 2))^0.5
  r2 = (st − sr)/st

END Regress
```

EXEMPLO 17.3 Regressão Linear Usando o Computador

Enunciado do Problema. Pode-se usar um software baseado na Figura 17.6 para resolver um problema de teste de hipótese associado com o pára-quedista em queda livre discutido no Capítulo 1. Um modelo matemático teórico para a velocidade do pára-quedista foi dado pelo seguinte [Equação (1.10)]

$$v(t) = \frac{gm}{c}\left(1 - e^{(-c/m)t}\right)$$

onde v é a velocidade (m/s), g é a constante gravitacional (9,8 m/s^2), m é a massa do pára-quedista, igual a 68,1 kg, e c é o coeficiente de arrasto, de 12,5 kg/s. O modelo prevê a velocidade do pára-quedista como uma função do tempo, como descrito no Exemplo 1.1.

Um modelo empírico alternativo para a velocidade do pára-quedista é dado por

$$v(t) = \frac{gm}{c}\left(\frac{t}{3{,}75 + t}\right) \tag{E17.3.1}$$

Suponha que você quisesse testar e comparar a adequação desses dois modelos matemáticos. Isso poderia ser conseguido medindo a velocidade real do pára-quedista em valores conhecidos do tempo e comparando tais resultados com as velocidades previstas por cada modelo.

TABELA 17.2 Velocidades medidas e calculadas de um pára-quedista em queda livre.

Tempo, s	v medida m/s (*a*)	v calculada pelo modelo m/s [Equação (1.10)] (*b*)	v calculada pelo modelo m/s [Equação (E17.3.1)] (*c*)
1	10,00	8,953	11,240
2	16,30	16,405	18,570
3	23,00	22,607	23,729
4	27,50	27,769	27,556
5	31,00	32,065	30,509
6	35,60	35,641	32,855
7	39,00	38,617	34,766
8	41,50	41,095	36,351
9	42,90	43,156	37,687
10	45,00	44,872	38,829
11	46,00	46,301	39,816
12	45,50	47,490	40,678
13	46,00	48,479	41,437
14	49,00	49,303	42,110
15	50,00	49,988	42,712

Um programa desse tipo de coleta de dados experimentais foi implementado e os resultados estão listados na coluna (*a*) da Tabela 17.2. As velocidades calculadas pelos modelos estão listadas nas colunas (*b*) e (*c*).

Solução. A adequação do modelo pode ser testada traçando-se a velocidade calculada pelo modelo em função da velocidade medida. A regressão linear pode ser usada para calcular a inclinação e a intersecção com o eixo y do gráfico. Essa reta terá uma inclinação 1, uma intersecção 0 com o eixo y e $r^2 = 1$ se o modelo se adequar perfeitamente aos dados. Um desvio significativo desses valores pode ser usado como uma indicação da inadequação do modelo.

As Figuras 17.7*a* e *b* são gráficos da reta e dos dados para a regressão das colunas (*b*) e (*c*), respectivamente, em função da coluna (*a*). Para o primeiro modelo [Equação (1.10) como descrita na Figura 17.7*a*],

$$v_{\text{modelo}} = -0{,}859 + 1{,}032 v_{\text{medida}}$$

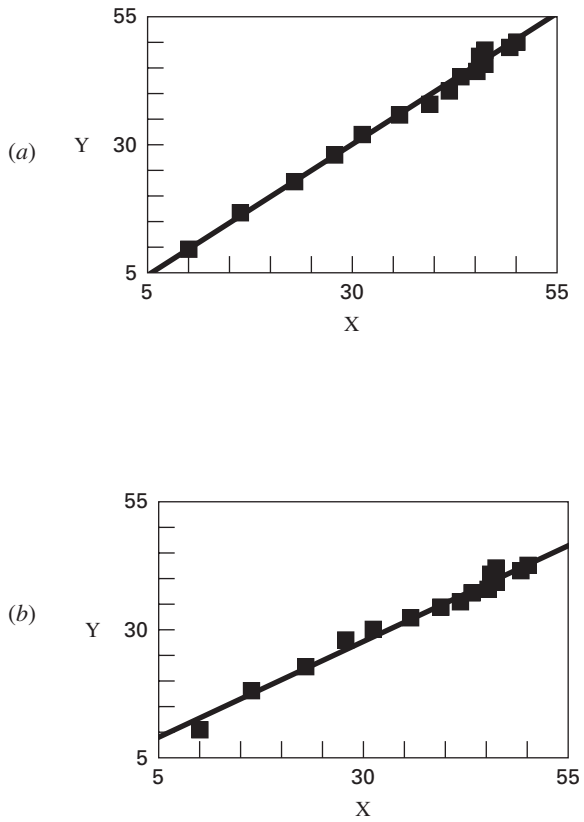


FIGURA 17.7
(*a*) Resultados usando regressão linear para comparar as previsões calculadas pelo modelo teórico [Equação (1.10)] com os valores medidos. (*b*) Resultados usando regressão linear para comparar as previsões calculadas pelo modelo empírico [Equação (E17.3.1)] com os valores medidos.

e para o segundo modelo [Equação (E17.3.1) como descrita na Figura 17.7*b*],

$$v_{\text{modelo}} = 5{,}776 + 0{,}752 v_{\text{medida}}$$

Esses gráficos indicam que a regressão linear entre os dados e cada um dos modelos é altamente significativa. Ambos os modelos se ajustam aos dados com um coeficiente de correlação maior do que 0,99.

Entretanto, o modelo descrito pela Equação (1.10) satisfaz o critério de teste de hipótese muito melhor do que o descrito pela Equação (E17.3.1) porque a inclinação e a intersecção com o eixo *y* estão mais próximas de 1 e 0. Logo, embora cada gráfico seja bem descrito por uma reta, a Equação (1.10) parece ser um modelo melhor do que a Equação (E17.3.1).

Testar e escolher o modelo são atividades comuns e extremamente importantes exercidas em todos os campos da engenharia. O material relativo a fundamentos dados neste capítulo, junto com seu software, deveria permitir a você atacar muitos problemas práticos desse tipo.

Há uma deficiência na análise do Exemplo 17.3. O exemplo não era ambíguo porque o modelo empírico [Equação (E17.3.1)] era claramente inferior à Equação (1.10). Logo,

a inclinação e a intersecção com o eixo y para o primeiro eram tão mais próximas do resultado desejado de 1 e 0 que se tornava óbvio qual modelo era superior.

Entretanto, suponha que a inclinação fosse 0,85 e que a intersecção com o eixo y fosse 2. Obviamente, isso deixaria aberta para debate a conclusão de que a inclinação e a intersecção com o eixo y eram 1 e 0. Claramente, em vez de depender de um julgamento subjetivo, seria preferível basear tal conclusão em um critério quantitativo.

Isso pode ser feito calculando-se intervalos de confiança para os parâmetros do modelo da mesma forma que desenvolvemos intervalos de confiança para a média na Seção PT5.2.3. Esse tópico será retomado no final do presente capítulo.

17.1.5 Linearização de Relações Não-lineares

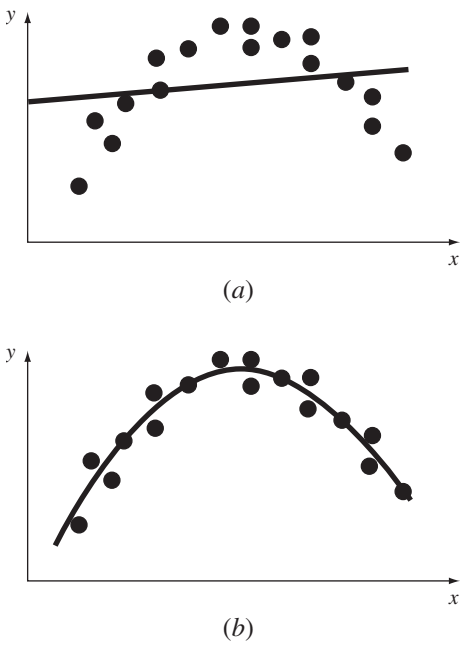
A regressão linear fornece uma técnica poderosa para ajustar a melhor reta aos dados. Entretanto, ela é baseada no fato de que a relação entre as variáveis dependentes e independentes é linear. Esse não é sempre o caso e o primeiro passo em qualquer análise de regressão deveria ser traçar e inspecionar visualmente os dados para verificar se um modelo linear se aplica. Por exemplo, a Figura 17.8 mostra alguns dados que são obviamente curvilíneos. Em alguns casos, técnicas como regressão polinomial, a qual é descrita na Seção 17.2, são apropriadas. Para outros, podem ser usadas transformações para expressar os dados em uma forma que seja compatível com a regressão linear.

Um exemplo é o *modelo exponencial*

$$y = \alpha_1 e^{\beta_1 x} \tag{17.12}$$

onde α_1 e β_1 são constantes. Esse modelo é usado em muitos campos da engenharia para caracterizar quantidades que aumentam (β_1 positivo) ou diminuem (β_1 negativo) a uma taxa que é diretamente proporcional a seu próprio valor absoluto. Por exemplo, o crescimento populacional ou o decaimento radioativo podem exibir tal comportamento. Como

FIGURA 17.8
(*a*) Os dados não são adequados para uma regressão linear por mínimos quadrados. (*b*) Indicação de que uma parábola é preferível.



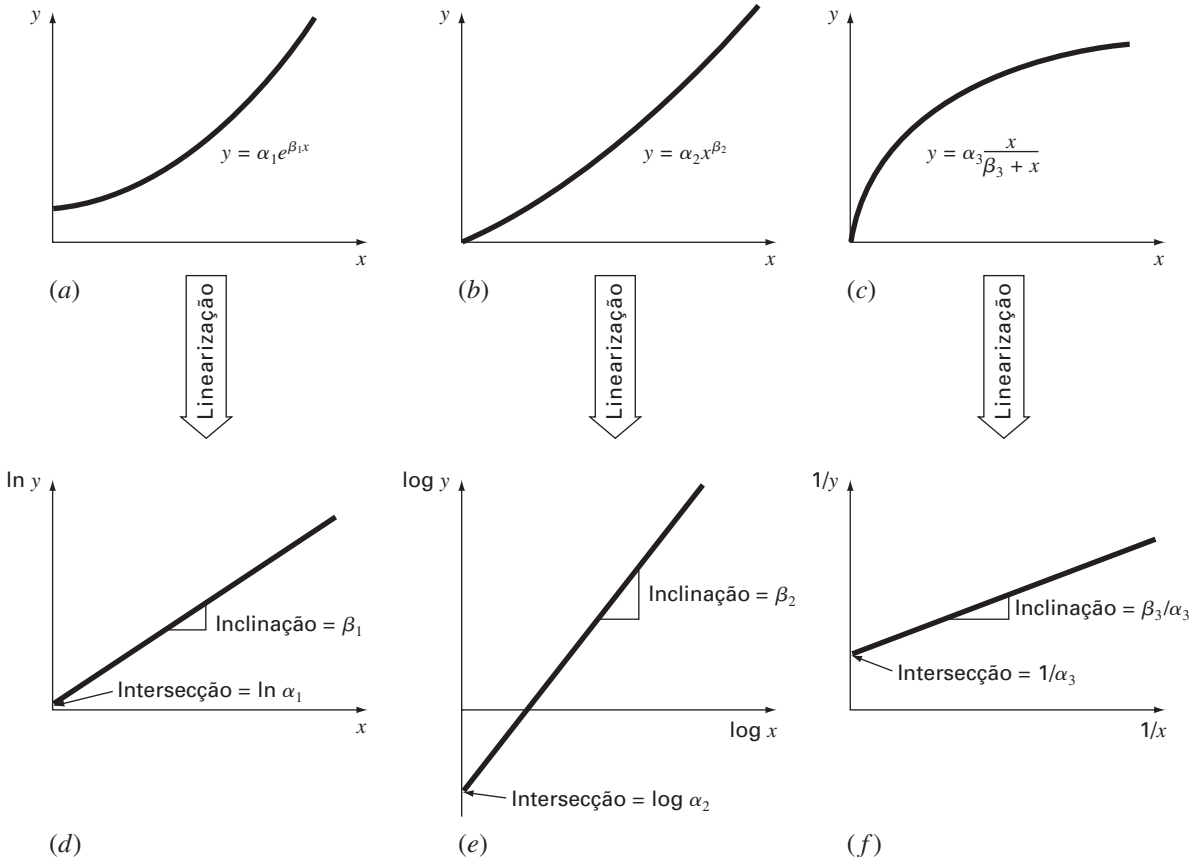


FIGURA 17.9
(*a*) A equação exponencial, (*b*) a equação de potência e (*c*) a equação da taxa de crescimento da saturação. As partes (*d*), (*e*) e (*f*) são versões linearizadas dessas equações que resultam de transformações simples.

descrito na Figura 17.9*a*, a equação representa uma relação não-linear (para $\beta_1 \neq 0$) entre y e x.

Outro exemplo de um modelo não-linear é a equação de potência simples

$$y = \alpha_2 x^{\beta_2} \tag{17.13}$$

onde α_2 e β_2 são coeficientes constantes. Esse modelo tem larga aplicabilidade em todos os campos da engenharia. Como descrito na Figura 17.9*b*, a equação (para $\beta_2 \neq 0$ ou 1) é não-linear.

Um terceiro exemplo de um modelo não-linear é a equação da taxa de crescimento da saturação [lembre-se da Equação (E17.3.1)]

$$y = \alpha_3 \frac{x}{\beta_3 + x} \tag{17.14}$$

onde α_3 e β_3 são coeficientes constantes. Esse modelo, que é particularmente adequado para caracterizar a taxa de crescimento populacional sob condições limitantes, também representa uma relação não-linear entre y e x (Figura 17.9*c*) que se nivela, ou "satura", conforme x aumenta.

As técnicas de regressão não-linear estão disponíveis para ajustar tais equações aos dados experimentais diretamente. (Observe que vamos discutir a regressão não-linear na Seção 17.5.) Entretanto, uma alternativa simples é usar manipulações matemáticas para

transformar as equações para uma forma linear. Então, uma simples regressão linear pode ser usada para ajustar as equações aos dados.

Por exemplo, a Equação (17.12) pode ser linearizada tomando-se seu logaritmo natural para obter

$$\ln y = \ln \alpha_1 + \beta_1 x \ln e$$

Mas, como $\ln e = 1$,

$$\ln y = \ln \alpha_1 + \beta_1 x \tag{17.15}$$

Logo, um gráfico de y em função de x irá fornecer uma reta com uma inclinação β_1 e uma intersecção com o eixo y em $\ln \alpha_1$ (Figura 17.9*d*).

A Equação (17.13) será linearizada tomando-se seu logaritmo na base 10 para obter

$$\log y = \beta_2 \log x + \log \alpha_2 \tag{17.16}$$

Logo, um gráfico de $\log y$ em função de $\log x$ irá fornecer uma reta com uma inclinação β_2 e uma intersecção com o eixo y em $\log \alpha_2$ (Figura 17.9*e*).

A Equação (17.14) será linearizada invertendo-a para obter

$$\frac{1}{y} = \frac{\beta_3}{\alpha_3}\frac{1}{x} + \frac{1}{\alpha_3} \tag{17.17}$$

Logo, um gráfico de $1/y$ em função de $1/x$ será linear, com uma inclinação β_3/α_3 e uma intersecção com o eixo y em $1/\alpha_3$ (Figura 17.9*f*).

Nas suas formas transformadas, esses modelos podem usar a regressão linear para calcular os coeficientes constantes. Eles poderiam, então, ser transformados de volta para seu estado original e usados para propósitos de previsão. O Exemplo 17.4 ilustra esse procedimento para a Equação (17.13). Além disso, a Seção 20.1 fornecerá um exemplo em engenharia do mesmo tipo de cálculo.

EXEMPLO 17.4 Linearização de uma Equação de Potência

Enunciado do Problema. Ajuste a Equação (17.13) aos dados da Tabela 17.3 usando uma transformação logarítmica dos dados.

Solução. A Figura 17.10*a* é um gráfico dos dados originais no seu estado não transformado. A Figura 17.10*b* mostra o gráfico dos dados transformados. Uma regressão linear para os dados transformados pelo log fornece o resultado

$$\log y = 1{,}75 \log x - 0{,}300$$

TABELA 17.3 Dados a serem ajustados pela equação de potência.

x	*y*	log *x*	log *y*
1	0,5	0	−0,301
2	1,7	0,301	0,226
3	3,4	0,477	0,534
4	5,7	0,602	0,753
5	8,4	0,699	0,922

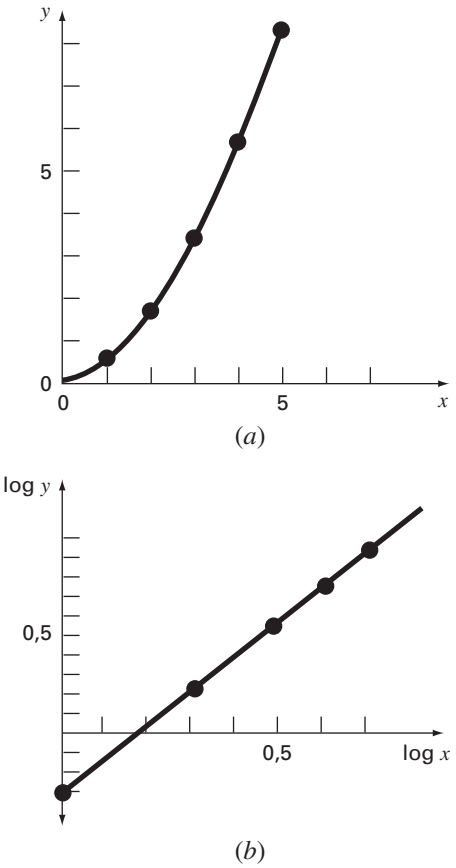


FIGURA 17.10
(*a*) Gráfico dos dados não transformados com a equação de potência que ajusta os dados. (*b*) Gráfico dos dados transformados usados para determinar os coeficientes da equação de potência.

Logo, a intersecção com o eixo y, $\log \alpha_2$, é igual a $-0{,}300$ e, portanto, calculando-se a inversa do logaritmo, $\alpha_2 = 10^{-0{,}3} = 0{,}5$. A inclinação é $\beta_2 = 1{,}75$. Conseqüentemente, a equação de potência é

$$y = 0{,}5x^{1{,}75}$$

Essa curva, traçada na Figura 17.10*a*, indica um bom ajuste.

17.1.6 Comentários Gerais sobre Regressão Linear

Antes de prosseguir para regressão linear múltipla e regressão não-linear, é necessário enfatizar a natureza introdutória do material anterior sobre regressão linear. Nós nos concentramos em deduções simples e usos práticos de equações para ajustar os dados. Você deve estar ciente de que existem aspectos teóricos da regressão que são de importância prática, mas que estão além do escopo deste livro. Por exemplo, algumas hipóteses estatísticas inerentes aos procedimentos por mínimos quadrados são:

1. Cada x tem um valor fixo; ele não é aleatório e é conhecido sem erros.
2. Os valores de y são variáveis aleatórias independentes e têm todos a mesma variância.
3. Os valores de y para um dado x devem estar normalmente distribuídos.

Tais hipóteses são relevantes para a dedução e o uso adequados da regressão. Por exemplo, a primeira hipótese significa que (1) os valores de x devem estar livres de erros e (2) a regressão de y em função de x não é a mesma que a de x em função de y (tente resolver o Problema 17.4 no final do capítulo). Recomendamos fortemente que você consulte outras referências, como Draper e Smith (1981), para apreciar aspectos e nuances da regressão que estão além do escopo deste livro.

17.2 REGRESSÃO POLINOMIAL

Na Seção 17.1, foi desenvolvido um procedimento para deduzir a equação da reta usando o critério dos mínimos quadrados. Alguns dados em engenharia, embora exibam um padrão definido tal como visto na Figura 17.8, são representados de forma insatisfatória por uma reta. Nesses casos, uma curva seria mais adequada para ajustar os dados. Como discutido na seção anterior, um método para atingir esse objetivo é usar transformações. Uma alternativa é ajustar polinômios aos dados usando *regressão polinomial*.

O procedimento dos mínimos quadrados pode ser prontamente estendido para ajustar dados por um polinômio de grau mais alto. Por exemplo, suponha que se queira ajustar um polinômio de segundo grau ou quadrático:

$$y = a_0 + a_1 x + a_2 x^2 + e$$

Nesse caso, a soma dos quadrados dos resíduos é [compare com a Equação (17.3)]

$$S_r = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)^2 \tag{17.18}$$

Seguindo o procedimento da seção anterior, toma-se a derivada da Equação (17.18) com relação a cada um dos coeficientes desconhecidos do polinômio, como em

$$\frac{\partial S_r}{\partial a_0} = -2 \sum \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum x_i \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)$$

$$\frac{\partial S_r}{\partial a_2} = -2 \sum x_i^2 \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)$$

Essas equações podem ser igualadas a zero e reorganizadas para determinar o seguinte conjunto de equações normais:

$$\begin{aligned}
(n)a_0 + \left(\sum x_i\right) a_1 + \left(\sum x_i^2\right) a_2 &= \sum y_i \\
\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 + \left(\sum x_i^3\right) a_2 &= \sum x_i y_i \\
\left(\sum x_i^2\right) a_0 + \left(\sum x_i^3\right) a_1 + \left(\sum x_i^4\right) a_2 &= \sum x_i^2 y_i
\end{aligned} \tag{17.19}$$

onde todas as somatórias vão de $i = 1$ a n. Observe que essas três equações são lineares e têm três incógnitas a_0, a_1 e a_2. Os coeficientes das incógnitas podem ser calculados diretamente dos dados observados.

Nesse caso, vê-se que o problema de determinar o polinômio de segundo grau por mínimos quadrados é equivalente a resolver um sistema de três equações lineares simultâneas. As técnicas para resolver tais equações foram discutidas na Parte Três.

O caso bidimensional pode ser facilmente estendido para um polinômio de grau m como

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m + e$$

A análise anterior pode ser facilmente estendida para esse caso mais geral. Portanto, pode-se perceber que determinar os coeficientes de um polinômio de grau m é equivalente

a resolver um sistema de $m + 1$ equações lineares simultâneas. Nesse caso, o erro padrão é formulado como

$$s_{y/x} = \sqrt{\frac{S_r}{n - (m + 1)}} \tag{17.20}$$

Essa quantidade está dividida por $n - (m + 1)$ porque $(m + 1)$ coeficientes provenientes dos dados — $a_0, a_1, \ldots, a_m$ — foram usados para calcular S_r; logo, foram perdidos $m + 1$ graus de liberdade. Além do erro padrão, um coeficiente de determinação também pode ser calculado para a regressão polinomial com a Equação (17.10).

EXEMPLO 17.5 Regressão Polinomial

Enunciado do Problema. Ajuste um polinômio do segundo grau aos dados nas primeiras duas colunas da Tabela 17.4.

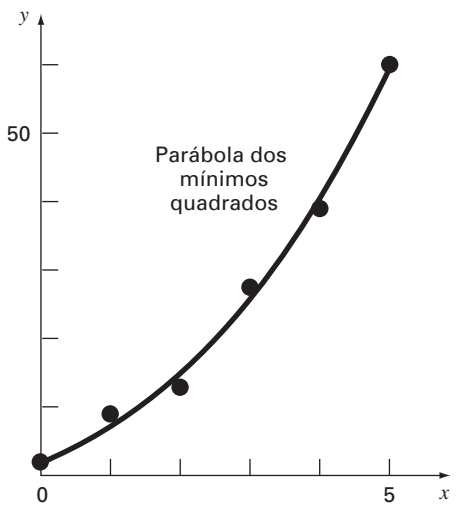
Solução. A partir dos dados fornecidos,

$$m = 2 \qquad \sum x_i = 15 \qquad \sum x_i^4 = 979$$
$$n = 6 \qquad \sum y_i = 152{,}6 \qquad \sum x_i y_i = 585{,}6$$
$$\overline{x} = 2{,}5 \qquad \sum x_i^2 = 55 \qquad \sum x_i^2 y_i = 2488{,}8$$
$$\overline{y} = 25{,}433 \qquad \sum x_i^3 = 225$$

TABELA 17.4 Cálculos para uma análise de erros do ajuste quadrático por mínimos quadrados.

x_i	y_i	$(y_i - \overline{y})^2$	$(y_i - a_0 - a_1 x_i - a_2 x_i^2)$
0	2,1	544,44	0,14332
1	7,7	314,47	1,00286
2	13,6	140,03	1,08158
3	27,2	3,12	0,80491
4	40,9	239,22	0,61951
5	61,1	1.272,11	0,09439
Σ	152,6	2.513,39	3,74657

FIGURA 17.11
Ajuste por um polinômio de segundo grau.



Portanto, as equações lineares simultâneas são

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152{,}6 \\ 585{,}6 \\ 2488{,}8 \end{Bmatrix}$$

Resolvendo essas equações por uma técnica como a eliminação de Gauss, obtêm-se $a_0 = 2{,}47857$, $a_1 = 2{,}35929$ e $a_2 = 1{,}86071$. Portanto, a equação quadrática por mínimos quadrados é

$$y = 2{,}47857 + 2{,}35929x + 1{,}86071x^2$$

O erro padrão da estimativa baseado na regressão polinomial é [Equação (17.20)]

$$s_{y/x} = \sqrt{\frac{3{,}74657}{6-3}} = 1{,}12$$

O coeficiente de determinação é

$$r^2 = \frac{2513{,}39 - 3{,}74657}{2513{,}39} = 0{,}99851$$

e o coeficiente de correlação é $r = 0{,}99925$.

Os resultados indicam que 99,851% da incerteza original foi explicada por esse modelo. Esse resultado confirma a conclusão de que a equação quadrática representa um excelente ajuste, como também é evidente a partir da Figura 17.11.

17.2.1 Algoritmo para a Regressão Polinomial

Um algoritmo para a regressão polinomial está delineado na Figura 17.12. Observe que a tarefa principal é a geração dos coeficientes das equações normais [Equação (17.19)]. (O pseudocódigo para conseguir isso está apresentado na Figura 17.13.) Então, técnicas

Passo 1: Entre a ordem do polinômio a ser ajustado, m.
Passo 2: Entre o número de pontos dados, n.
Passo 3: Se $n < m + 1$, imprima uma mensagem de erro dizendo que a regressão é impossível e pare o processo. Se $n \geq m + 1$, continue.
Passo 4: Calcule os elementos da equação normal, na forma de uma matriz aumentada.
Passo 5: Resolva a matriz aumentada e determine os coeficientes $a_0, a_1, a_2, \ldots, a_m$, usando um método de eliminação.
Passo 6: Imprima os coeficientes.

FIGURA 17.12
Algoritmo para a implementação de regressão polinomial e linear múltipla.

FIGURA 17.13
Pseudocódigo para calcular os elementos da equação normal para a regressão polinomial.

```
DOFOR i = 1, order + 1
  DOFOR j = 1, i
    k = i + j − 2
    soma = 0
    DOFOR ℓ = 1, n
      soma = soma + x_ℓ^k
    END DO
    a_i,j = soma
    a_j,i = soma
  END DO
  soma = 0
  DOFOR ℓ = 1, n
    soma = soma + y_ℓ · x_ℓ^(i−1)
  END DO
  a_i,order+2 = soma
END DO
```

da Parte Três podem ser aplicadas para resolver essas equações simultâneas nos coeficientes do polinômio.

Um problema em potencial associado com a implementação da regressão polinomial no computador é que as equações normais são, às vezes, mal condicionadas. Isso é particularmente verdadeiro para versões de grau mais alto. Nesses casos, os coeficientes calculados podem ser altamente suscetíveis a erros de arredondamento e, conseqüentemente, os resultados podem não ser acurados. Entre outras coisas, o problema está relacionado à estrutura das equações normais e ao fato de que, para polinômios de grau mais alto, as equações normais podem ter coeficientes muito grandes e muito pequenos. Isso ocorre porque os coeficientes são somas de dados elevados a potências.

Embora as técnicas para abrandar os erros de arredondamento discutidas na Parte Três, como o pivotamento, possam ajudar a remediar parcialmente o problema, uma alternativa simples é usar um computador com precisão maior. Felizmente, a maioria dos problemas práticos é limitada a polinômios de grau baixo para os quais os erros de arredondamento são usualmente desprezíveis. Em situações nas quais versões de grau mais alto forem necessárias, existem alternativas disponíveis para certos tipos de dados. Entretanto, essas técnicas (como polinômios ortogonais) estão além do escopo deste livro. O leitor deveria consultar textos sobre regressão, como Draper e Smith (1981), para informação adicional sobre o problema e possíveis alternativas.

17.3 REGRESSÃO LINEAR MÚLTIPLA

Uma extensão útil da regressão linear é o caso no qual y é uma função linear de duas ou mais variáveis independentes. Por exemplo, y poderia ser uma função linear de x_1 e x_2, como em

$$y = a_0 + a_1 x_1 + a_2 x_2 + e$$

Tal equação é particularmente útil para ajustar dados experimentais, para os quais as variáveis sendo estudadas são freqüentemente uma função de duas outras variáveis. Para esse caso bidimensional, a "reta" de regressão se torna o "plano" de regressão (Figura 17.14).

Como nos casos anteriores, os "melhores" valores dos coeficientes são determinados escrevendo-se a soma dos quadrados dos resíduos,

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})^2 \tag{17.21}$$

e derivando-a com relação a cada um dos coeficientes desconhecidos,

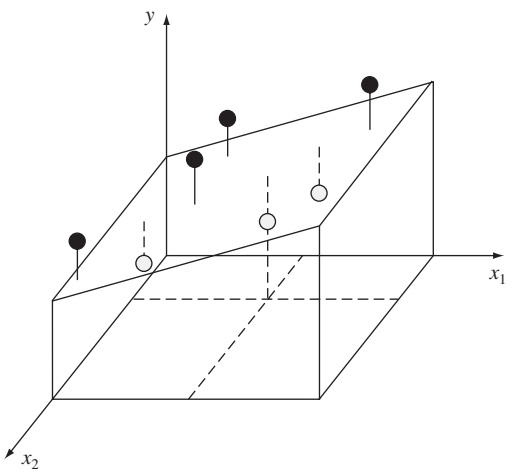


FIGURA 17.14
Descrição gráfica da regressão linear múltipla na qual y é uma função linear de x_1 e x_2.

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})$$
$$\frac{\partial S_r}{\partial a_1} = -2\sum x_{1i}\,(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})$$
$$\frac{\partial S_r}{\partial a_2} = -2\sum x_{2i}\,(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})$$

Os coeficientes fornecendo a soma mínima dos quadrados dos resíduos são obtidos igualando-se as derivadas parciais a zero e expressando o resultado na forma matricial como

$$\begin{bmatrix} n & \Sigma x_{1i} & \Sigma x_{2i} \\ \Sigma x_{1i} & \Sigma x_{1i}^2 & \Sigma x_{1i}x_{2i} \\ \Sigma x_{2i} & \Sigma x_{1i}x_{2i} & \Sigma x_{2i}^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \Sigma y_i \\ \Sigma x_{1i} y_i \\ \Sigma x_{2i} y_i \end{Bmatrix} \tag{17.22}$$

EXEMPLO 17.6 Regressão Linear Múltipla

Enunciado do Problema. Os seguintes dados foram calculados a partir da equação $y = 5 + 4x_1 - 3x_2$:

x_1	x_2	y
0	0	5
2	1	10
2,5	2	9
1	3	0
4	6	3
7	2	27

Use regressão linear múltipla para ajustar esses dados.

Solução. As somatórias necessárias para desenvolver a Equação (17.22) foram calculadas na Tabela 17.5. O resultado é

$$\begin{bmatrix} 6 & 16{,}5 & 14 \\ 16{,}5 & 76{,}25 & 48 \\ 14 & 48 & 54 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 54 \\ 243{,}5 \\ 100 \end{Bmatrix}$$

que pode ser resolvida usando-se um método como a eliminação de Gauss, obtendo-se

$$a_0 = 5 \qquad a_1 = 4 \qquad a_2 = -3$$

o que é consistente com a equação original a partir da qual os dados foram obtidos.

TABELA 17.5 Cálculos necessários para desenvolver as equações normais para o Exemplo 17.6.

	y	x_1	x_2	x_1^2	x_2^2	x_1x_2	x_1y	x_2y
	5	0	0	0	0	0	0	0
	10	2	1	4	1	2	20	10
	9	2,5	2	6,25	4	5	22,5	18
	0	1	3	1	9	3	0	0
	3	4	6	16	36	24	12	18
	27	7	2	49	4	14	189	54
Σ	54	16,5	14	76,25	54	48	243,5	100

O caso bidimensional anterior pode ser facilmente estendido para m dimensões, como em

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m + e$$

para o qual o erro padrão é formulado como

$$s_{y/x} = \sqrt{\frac{S_r}{n - (m + 1)}}$$

e o coeficiente de determinação é calculado como na Equação (17.10). Um algoritmo para determinar as equações normais está listado na Figura 17.15.

Embora possam existir certos casos em que a variável seja linearmente relacionada a duas ou mais outras variáveis, a regressão linear múltipla tem utilidade adicional na dedução das equações de potência na forma geral

$$y = a_0 x_1^{a_1} x_2^{a_2} \cdots x_m^{a_m}$$

Tais equações são extremamente úteis para ajustar dados experimentais. Para usar regressão linear múltipla, a equação é transformada tomando seu logaritmo para obter

$$\log y = \log a_0 + a_1 \log x_1 + a_2 \log x_2 + \cdots + a_m \log x_m$$

Essa transformação é parecida em espírito com a que foi usada na Seção 17.1.5 e no Exemplo 17.4 para ajustar uma equação de potência quando y fosse uma função de uma única variável x. A Seção 20.4 fornece um exemplo de tal aplicação para duas variáveis independentes.

17.4 MÍNIMOS QUADRADOS LINEAR GERAL

Até agora, nós nos concentramos na mecânica da obtenção de ajustes por mínimos quadrados de algumas funções simples aos dados. Antes de nos voltarmos para a regressão não-linear, há diversas questões que gostaríamos de discutir para enriquecer sua compreensão do material anterior.

FIGURA 17.15
Pseudocódigo para determinar os elementos das equações normais para a regressão múltipla. Observe que, além do fato de armazenar as variáveis independentes $x_{1,i}$, $x_{2,i}$, e assim por diante, os 1's precisam ser armazenados em $x_{0,i}$ para que o algoritmo funcione.

```
DOFOR i = 1, order + 1
  DOFOR j = 1, i
    soma = 0
    DOFOR ℓ = 1, n
      soma = soma + x_{i-1,ℓ} · x_{j-1,ℓ}
    END DO
    a_{i,j} = soma
    a_{j,i} = soma
  END DO
  soma = 0
  DOFOR ℓ = 1, n
    soma = soma + y_ℓ · x_{i-1,ℓ}
  END DO
  a_{i,order+2} = soma
END DO
```

17.4.1 Formulação Matricial Geral dos Mínimos Quadrados Linear

Nas páginas precedentes, foram introduzidos três tipos de regressão: linear simples, polinomial e linear múltipla. Na realidade, todas as três pertencem ao seguinte modelo geral de mínimos quadrados linear:

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots + a_m z_m + e \tag{17.23}$$

onde $z_0, z_1, \ldots, z_m$ são $m + 1$ funções base. Pode ser visto facilmente como a regressão linear simples e a linear múltipla se encaixam nesse modelo — isto é, $z_0 = 1$, $z_1 = x_1$, $z_2 = x_2, \ldots, z_m = x_m$. Além disso, a regressão polinomial também é incluída se as funções-base forem monômios simples como em $z_0 = x^0 = 1$, $z_1 = x$, $z_2 = x^2, \ldots, z_m = x^m$.

Observe que a terminologia "linear" se refere apenas a como o modelo depende de seus parâmetros — ou seja, os a's. Como no caso da regressão polinomial, as próprias funções podem ser altamente não-lineares. Por exemplo, os z's podem ser funções senoidais, como em

$$y = a_0 + a_1 \cos(\omega t) + a_2 \operatorname{sen}(\omega t)$$

Tal forma é a base da análise de Fourier descrita no Capítulo 19.

Por outro lado, um modelo aparentemente simples como

$$f(x) = a_0(1 - e^{-a_1 x})$$

é verdadeiramente não-linear porque não pode ser manipulado na forma da Equação (17.23). Tais modelos serão retomados no final deste capítulo.

Por agora, a Equação (17.23) pode ser expressa na notação matricial como

$$\{Y\} = [Z]\{A\} + \{E\} \tag{17.24}$$

onde $[Z]$ é a matriz dos valores calculados das funções base nos valores medidos das variáveis independentes,

$$[Z] = \begin{bmatrix} z_{01} & z_{11} & \cdots & z_{m1} \\ z_{02} & z_{12} & \cdots & z_{m2} \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ z_{0n} & z_{1n} & \cdots & z_{mn} \end{bmatrix}$$

onde m é o número de variáveis no modelo e n é o número de pontos dados. Como $n \geq m + 1$, você deve perceber que, na maioria das vezes, $[Z]$ não é uma matriz quadrada.

O vetor coluna $\{Y\}$ contém os valores observados das variáveis dependentes

$$\{Y\}^T = \lfloor y_1 \quad y_2 \quad \cdots \quad y_n \rfloor$$

O vetor coluna $\{A\}$ contém os coeficientes desconhecidos

$$\{A\}^T = \lfloor a_0 \quad a_1 \quad \cdots \quad a_m \rfloor$$

e o vetor coluna $\{E\}$ contém os resíduos

$$\{E\}^T = \lfloor e_1 \quad e_2 \quad \cdots \quad e_n \rfloor$$

Como foi feito por todo este capítulo, a soma dos quadrados dos resíduos para esse caso pode ser definida por

$$S_r = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{m} a_j z_{ji} \right)^2$$

Essa quantidade pode ser minimizada tomando-se suas derivadas parciais em relação a cada um dos coeficientes e igualando-se a equação resultante a zero. O resultado desse processo são as equações normais que podem ser expressas concisamente na forma matricial como

$$\left[[Z]^T [Z] \right] \{A\} = \left\{ [Z]^T \{Y\} \right\} \tag{17.25}$$

Pode ser mostrado que, de fato, a Equação (17.25) é equivalente às equações normais desenvolvidas anteriormente para a regressão linear simples, polinomial e linear múltipla.

Nossa motivação principal na discussão anterior foi ilustrar a unidade entre as três abordagens e mostrar como elas podem ser todas expressas simplesmente na mesma notação matricial. Ela também prepara o terreno para a próxima seção, na qual será adquirida alguma percepção sobre as estratégias preferidas para resolver a Equação (17.25). A notação matricial também terá relevância quando se abordar a regressão não-linear na última seção deste capítulo.

17.4.2 Técnicas de Solução

Em discussões anteriores deste capítulo, foi abordada rapidamente a questão das técnicas numéricas específicas para resolver as equações normais. Agora que foi estabelecida a unidade entre os diversos modelos, pode-se explorar essa questão com mais detalhe.

Primeiro, deve ficar claro que o método de Gauss-Seidel não pode ser usado porque as equações normais não são diagonalmente dominantes. Restam, portanto, os métodos de eliminação. Para o presente propósito, podem-se dividir essas técnicas em três categorias: (1) métodos de decomposição *LU*, incluindo a eliminação de Gauss, (2) método de Cholesky e (3) abordagens por inversão de matrizes. Existem sobreposições óbvias envolvendo essa divisão. Por exemplo, o método de Cholesky é, na realidade, uma decomposição *LU* e todas as abordagens podem ser formuladas de modo a gerar a matriz inversa. Entretanto, essa divisão tem o mérito de que cada categoria oferece vantagens com relação à solução das equações normais.

Decomposição LU. Se você estiver simplesmente interessado em aplicar o ajuste por mínimos quadrados no caso em que o modelo apropriado é conhecido *a priori*, qualquer uma das abordagens por decomposição *LU* descritas no Capítulo 9 é perfeitamente aceitável. De fato, as formulações que não são do tipo decomposição *LU* da eliminação de Gauss também podem ser usadas. É uma tarefa de programação relativamente simples incorporar qualquer um deles em um algoritmo para mínimos quadrados linear. Realmente, se tiver sido seguida uma abordagem modular, isso é quase trivial.

Método de Cholesky. O algoritmo de decomposição de Cholesky tem diversas vantagens com relação à solução do problema de regressão linear geral. Primeiro, é planejado especificamente para resolver matrizes simétricas como as equações normais. Logo, é rápido e exige menos espaço de armazenamento para resolver tais sistemas. Segundo, ele é apropriado de forma ideal para os casos nos quais a ordem do modelo [ou seja, o valor de m na Equação (17.23)] não for conhecida *a priori* (ver Ralston e Rabinowitz, 1978). Um desses casos seria a regressão polinomial. Nesse caso, poderíamos não saber *a priori* se um polinômio linear, quadrático, cúbico ou de grau superior é o "melhor" modelo para descrever nossos dados. Por causa da forma tanto como as equações normais são construídas quanto como o algoritmo de Cholesky procede (Figura 11.3), pode-se desenvolver modelos de ordem sucessivamente superior de uma maneira extremamente eficiente. Em cada passo, podemos verificar a soma dos quadrados dos resíduos (e um gráfico) para examinar se a inclusão de termos de grau mais alto melhora significativamente o ajuste.

A situação análoga para regressão linear múltipla ocorre quando forem adicionadas variáveis independentes ao modelo, uma de cada vez. Suponha que a variável dependente de interesse seja uma função de diversas variáveis independentes, digamos, temperatura, conteúdo de umidade, pressão etc. Poderíamos primeiro fazer uma regressão linear com a temperatura e calcular um erro residual. A seguir, poderíamos incluir o conteúdo de umidade fazendo uma regressão múltipla com duas variáveis e ver se a variável adicional resultou em um ajuste melhor. O método de Cholesky torna esse processo eficiente porque a decomposição do modelo linear iria simplesmente ser suplementada para incorporar uma nova variável.

Abordagens por Inversão de Matrizes. Da Equação (PT3.6), lembre-se de que a inversa de uma matriz pode ser usada para resolver a Equação (17.25), como em

$$\{A\} = \left[[Z]^T[Z]\right]^{-1}\left\{[Z]^T\{Y\}\right\} \tag{17.26}$$

Cada um dos métodos de eliminação pode ser usado para determinar a inversa e, portanto, pode ser usado para implementar a Equação (17.26). Entretanto, como você aprendeu na Parte Três, essa é uma abordagem ineficiente para resolver um conjunto de equações simultâneas. Logo, se o interesse fosse simplesmente determinar os coeficientes da regressão, seria preferível empregar uma abordagem por decomposição *LU*, sem inversão. Entretanto, do ponto de vista estatístico, há diversas razões pelas quais poderia haver interesse em obter a inversa e examinar seus coeficientes. Essas razões serão discutidas a seguir.

17.4.3 Aspectos Estatísticos da Teoria dos Mínimos Quadrados

Na Seção PT5.2.1, revisamos um pouco da estatística descritiva que pode ser usada para descrever uma amostra. Isso incluía a média aritmética, o desvio padrão e a variância.

Além de fornecer uma solução para os coeficientes da regressão, a formulação matricial da Equação (17.26) fornece estimativas da estatística deles. Pode ser mostrado (Draper e Smith, 1981) que os termos diagonais e fora da diagonal da matriz $[[Z]^T[Z]]^{-1}$ dão, respectivamente, a variância e a covariância[1] dos a's. Se os elementos da diagonal de $[[Z]^T[Z]]^{-1}$ forem designados por $z_{i,i}^{-1}$,

$$\text{var}(a_{i-1}) = z_{i,i}^{-1} s_{y/x}^2 \tag{17.27}$$

e

$$\text{cov}(a_{i-1}, a_{j-1}) = z_{i,j}^{-1} s_{y/x}^2 \tag{17.28}$$

Essas estatísticas têm diversas aplicações importantes. Para o presente propósito, será ilustrado como elas podem ser usadas para desenvolver intervalos de confiança para a intersecção com o eixo y e a inclinação.

Usando uma abordagem parecida com a da Seção PT5.2.3, pode ser mostrado que limites superior e inferior para a intersecção com o eixo y podem ser formulados como (ver Milton e Arnold, 1995, para detalhes)

$$L = a_0 - t_{\alpha/2,n-2}\, s(a_0) \qquad U = a_0 + t_{\alpha/2,n-2}\, s(a_0) \tag{17.29}$$

onde $s(a_j)$ é o erro padrão do coeficiente $a_j = \sqrt{\text{var}(a_j)}$. De forma parecida, os limites superior e inferior da inclinação podem ser formulados como

$$L = a_1 - t_{\alpha/2,n-2}\, s(a_1) \qquad U = a_1 + t_{\alpha/2,n-2}\, s(a_1) \tag{17.30}$$

O seguinte exemplo ilustra como esses intervalos podem ser usados para fazer inferências quantitativas relacionadas à regressão linear.

EXEMPLO 17.7 Intervalos de Confiança para a Regressão Linear

Enunciado do Problema. No Exemplo 17.3, usamos a regressão para desenvolver as seguintes relações entre as medidas e as previsões do modelo:

$$y = -0{,}859 + 1{,}032x$$

onde y são as previsões do modelo e x são as medidas. Concluímos que existe uma boa concordância entre as duas porque a intersecção com o eixo y era aproximadamente igual a zero e a inclinação, aproximadamente igual a 1. Recalcule a regressão, mas use a abor-

[1] A covariância é uma quantidade estatística que mede a dependência de uma variável com outra. Logo, $\text{cov}(x, y)$ indica a dependência de x e y. Por exemplo, $\text{cov}(x, y) = 0$ indicaria que x e y são totalmente independentes.

dagem matricial para estimar os erros padrão para os parâmetros. A seguir, use esses erros para desenvolver intervalos de confiança e utilize-os para fazer uma afirmação probabilística com relação à adequação do ajuste.

Solução. Os dados podem ser escritos na forma matricial para uma regressão linear simples como:

$$[Z] = \begin{bmatrix} 1 & 10 \\ 1 & 16{,}3 \\ 1 & 23 \\ . & . \\ . & . \\ . & . \\ 1 & 50 \end{bmatrix} \qquad \{Y\} = \begin{Bmatrix} 8{,}953 \\ 16{,}405 \\ 22{,}607 \\ . \\ . \\ . \\ 49{,}988 \end{Bmatrix}$$

A transposição e a multiplicação de matrizes podem então ser usadas para gerar as equações normais como:

$$\begin{bmatrix} [Z]^T[Z] \end{bmatrix} \qquad \{A\} = \left\{[Z]^T\{Y\}\right\}$$

$$\begin{bmatrix} 15 & 548{,}3 \\ 548{,}3 & 2.2191{,}21 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} 552{,}741 \\ 2.2421{,}43 \end{Bmatrix}$$

A inversão de matrizes pode ser usada para obter a inclinação e a intersecção com o eixo y como

$$\{A\} = \begin{bmatrix} [Z]^T[Z] \end{bmatrix}^{-1} \qquad \left\{[Z]^T\{Y\}\right\}$$

$$= \begin{bmatrix} 0{,}688414 & -0{,}01701 \\ -0{,}01701 & 0{,}000465 \end{bmatrix} \begin{Bmatrix} 552{,}741 \\ 2.2421{,}43 \end{Bmatrix} = \begin{Bmatrix} -0{,}85872 \\ 1{,}031592 \end{Bmatrix}$$

Logo, a intersecção com o eixo y e a inclinação são determinadas como $a_0 = -0{,}85872$ e $a_1 = 1{,}031592$, respectivamente. Esses valores, por sua vez, podem ser usados pra calcular o erro padrão das estimativas como $s_{y/x} = 0{,}863403$. Esse valor pode ser utilizado junto com os elementos da diagonal da matriz inversa para calcular os erros padrão dos coeficientes,

$$s(a_0) = \sqrt{z_{11}^{-1} s_{y/x}^2} = \sqrt{0{,}688414(0{,}863403)^2} = 0{,}716372$$

$$s(a_1) = \sqrt{z_{22}^{-1} s_{y/x}^2} = \sqrt{0{,}000465(0{,}863403)^2} = 0{,}018625$$

A estatística $t_{\alpha/2,n-1}$ necessária para um intervalo de confiança de 95% com $n - 2 = 15 - 2 = 13$ graus de liberdade pode ser determinada de uma tabela de estatística ou usando um software. Usamos uma função do Excel TINV, para chegar ao valor adequado, como em

$$= \text{TINV}(0{,}05,\ 13)$$

a qual fornece um valor de 2,160368. As Equações (17.29) e (17.30) podem ser usadas para calcular os intervalos de confiança como

$$a_0 = -0{,}85872 \pm 2{,}160368(0{,}716372)$$
$$= -0{,}85872 \pm 1{,}547627 = [-2{,}40634,\ 0{,}688912]$$
$$a_1 = 1{,}031592 \pm 2{,}160368(0{,}018625)$$
$$= 1{,}031592 \pm 0{,}040237 = [0{,}991355,\ 1{,}071828]$$

Observe que os valores desejados (0 para a intersecção e 1 para a inclinação) caem dentro desses intervalos. Com base nessa análise, pode-se fazer a seguinte afirmação estatística com relação à inclinação: temos fortes indicações para acreditar que a inclinação da reta de regressão verdadeira cai dentro do intervalo de 0,991355 a 1,071828. Como

1 cai dentro desse intervalo, também temos fortes indicações para acreditar que o resultado confirma a concordância entre as medidas e o modelo. Como zero cai dentro do intervalo da intersecção, uma afirmação parecida pode ser feita com relação à intersecção com o eixo y.

A discussão anterior é uma introdução limitada ao rico tópico da inferência estatística e sua relação com a regressão. Há muitas sutilezas que estão além do escopo deste livro. Nossa principal motivação foi ilustrar o poder da abordagem matricial para gerar mínimos quadrados linear. Você deveria consultar alguns dos excelentes livros sobre o assunto (por exemplo, Draper e Smith, 1981) para informação adicional. Além disso, deve ser observado que os pacotes e bibliotecas de software podem gerar ajustes por regressão por mínimos quadrados junto com informações relevantes para a inferência estatística. Vamos explorar alguns desses recursos quando descrevermos esses pacotes no final do Capítulo 19.

17.5 REGRESSÃO NÃO-LINEAR

Há muitos casos em engenharia nos quais modelos não-lineares devem ser ajustados aos dados. No presente contexto, esses modelos são definidos como aqueles que têm uma dependência não-linear com seus parâmetros. Por exemplo,

$$f(x) = a_0(1 - e^{-a_1 x}) + e \tag{17.31}$$

Essa equação não pode ser manipulada de modo que fique de acordo com a forma geral da Equação (17.23).

Como nos casos dos mínimos quadrados linear, a regressão não-linear é baseada na determinação dos valores dos parâmetros que minimizem a soma dos quadrados dos resíduos. Entretanto, para o caso não-linear, a solução deve prosseguir de uma forma iterativa.

O *método de Gauss-Newton* é um algoritmo para minimizar a soma dos quadrados dos resíduos entre os dados e as equações não-lineares. Os conceito chave subjacente a essa técnica é que uma expansão em série de Taylor é usada para expressar a equação não-linear original em uma forma aproximada linear. A seguir, a teoria dos mínimos quadrados pode ser usada para se obter novas estimativas dos parâmetros que se movam na direção de minimizar o resíduo.

Para ilustrar como isso é feito, primeiro a relação entre a equação não-linear e os dados pode ser expressa genericamente como

$$y_i = f(x_i; a_0, a_1, \ldots, a_m) + e_i$$

onde y_i é um valor medido da variável dependente, $f(x_i; a_0, a_1, \ldots, a_m)$ é a equação que é uma função da variável independente x_i e uma função não-linear dos parâmetros $a_0, a_1, \ldots, a_m$ e e_i é um erro aleatório. Por conveniência, esse modelo pode ser expresso na forma abreviada pela omissão dos parâmetros,

$$y_i = f(x_i) + e_i \tag{17.32}$$

O modelo não-linear pode ser expandido por uma série de Taylor em torno dos valores dos parâmetros e truncada depois da primeira derivada. Por exemplo, para um caso com 2 parâmetros,

$$f(x_i)_{j+1} = f(x_i)_j + \frac{\partial f(x_i)_j}{\partial a_0}\,\Delta a_0 + \frac{\partial f(x_i)_j}{\partial a_1}\,\Delta a_1 \tag{17.33}$$

onde j é a aproximação inicial, $j + 1$ é a previsão, $\Delta a_0 = a_{0,j+1} - a_{0,j}$ e $\Delta a_1 = a_{1,j+1} - a_{1,j}$. Logo, linearizamos o modelo inicial com relação aos parâmetros. A Equação (17.33) pode ser substituída na Equação (17.32) para fornecer

$$y_i - f(x_i)_j = \frac{\partial f(x_i)_j}{\partial a_0}\,\Delta a_0 + \frac{\partial f(x_i)_j}{\partial a_1}\,\Delta a_1 + e_i$$

ou, na forma matricial [compare com a Equação (17.24)],

$$\{D\} = \left[Z_j\right]\{\Delta A\} + \{E\} \tag{17.34}$$

onde $[Z_j]$ é a matriz das derivadas parciais da função calculada na aproximação inicial j,

$$[Z_j] = \begin{bmatrix} \partial f_1/\partial a_0 & \partial f_1/\partial a_1 \\ \partial f_2/\partial a_0 & \partial f_2/\partial a_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \partial f_n/\partial a_0 & \partial f_n/\partial a_1 \end{bmatrix}$$

onde n é o número de pontos dados e $\partial f_i/\partial a_k$ é a derivada parcial da função com relação ao k-ésimo parâmetro calculada no i-ésimo ponto dado. O vetor $\{D\}$ contém as diferenças entre as medidas e os valores da função,

$$\{D\} = \begin{Bmatrix} y_1 - f(x_1) \\ y_2 - f(x_2) \\ \cdot \\ \cdot \\ \cdot \\ y_n - f(x_n) \end{Bmatrix}$$

e o vetor $\{\Delta A\}$ contém a variação nos valores dos parâmetros,

$$\{\Delta A\} = \begin{Bmatrix} \Delta a_0 \\ \Delta a_1 \\ \cdot \\ \cdot \\ \cdot \\ \Delta a_m \end{Bmatrix}$$

A aplicação da teoria linear dos mínimos quadrados à Equação (17.34) resulta nas seguintes equações normais [lembre-se da Equação (17.25)]:

$$\left[[Z_j]^T[Z_j]\right]\{\Delta A\} = \left\{[Z_j]^T\{D\}\right\} \tag{17.35}$$

Logo, a abordagem consiste em resolver a Equação (17.35), determinando $\{\Delta A\}$, o qual pode ser usado para calcular valores melhorados para os parâmetros como em

$$a_{0,j+1} = a_{0,j} + \Delta a_0$$

e

$$a_{1,j+1} = a_{1,j} + \Delta a_1$$

Esse procedimento é repetido até que a solução convirja — ou seja, até que

$$|\varepsilon_a|_k = \left|\frac{a_{k,j+1} - a_{k,j}}{a_{k,j+1}}\right| 100\% \tag{17.36}$$

fique abaixo de um critério de parada aceitável.

EXEMPLO 17.9 Método de Gauss-Newton

Enunciado do Problema. Ajuste a função $f(x; a_0, a_1) = a_0(1 - e^{-a_1 x})$ aos dados

x	0,25	0,75	1,25	1,75	2,25
y	0,28	0,57	0,68	0,74	0,79

Use aproximações iniciais de $a_0 = 1{,}0$ e $a_1 = 1{,}0$ para os parâmetros. Observe que, para essas aproximações, a soma inicial dos quadrados dos resíduos é 0,0248.

Solução. As derivadas parciais da função com relação aos parâmetros são

$$\frac{\partial f}{\partial a_0} = 1 - e^{-a_1 x} \tag{E17.9.1}$$

e

$$\frac{\partial f}{\partial a_1} = a_0 x e^{-a_1 x} \tag{E17.9.2}$$

As Equações (E17.9.1) e (E17.9.2) podem ser usadas para calcular a matriz

$$[Z_0] = \begin{bmatrix} 0{,}2212 & 0{,}1947 \\ 0{,}5276 & 0{,}3543 \\ 0{,}7135 & 0{,}3581 \\ 0{,}8262 & 0{,}3041 \\ 0{,}8946 & 0{,}2371 \end{bmatrix}$$

Essa matriz multiplicada pela sua transposta resulta em

$$[Z_0]^T[Z_0] = \begin{bmatrix} 2{,}3193 & 0{,}9489 \\ 0{,}9489 & 0{,}4404 \end{bmatrix}$$

a qual, por sua vez, pode ser invertida para fornecer

$$\left[[Z_0]^T[Z_0]\right]^{-1} = \begin{bmatrix} 3{,}6397 & -7{,}8421 \\ -7{,}8421 & 19{,}1678 \end{bmatrix}$$

O vetor $\{D\}$ consiste nas diferenças entre as medidas e as previsões do modelo,

$$\{D\} = \begin{Bmatrix} 0{,}28 - 0{,}2212 \\ 0{,}57 - 0{,}5276 \\ 0{,}68 - 0{,}7135 \\ 0{,}74 - 0{,}8262 \\ 0{,}79 - 0{,}8946 \end{Bmatrix} = \begin{Bmatrix} 0{,}0588 \\ 0{,}0424 \\ -0{,}0335 \\ -0{,}0862 \\ -0{,}1046 \end{Bmatrix}$$

Isso é multiplicado por $[Z_0]^T$ para fornecer

$$[Z_0]^T\{D\} = \begin{bmatrix} -0{,}1533 \\ -0{,}0365 \end{bmatrix}$$

O vetor $\{\Delta A\}$ é então calculado resolvendo-se a Equação (17.35),

$$\Delta A = \begin{Bmatrix} -0{,}2714 \\ 0{,}5019 \end{Bmatrix}$$

o que pode ser somado às aproximações iniciais dos parâmetros para fornecer

$$\begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} 1{,}0 \\ 1{,}0 \end{Bmatrix} + \begin{Bmatrix} -0{,}2714 \\ 0{,}5019 \end{Bmatrix} = \begin{Bmatrix} 0{,}7286 \\ 1{,}5019 \end{Bmatrix}$$

Logo, as estimativas melhoradas dos parâmetros são $a_0 = 0{,}7286$ e $a_1 = 1{,}5019$. Os novos parâmetros resultam em uma soma dos quadrados dos resíduos igual a 0,0242. A Equação (17.36) pode ser usada para calcular ε_0 e ε_1 iguais a 37% e 33%, respectivamente. O cálculo seria então repetido até que esses valores ficassem abaixo do critério de parada prescrito. O resultado final é $a_0 = 0{,}79186$ e $a_1 = 1{,}6751$. Esses coeficientes dão uma soma dos quadrados dos resíduos de 0,000662.

Um problema em potencial com o método de Gauss-Newton como desenvolvido até esse ponto é que as derivadas parciais da função podem ser difíceis de calcular. Conseqüentemente, muitos programas de computador usam equações de diferenças para aproximar as derivadas parciais. Um método é

$$\frac{\partial f_i}{\partial a_k} \cong \frac{f(x_i; a_0, \ldots, a_k + \delta a_k, \ldots, a_m) - f(x_i; a_0, \ldots, a_k, \ldots, a_m)}{\delta a_k} \tag{17.37}$$

onde δ é uma pequena fração de perturbação.

O método de Gauss-Newton apresenta diversas outras possíveis deficiências:

1. Ele pode convergir lentamente.
2. Ele pode oscilar muito, ou seja, mudar de direção continuamente.
3. Ele pode simplesmente não convergir.

Modificações do método (Booth e Peterson, 1958; Hartley, 1961) têm sido desenvolvidas para remediar as deficiências.

Além disso, embora existam diversas abordagens especialmente planejadas para a regressão, uma abordagem mais geral é usar rotinas de otimização não-lineares como descrito na Parte Quatro. Para fazê-lo, realiza-se uma aproximação para os parâmetros e a soma dos quadrados dos resíduos é calculada. Por exemplo, para a Equação (17.31), isso seria calculado como

$$S_r = \sum_{i=1}^{n} [y_i - a_0(1 - e^{-a_1 x_i})]^2 \tag{17.38}$$

Então, os parâmetros seriam ajustados sistematicamente para minimizar S_r usando-se técnicas de busca do tipo descrito anteriormente no Capítulo 14. Vamos ilustrar como isso é feito quando descrevermos aplicações de software no final do Capítulo 19.

PROBLEMAS

17.1 Considere os dados

8,8	9,5	9,8	9,4	10,0
9,4	10,1	9,2	11,3	9,4
10,0	10,4	7,9	10,4	9,8
9,8	9,5	8,9	8,8	10,6
10,1	9,5	9,6	10,2	8,9

Determine (**a**) a média, (**b**) o desvio padrão, (**c**) a variância, (**d**) o coeficiente de variação e (**e**) o intervalo de confiança de 95% para a média.

17.2 Construa um histograma a partir dos dados do Problema 17.1. Use uma variação de 7,5 a 11,5 com intervalos de 0,5.

17.3 Considere os dados

28,65	26,55	26,65	27,65	27,35	28,35	26,85
28,65	29,65	27,85	27,05	28,25	28,85	26,75
27,65	28,45	28,65	28,45	31,65	26,35	27,75
29,25	27,65	28,65	27,65	28,55	27,65	27,25

Determine (**a**) a média, (**b**) o desvio padrão, (**c**) a variância, (**d**) o coeficiente de variação e (**e**) o intervalo de confiança de 90% para a média. (**f**) Construa um histograma. Use um intervalo de 26 a 32 com incrementos de 0,5. (**g**) Supondo que a distribuição seja normal e que sua estimativa do desvio padrão seja válida, calcule o intervalo (isto é, as extremidades inferior e superior) que englobe 68% das leituras. Determine se essa é uma estimativa válida para os dados desse problema.

17.4 Use regressão por mínimos quadrados para ajustar uma reta a

x	0	2	4	6	9	11	12	15	17	19
y	5	6	7	6	9	8	7	10	12	12

Junto com a inclinação e a intersecção com o eixo y, calcule o erro padrão da estimativa e o coeficiente de correlação. Faça um gráfico dos dados e da reta de regressão. A seguir, repita o problema, mas faça a regressão de x em função de y — isto é, troque as variáveis. Interprete seus resultados.

17.5 Use regressão por mínimos quadrados para ajustar uma reta a

x	6	7	11	15	17	21	23	29	29	37	39
y	29	21	29	14	21	15	7	7	13	0	3

Junto com a inclinação e a intersecção com o eixo y, calcule o erro padrão da estimativa e o coeficiente de correlação. Faça um gráfico dos dados e da reta de regressão. Se alguém fizesse uma medida adicional de $x = 10$, $y = 10$, você suspeitaria, com base em uma avaliação visual e no erro padrão, que a medida é válida ou não? Justifique sua conclusão.

17.6 Usando a mesma abordagem que foi empregada para deduzir as Equações (17.15) e (17.16), deduza um ajuste por mínimos quadrados para o seguinte modelo:

$$y = a_1 x + e$$

Isto é, determine a inclinação que resulta no ajuste por mínimos quadrados para uma reta com intersecção nula com o eixo y. Ajuste os seguintes dados com esse modelo e mostre o resultado graficamente:

x	2	4	6	7	10	11	14	17	20
y	1	2	5	2	8	7	6	9	12

17.7 Use regressão por mínimos quadrados para ajustar uma reta a

x	1	2	3	4	5	6	7	8	9
y	1	1,5	2	3	4	5	8	10	13

(a) Junto com a inclinação e a intersecção com o eixo y, calcule o erro padrão da estimativa e o coeficiente de correlação. Faça o gráfico dos dados e da reta. Avalie o ajuste.
(b) Recalcule (**a**), mas use uma regressão polinomial para ajustar uma parábola aos dados. Compare os resultados com aqueles em (**a**).

17.8 Ajuste os seguintes dados com (**a**) um modelo de taxa de crescimento de saturação, (**b**) uma equação de potência e (**c**) uma parábola. Em cada caso, trace os dados e a equação.

x	0,75	2	3	4	6	8	8,5
y	1,2	1,95	2	2,4	2,4	2,7	2,6

17.9 Ajuste os seguintes dados com um modelo de potências ($y = ax^b$). Use a equação de potência resultante para prever y em $x = 9$:

x	2,5	3,5	5	6	7,5	10	12,5	15	17,5	20
y	13	11	8,5	8,2	7	6,2	5,2	4,8	4,6	4,3

17.10 Ajuste um modelo exponencial a

x	0,4	0,8	1,2	1,6	2	2,3
y	800	975	1.500	1.950	2.900	3.600

Trace os dados e a equação tanto em papel de gráfico padrão quanto logarítmico.

17.11 Em vez de usar um modelo exponencial na base e [Equação (17.22)], uma alternativa comum é usar um modelo na base 10,

$$y = \alpha_5 10^{\beta_5 x}$$

Quando usada para ajuste de curvas, essa equação fornece resultados idênticos àqueles na base e, mas o valor do parâmetro no expoente (β_5) será diferente do valor estimado com a Equação (17.22) (β_1). Use a versão na base 10 para resolver o Problema 17.10. Além disso, deduza uma fórmula para relacionar β_1 a β_5.

17.12 Além dos exemplos na Figura 17.10, existem outros mo-delos que podem ser linearizados usando-se transformações. Por exemplo,

$$y = \alpha_4 x e^{\beta_4 x}$$

Linearize esse modelo e use-o para fazer uma estimativa de α_4 e β_4 com base nos seguintes dados.

x	0,1	0,2	0,4	0,6	0,9	1,3	1,5	1,7	1,8
y	0,75	1,25	1,45	1,25	0,85	0,55	0,35	0,28	0,18

17.13 Um investigador relatou os dados tabulados a seguir para uma experiência para determinar a taxa de crescimento da bactéria k (por dia), como uma função da concentração de oxigênio c (mg/L). Sabe-se que tais dados podem ser modelados pela seguinte equação:

$$k = \frac{k_{máx} c^2}{c_s + c^2}$$

onde c_s e k_{max} são parâmetros. Use uma transformação para linearizar essa equação. A seguir, use regressão linear para fazer uma estimativa de c_s e k_{max} e prever a taxa de crescimento em $c = 2$ mg/L.

c	0,5	0,8	1,5	2,5	4
k	1,1	2,4	5,3	7,6	8,9

17.14 Considere os dados

x	5	10	15	20	25	30	35	40	45	50
y	17	24	31	33	37	37	40	40	42	41

e use regressão por mínimos quadrados para ajustar (**a**) uma reta, (**b**) uma equação de potência, (**c**) uma equação de taxa de crescimento de saturação e (**d**) uma parábola. Faça o gráfico dos dados junto com todas as curvas. Uma das curvas é superior?
Se esse for o caso, justifique.

17.15 Ajuste uma equação cúbica aos seguintes dados:

x	3	4	5	7	8	9	11	12
y	1,6	3,6	4,4	3,4	2,2	2,8	3,8	4,6

Junto com os coeficientes, determine r^2 e $s_{y/x}$.

17.16 Use regressão linear múltipla para ajustar

x_1	0	1	1	2	2	3	3	4	4
x_2	0	1	2	1	2	1	2	1	2
y	15,1	17,9	12,7	25,6	20,5	35,1	29,7	45,4	40,2

Calcule os coeficientes, o erro padrão da estimativa e o coeficiente de correlação.

17.17 Use regressão linear múltipla para ajustar

x_1	0	0	1	2	0	1	2	2	1
x_2	0	2	2	4	4	6	6	2	1
y	14	21	11	12	23	23	14	6	11

Calcule os coeficientes, o erro padrão da estimativa e o coeficiente de correlação.

17.18 Use regressão não-linear para ajustar uma parábola aos seguintes dados:

x	0,2	0,5	0,8	1,2	1,7	2	2,3
y	500	700	1.000	1.200	2.200	2.650	3.750

17.19 Use regressão não-linear para ajustar uma equação de taxa de crescimento de saturação aos dados do Problema 17.14.

17.20 Recalcule o ajuste por regressão dos Problemas (**a**) 17.4 e (**b**) 17.15, usando a abordagem matricial. Faça uma estimativa dos erros padrão e desenvolva um intervalo de confiança de 90% para os coeficientes.

17.21 Desenvolva, debug e teste um programa em uma linguagem de alto nível ou em uma linguagem de macro de sua escolha para implementar regressão linear. Entre outras coisas: (**a**) inclua comentários para documentar o código e (**b**) determine o erro padrão e o coeficiente de determinação.

17.22 Um material é testado para falha por fadiga cíclica enquanto uma tensão, em MPa, é aplicada ao material e o número de ciclos necessários para causar falha é medido. Os resultados estão na tabela a seguir. Quando um gráfico log-log da tensão em função dos ciclos é gerado, a tendência dos dados mostra uma relação linear. Use regressão por mínimos quadrados para determinar a equação de melhor ajuste para estes dados.

Nº de ciclos	1	10	100	1.000	10.000	100.000	1.000.000
Tensão, MPa	1.100	1.000	925	800	625	550	420

17.23 Os seguintes dados mostram a relação entre a viscosidade do óleo SAE 70 e a temperatura. Depois de tomar o log dos dados, use uma regressão linear para encontrar uma equação da reta que melhor ajusta os dados e o valor de r^2.

Temperatura, °C	26,67	93,33	148,89	315,56
Viscosidade, μ, N · s/m^2	1,35	0,085	0,012	0,00075

17.24 Os dados a seguir representam o crescimento bacteriológico em uma cultura líquida durante certo número de dias.

Dia	0	4	8	12	16	20
Quantidade × 10^6	67	84	98	125	149	185

Encontre a equação do melhor ajuste para a tendência dos dados. Tente várias possibilidades — linear, parabólica e exponencial. Use um pacote de software de sua escolha para encontrar a melhor equação para prever a quantidade de bactéria após 40 dias.

17.25 A concentração da bactéria *E. coli* em uma área de natação é monitorada depois de uma tempestade:

t (h)	4	8	12	16	20	24
c (UFC/100 mL)	1590	1320	1000	900	650	560

O tempo é medido em horas depois do final da tempestade e a unidade UFC é uma "unidade de formação de colônias". Use esses dados para fazer uma estimativa (**a**) da concentração no final da tempestade ($t = 0$) e (**b**) do instante no qual a concentração atingirá 200 UFC/100 mL. Observe que sua escolha do modelo deveria ser consistente com o fato de que uma concentração negativa é impossível e que a concentração de bactérias sempre decresce com o tempo.

17.26 Um objeto é suspenso em um túnel de vento e a força é medida para diversos níveis de velocidade do vento. Os resultados estão tabulados a seguir.

v, m/s	10	20	30	40	50	60	70	80
F, N	25	70	380	550	610	1.220	830	1.450

Use regressão por mínimos quadrados para ajustar esses dados com (**a**) uma reta, (**b**) uma equação de potência baseada em transformações log e (**c**) um modelo de potências baseado em regressão não-linear. Mostre os seus resultados graficamente.

17.27 Ajuste um modelo de potências aos dados do Problema 17.26, mas use logaritmos naturais para fazer as transformações.

17.28 Usando a mesma abordagem que foi utilizada para deduzir as Equações (17.15) e (17.16), deduza um ajuste por mínimos quadrados do seguinte modelo:

$$y = a_1 x + a_2 x^2 + e$$

Isto é, determine os coeficientes que resultam no ajuste por mínimos quadrados para um polinômio de 2º grau com intersecção com o eixo y igual a zero. Teste essa abordagem usando-a para ajustar os dados do Problema 17.26.

17.29 No Problema 17.12, utilizamos transformações para ajustar e linearizar o seguinte modelo:

$$y = \alpha_4 x e^{\beta_4 x}$$

Use regressão não-linear para fazer uma estimativa de α_4 e β_4 com base nos seguintes dados. Faça um gráfico de seu ajuste junto com os dados.

x	0,1	0,2	0,4	0,6	0,9	1,3	1,5	1,7	1,8
y	0,75	1,25	1,45	1,25	0,85	0,55	0,35	0,28	0,18

CAPÍTULO 18

Interpolação

Você freqüentemente terá a oportunidade de fazer estimativas de valores intermediários entre dados precisos. O método mais comum usado para esse propósito é a interpolação polinomial. Lembre-se de que a fórmula geral para um polinômio de grau n é

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{18.1}$$

Para $n + 1$ pontos dados, existe um e somente um polinômio de grau n[1] que passa por todos os pontos. Por exemplo, existe uma única reta (isto é, um polinômio de primeiro grau) que liga dois pontos (Figura 18.1*a*). Analogamente, existe uma única parábola ligando um conjunto de três pontos (Figura 18.1*b*). A *interpolação polinomial* consiste em determinar o único polinômio de grau n que passa pelos $n + 1$ pontos dados. Esse polinômio, então, fornece uma fórmula para calcular valores intermediários.

Embora exista um e só um polinômio de grau n que passa por $n + 1$ pontos, há diversas fórmulas matemáticas nas quais esse polinômio pode ser expresso. Neste capítulo, serão descritas duas alternativas que são adequadas para a implementação computacional: polinômios de Newton e de Lagrange.

18.1 POLINÔMIOS INTERPOLADORES POR DIFERENÇAS DIVIDIDAS DE NEWTON

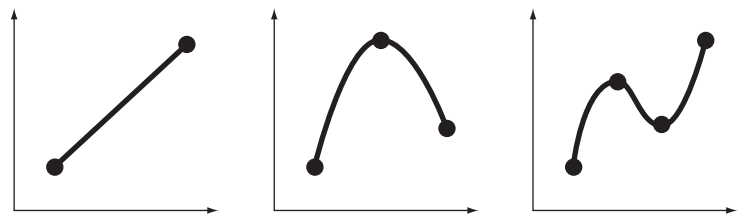
Como afirmado, há uma grande variedade de formas alternativas para expressar um polinômio interpolador. O *polinômio interpolador por diferenças divididas de Newton* está entre as fórmulas mais populares e úteis. Antes de se apresentar a equação geral, serão introduzidas as versões de primeiro e segundo graus por causa de sua interpretação visual simples.

18.1.1 Interpolação Linear

A forma mais simples de interpolação é ligar dois pontos dados com uma reta. A técnica, chamada de *interpolação linear,* é descrita graficamente na Figura 18.2. Usando semelhança de triângulos,

$$\frac{f_1(x) - f(x_0)}{x - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

FIGURA 18.1
Exemplos de polinômios interpoladores: (*a*) de primeiro grau (linear) ligando dois pontos, (*b*) de segundo grau (quadrático ou parabólico) ligando três pontos, e (*c*) de terceiro grau (cúbico) ligando quatro pontos.



[1] N.R.T.: Observe que com esta forma geral, é possível que $a_n = 0$

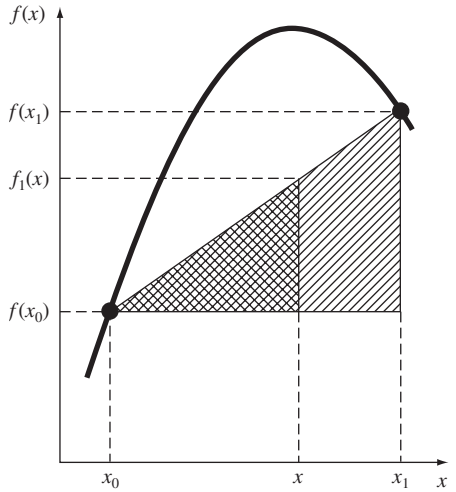


FIGURA 18.2
Descrição gráfica de uma interpolação linear. As áreas hachuradas indicam os triângulos semelhantes usados para deduzir a fórmula de interpolação linear [Equação (18.2)].

a qual pode ser reorganizada para fornecer

$$f_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) \tag{18.2}$$

que é a *fórmula de interpolação linear*. A notação $f_1(x)$ indica que esse é um polinômio interpolador de primeiro grau. Observe que, além de representar a inclinação da reta ligando os pontos, o termo $[f(x_1) - f(x_0)]/(x_1 - x_0)$ é uma aproximação por diferenças divididas finitas da primeira derivada [lembre-se da Equação (4.17)]. Em geral, quanto menor o intervalo entre os pontos dados, melhor a aproximação, o que se deve ao fato de que, conforme o intervalo diminui, uma função contínua será mais bem aproximada por uma reta. Essa característica é ilustrada no seguinte exemplo.

EXEMPLO 18.1 Interpolação Linear

Enunciado do Problema. Faça uma estimativa do logaritmo natural de 2 usando uma interpolação linear. Primeiro, faça o cálculo interpolando entre $\ln 1 = 0$ e $\ln 6 = 1{,}791759$. Então, repita o procedimento, mas use o intervalo menor de ln 1 a ln 4 (1,386294). Observe que o valor verdadeiro de ln 2 é 0,6931472.

Solução. Usamos a Equação (18.2) e a interpolação linear para ln(2) de $x_0 = 1$ a $x_1 = 6$ para obter

$$f_1(2) = 0 + \frac{1{,}791759 - 0}{6 - 1}(2 - 1) = 0{,}3583519$$

o que representa um erro de $\varepsilon_t = 48{,}3\%$. O uso do intervalo menor de $x_0 = 1$ a $x_1 = 4$ fornece

$$f_1(2) = 0 + \frac{1{,}386294 - 0}{4 - 1}(2 - 1) = 0{,}4620981$$

Logo, o uso do intervalo menor reduz o erro relativo porcentual para $\varepsilon_t = 33{,}3\%$. Ambas as interpolações estão mostradas na Figura 18.3, junto com a função verdadeira.

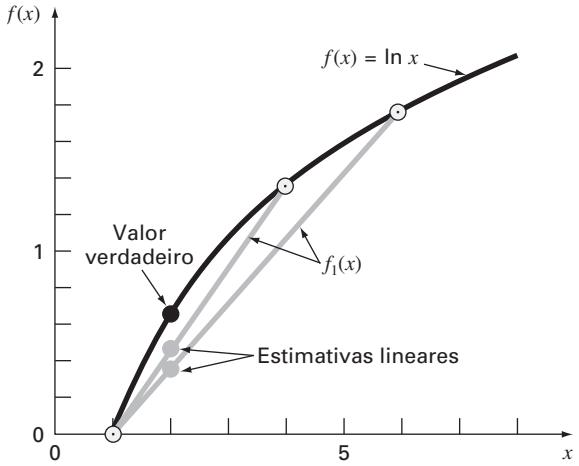


FIGURA 18.3
Duas interpolações lineares para estimar ln 2. Observe como o intervalo menor fornece uma estimativa melhor.

18.1.2 Interpolação Quadrática

O erro no Exemplo 18.1 resultou de se aproximar uma curva por uma reta. Conseqüentemente, uma estratégia para melhorar a estimativa é introduzir alguma curvatura na curva ligando os pontos. Se estiverem disponíveis três pontos, isso pode ser conseguido com um polinômio de segundo grau (também chamado de polinômio quadrático ou uma *parábola*). Uma fórmula particularmente conveniente para esse propósito é

$$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1) \tag{18.3}$$

Observe que, embora a Equação (18.3) possa parecer diferente do polinômio geral [Equação (18.1)], as duas equações são equivalentes, o que se pode mostrar multiplicando os termos na Equação (18.3) para fornecer

$$f_2(x) = b_0 + b_1x - b_1x_0 + b_2x^2 + b_2x_0x_1 - b_2xx_0 - b_2xx_1$$

ou, colecionando os termos,

$$f_2(x) = a_0 + a_1x + a_2x^2$$

onde

$$a_0 = b_0 - b_1x_0 + b_2x_0x_1$$
$$a_1 = b_1 - b_2x_0 - b_2x_1$$
$$a_2 = b_2$$

Logo, as Equações (18.1) e (18.3) são formulações alternativas equivalentes do único polinômio de segundo grau ligando os três pontos.

Um procedimento simples pode ser usado para determinar os valores dos coeficientes. Para b_0, a Equação (18.3) com $x = x_0$ pode ser usada para calcular

$$b_0 = f(x_0) \tag{18.4}$$

A Equação (18.4) pode ser substituída na Equação (18.3), a qual pode ser calculada em $x = x_1$ para

$$b_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \tag{18.5}$$

Finalmente, as Equações (18.4) e (18.5) podem ser substituídas na Equação (18.3), a qual pode ser calculada em $x = x_2$ e resolvida (depois de algumas manipulações algébricas) por

$$b_2 = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} \tag{18.6}$$

Observe que, como foi o caso com a interpolação linear, b_1 ainda representa a inclinação da reta ligando os pontos x_0 e x_1. Logo, os dois primeiros termos da Equação (18.3) são equivalentes à interpolação linear de x_0 a x_1, como especificado anteriormente na Equação (18.2). O último termo, $b_2(x - x_0)(x - x_1)$, introduz a curvatura de segundo grau na fórmula.

Antes de ilustrar como usar a Equação (18.3), é preciso examinar a forma do coeficiente b_2. Ela é muito parecida com a aproximação por diferença dividida finita da segunda derivada introduzida anteriormente na Equação (4.24). Logo, a Equação (18.3) começa a manifestar uma estrutura que é muito parecida com a expansão em série de Taylor. Essa observação será mais explorada quando relacionarmos o polinômio interpolador de Newton com a série de Taylor, na Seção 18.1.4. Mas, primeiro, vamos fazer um exemplo que mostra como a Equação (18.3) é usada para interpolar entre três pontos.

EXEMPLO 18.2 Interpolação Quadrática

Enunciado do Problema. Ajuste um polinômio de segundo grau aos três pontos usados no Exemplo 18.1:

$x_0 = 1 \qquad f(x_0) = 0$

$x_1 = 4 \qquad f(x_1) = 1{,}386294$

$x_2 = 6 \qquad f(x_2) = 1{,}791759$

Use o polinômio para calcular ln 2.

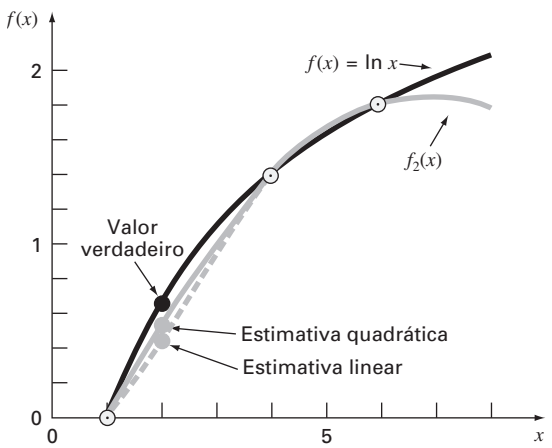
Solução. Aplicando a Equação (18.4), obtém-se

$$b_0 = 0$$

A Equação (18.5) fornece

$$b_1 = \frac{1{,}386294 - 0}{4 - 1} = 0{,}4620981$$

FIGURA 18.4
O uso da interpolação quadrática para fazer uma estimativa de ln 2. A interpolação linear de $x = 1$ a 4 também está incluída para comparação.



e a Equação (18.6) dá

$$b_2 = \frac{\dfrac{1{,}791759 - 1{,}386294}{6 - 4} - 0{,}4620981}{6 - 1} = -0{,}0518731$$

Substituindo esses valores na Equação (18.3), obtém-se a fórmula quadrática

$$f_2(x) = 0 + 0{,}4620981(x - 1) - 0{,}0518731(x - 1)(x - 4)$$

a qual pode ser calculada em $x = 2$ para fornecer

$$f_2(2) = 0{,}5658444$$

o que representa um erro relativo de $\varepsilon_t = 18{,}4\%$. Logo, a curvatura introduzida pela fórmula quadrática (Figura 18.4) melhora a interpolação quando comparada com o resultado obtido usando-se retas nos Exemplos 18.1 e 18.3.

18.1.3 Forma Geral dos Polinômios Interpoladores de Newton

A análise anterior pode ser generalizada para ajustar um polinômio de grau n a $n + 1$ pontos dados. O polinômio de grau n é

$$f_n(x) = b_0 + b_1(x - x_0) + \cdots + b_n(x - x_0)(x - x_1)\cdots(x - x_{n-1}) \tag{18.7}$$

Como foi feito anteriormente com as interpolações linear e quadrática, os pontos dados podem ser usados para calcular os coeficientes $b_0, b_1, \ldots, b_n$. Para um polinômio de grau n, $n + 1$ pontos dados são necessários: $(x_0, f(x_0))$, $(x_1, f(x_1))$, . . . , $(x_n, f(x_n))$. Usamos esses pontos dados e as seguintes equações para calcular os coeficientes:

$$b_0 = f(x_0) \tag{18.8}$$

$$b_1 = f[x_1, x_0] \tag{18.9}$$

$$b_2 = f[x_2, x_1, x_0] \tag{18.10}$$

$$\vdots$$

$$b_n = f[x_n, x_{n-1}, \ldots, x_1, x_0] \tag{18.11}$$

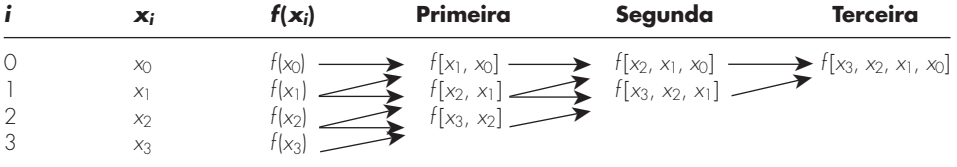
onde a função com colchetes corresponde a diferenças divididas finitas. Por exemplo, a primeira diferença dividida finita é representada em geral por

$$f[x_i, x_j] = \frac{f(x_i) - f(x_j)}{x_i - x_j} \tag{18.12}$$

A *segunda diferença dividida finita*, que representa a diferença das duas primeiras diferenças divididas, é expressa em geral por

$$f[x_i, x_j, x_k] = \frac{f[x_i, x_j] - f[x_j, x_k]}{x_i - x_k} \tag{18.13}$$

FIGURA 18.5
Descrição gráfica da natureza recursiva das diferenças divididas finitas.



Analogamente, a *n-ésima diferença dividida finita* é

$$f[x_n, x_{n-1}, \ldots, x_1, x_0] = \frac{f[x_n, x_{n-1}, \ldots, x_1] - f[x_{n-1}, x_{n-2}, \ldots, x_0]}{x_n - x_0} \tag{18.14}$$

Tais diferenças podem ser usadas para calcular os coeficientes nas Equações (18.8) a (18.11), os quais podem então ser substituídos na Equação (18.7) para fornecer o polinômio interpolador

$$\begin{aligned} f_n(x) = f(x_0) + (x - x_0) f[x_1, x_0] + (x - x_0)(x - x_1) f[x_2, x_1, x_0] \\ + \cdots + (x - x_0)(x - x_1) \cdots (x - x_{n-1}) f[x_n, x_{n-1}, \ldots, x_0] \end{aligned} \tag{18.15}$$

que é chamado *polinômio interpolador por diferenças divididas de Newton*. Deve ser observado que não é necessário que os dados usados na Equação (18.15) sejam igualmente espaçados ou que os valores das abscissas estejam necessariamente em ordem crescente, como ilustrado no exemplo a seguir. Observe também como as Equações (18.12) a (18.14) são recursivas — isto é, diferenças de ordem mais alta são calculadas tomando-se diferenças das diferenças de ordem mais baixa (Figura 18.5). Essa propriedade será explorada quando for desenvolvido um programa computacional eficiente na Seção 18.1.5 para implementar o método.

EXEMPLO 18.3 Polinômio Interpolador por Diferenças Divididas de Newton

Enunciado do Problema. No Exemplo 18.2, dados em $x_0 = 1$, $x_1 = 4$ e $x_2 = 6$ foram usados para fazer uma estimativa de $\ln 2$ com uma parábola. Agora, adicionando um quarto ponto [$x_3 = 5$; $f(x_3) = 1{,}609438$], faça uma estimativa de $\ln 2$ com um polinômio interpolador de Newton de terceiro grau.

Solução. O polinômio de terceiro grau, a Equação (18.7) com $n = 3$, é

$$f_3(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1) + b_3(x - x_0)(x - x_1)(x - x_2)$$

As primeiras diferenças divididas para o problema são [Equação (18.12)]

$$f[x_1, x_0] = \frac{1{,}386294 - 0}{4 - 1} = 0{,}4620981$$

$$f[x_2, x_1] = \frac{1{,}791759 - 1{,}386294}{6 - 4} = 0{,}2027326$$

$$f[x_3, x_2] = \frac{1{,}609438 - 1{,}791759}{5 - 6} = 0{,}1823216$$

As segundas diferenças divididas são [Equação (18.13)]

$$f[x_2, x_1, x_0] = \frac{0{,}2027326 - 0{,}4620981}{6 - 1} = -0{,}05187311$$

$$f[x_3, x_2, x_1] = \frac{0{,}1823216 - 0{,}2027326}{5 - 4} = -0{,}02041100$$

A terceira diferença dividida é [Equação (18.4) com $n = 3$]

$$f[x_3, x_2, x_1, x_0] = \frac{-0{,}02041100 - (-0{,}05187311)}{5 - 1} = 0{,}007865529$$

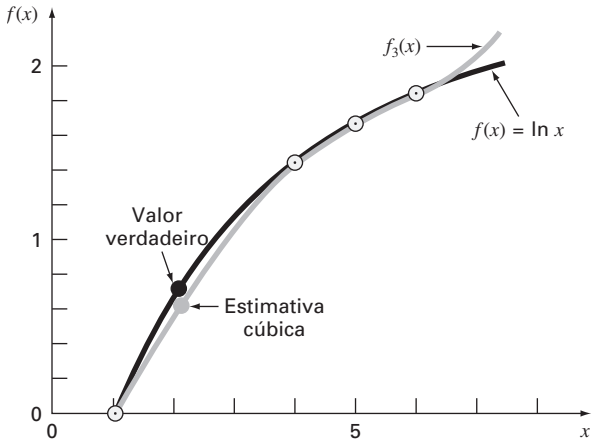


FIGURA 18.6
O uso da interpolação cúbica para estimar ln 2.

Os resultados para $f[x_1, x_0]$, $f[x_2, x_1, x_0]$ e $f[x_3, x_2, x_1, x_0]$ representam os coeficientes b_1, b_2 e b_3, respectivamente, da Equação (18.7). Junto com $b_0 = f(x_0) = 0{,}0$, a Equação (18.7) é

$$f_3(x) = 0 + 0{,}4620981(x - 1) - 0{,}05187311(x - 1)(x - 4) + 0{,}007865529(x - 1)(x - 4)(x - 6)$$

a qual pode ser usada para calcular $f_3(2) = 0{,}6287686$, o que representa um erro relativo de $\varepsilon_t = 9{,}3\%$. O polinômio cúbico completo está mostrado na Figura 18.6.

18.1.4 Erros nos Polinômios Interpoladores de Newton

Observe que a estrutura da Equação (18.15) é parecida com a expansão em série de Taylor no sentido que os termos são adicionados seqüencialmente para capturar o comportamento de ordem superior da função subjacente. Esses termos são diferenças divididas finitas e, portanto, representam aproximações de derivadas de ordem superior. Conseqüentemente, como no caso da série de Taylor, se a função subjacente verdadeira for um polinômio de grau n, o polinômio interpolador de grau n baseado em $n + 1$ pontos dados fornecerá resultados exatos.

Além disso, como no caso da série de Taylor, pode ser obtida uma formulação para o erro de truncamento. Lembre-se, a partir da Equação (4.6), de que o erro de truncamento para a série de Taylor podia ser, no caso geral, expresso por

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x_{i+1} - x_i)^{n+1} \tag{4.6}$$

onde ξ é algum ponto no intervalo x_i a x_{i+1}. Para um polinômio interpolador de grau n, uma relação análoga para o erro é

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)(x - x_1)\cdots(x - x_n) \tag{18.16}$$

onde ξ é algum ponto no intervalo contendo a variável e os dados. Para essa fórmula ser útil, a função em questão deve ser conhecida e $(n + 1)$ vezes diferenciável. Em geral, esse não é o caso. Felizmente, está disponível uma formulação alternativa que não requer conhecimento anterior da função. Em vez disso, ela usa diferenças divididas finitas para aproximar a $(n + 1)$-ésima derivada,

$$R_n = f[x, x_n, x_{n-1}, \ldots, x_0](x - x_0)(x - x_1)\cdots(x - x_n) \tag{18.17}$$

onde $f[x, x_n, x_{n-1}, \ldots, x_0]$ é a $(n + 1)$-ésima diferença dividida finita. Como a Equação (18.17) contém a incógnita $f(x)$, ela não determina o erro. Entretanto, se estiver disponível um ponto dado adicional $f(x_{n+1})$, a Equação (18.17) pode ser usada para estimar o erro, como em

$$R_n \cong f[x_{n+1}, x_n, x_{n-1}, \ldots, x_0](x - x_0)(x - x_1) \cdots (x - x_n) \tag{18.18}$$

EXEMPLO 18.4 Estimativa do Erro para o Polinômio de Newton

Enunciado do Problema. Use a Equação (18.18) para fazer uma estimativa do erro para o polinômio interpolador de segundo grau do Exemplo 18.2. Utilize o ponto adicional $f(x_3) = f(5) = 1{,}609438$ para obter seus resultados.

Solução. Lembre-se de que, no Exemplo 18.2, o polinômio interpolador de segundo grau fornecia uma estimativa de $f_2(2) = 0{,}5658444$, a qual representa um erro de $0{,}6931472 - 0{,}5658444 = 0{,}1273028$. Se não soubéssemos o valor verdadeiro, que é o caso mais comum, a Equação (18.18), junto com o valor adicional em x_3, poderia ter sido usada para fazer uma estimativa do erro, como em

$$R_2 = f[x_3, x_2, x_1, x_0](x - x_0)(x - x_1)(x - x_2)$$

ou

$$R_2 = 0{,}007865529(x - 1)(x - 4)(x - 6)$$

onde o valor da diferença dividida finita de terceira ordem é como calculado anteriormente no Exemplo 18.3. Essa relação pode ser calculada em $x = 2$ por

$$R_2 = 0{,}007865529(2 - 1)(2 - 4)(2 - 6) = 0{,}0629242$$

que é da mesma ordem de grandeza que o erro verdadeiro.

Dos exemplos anteriores e da Equação (18.18), deveria ficar claro que a estimativa de erro para o polinômio de grau n é equivalente à diferença entre as previsões de ordem $(n + 1)$ e de ordem n. Isto é,

$$R_n = f_{n+1}(x) - f_n(x) \tag{18.19}$$

Em outras palavras, o incremento que é somado ao caso de ordem n para criar o caso de ordem $(n + 1)$ [ou seja, a Equação (18.19)] é interpretado como uma estimativa do erro de ordem n. Isso pode ser visto claramente reorganizando-se a Equação (18.19) para obter

$$f_{n+1}(x) = f_n(x) + R_n$$

A validade dessa aproximação se baseia no fato de que a série é fortemente convergente. Para tal situação, a previsão de ordem $(n + 1)$ deveria estar muito mais próxima do valor verdadeiro do que a previsão de ordem n. Conseqüentemente, a Equação (18.19) está de acordo com nossa definição padrão de erro como representando a diferença entre o valor verdadeiro e uma aproximação. Entretanto, observe que, enquanto todas as outras estimativas de erro para processos iterativos introduzidos até esse ponto foram determinadas como a previsão presente menos uma previsão anterior, a Equação (18.19) representa uma previsão futura menos a previsão atual. Isso significa que, para uma série que está convergindo rapidamente, a estimativa de erro da Equação (18.19) poderia ser menor do que o erro verdadeiro, o que representaria uma característica altamente insatisfatória se a estimativa de erro estivesse sendo usada como critério de parada. Contudo, como será descrito na próxima seção, polinômios interpoladores de grau mais alto são muito sensíveis a erros nos dados — ou seja, são muito mal condicionados. Quando usados para interpolação, em geral fornecem previsões que divergem significativamente do valor verdadeiro. Como "olha à frente" para perceber os erros, a Equação (18.19) é mais sensível a tal divergência. Por isso, é mais valiosa para o tipo de análise de dados exploratória para a qual o polinômio de Newton é mais bem adequado.

18.1.5 Algoritmo Computacional para o Polinômio Interpolador de Newton

Três propriedades tornam o polinômio interpolador de Newton extremamente atrativo para aplicações computacionais:

1. Como na Equação (18.7), versões de ordem mais alta podem ser desenvolvidas seqüencialmente adicionando-se um único termo à próxima equação de ordem mais baixa. Isso facilita o cálculo das diversas versões diferentes no mesmo programa. Tal recurso é extremamente valioso quando o grau do polinômio não é conhecido *a priori*. Adicionando-se novos termos seqüencialmente, pode-se determinar quando um ponto de retorno diminuído é atingido — isto é, quando a adição de termos de grau mais alto não melhora significativamente a estimativa ou, em certas situações, a piora. As equações de erro discutidas a seguir em (3) são úteis no desenvolvimento de um critério objetivo para identificar esse ponto de retorno diminuído.
2. As diferenças divididas finitas que constituem os coeficientes do polinômio [Equações (18.8) a (18.11)] podem ser eficientemente calculadas. Ou seja, como na Equação (18.14) e na Figura 18.5, diferenças de ordem mais baixa são usadas para calcular diferenças de ordem mais alta. Utilizando essa informação determinada anteriormente, os coeficientes podem ser calculados eficientemente. O algoritmo na Figura 18.7 contém tal esquema.
3. A estimativa de erro [Equação (18.18)] pode ser incorporada de modo muito simples em um algoritmo computacional por causa da forma seqüencial na qual a previsão é construída.

Todas essas características podem ser exploradas e incorporadas em um algoritmo geral para implementar o polinômio de Newton (Figura 18.7). Observe que o algoritmo consiste em duas partes: a primeira determina os coeficientes a partir da Equação (18.7), e a segunda, as previsões e seus erros associados. A utilidade desse algoritmo é ilustrada no próximo exemplo.

FIGURA 18.7
Um algoritmo para o polinômio interpolador de Newton escrito em pseudocódigo.

```
SUBROUTINE NewtInt (x, y, n, xi, yint, ea)
  LOCAL fdd(n,n)
  DOFOR i = 0, n
    fdd(i,0) = y(i)
  END DO
  DOFOR j = 1, n
    DOFOR i = 0, n − j
      fdd(i,j) = (fdd(i+1,j−1) − fdd(i,j−1))/(x(i+j) − x(i))
    END DO
  END DO
  xterm = 1
  yint(0) = fdd(0,0)
  DOFOR order = 1, n
    xterm = xterm * (xi − x(order−1))
    yint2 = yint(order−1) + fdd(0,order) * xterm
    Ea(order−1) = yint2 − yint(order−1)
    yint(order) = yint2
  END order
END NewtInt
```

EXEMPLO 18.5 Estimativa de Erro para Determinar o Grau Adequado da Interpolação

Enunciado do Problema. Depois de incorporar o erro [Equação (18.18)], utilize o algoritmo computacional dado na Figura 18.7 e a seguinte informação para calcular $f(x) = \ln x$ em $x = 2$:

x	$f(x) = \ln x$
1	0
4	1,3862944
6	1,7917595
5	1,6094379
3	1,0986123
1,5	0,4054641
2,5	0,9162907
3,5	1,2527630

Solução. Os resultados da utilização do algoritmo na Figura 18.7 para obter uma solução estão mostrados na Figura 18.8. As estimativas de erros, junto com o erro verdadeiro (baseado no fato de que $\ln 2 = 0{,}6931472$), são descritos na Figura 18.9. Observe que a estimativa do erro e o erro verdadeiro são parecidos e sua concordância melhora conforme o grau aumenta. Desses resultados, pode-se concluir que a versão de quinto grau fornece uma boa estimativa e que termos de grau mais alto não melhoram significativamente a previsão.

Esse exercício também ilustra a importância do posicionamento e da ordenação dos pontos. Por exemplo, até a estimativa de terceiro grau, a taxa de melhoria é lenta, pois os pontos que foram adicionados (em $x = 4$, 6 e 5) estão distantes e para um mesmo lado do ponto em questão, $x = 2$. A estimativa de quarto grau mostra uma melhora um pouco maior, pois o novo ponto em $x = 3$ está mais próximo da incógnita. Entretanto, a diminuição mais dramática no erro está associada com a inclusão do termo de quinto grau usando o ponto em $x = 1{,}5$. Esse ponto não só está próximo da incógnita, mas está posicionado do lado oposto da maioria dos outros pontos. Como conseqüência, o erro é reduzido de quase uma ordem de grandeza.

FIGURA 18.8
A saída do programa, com base no algoritmo da Figura 18.7 para o cálculo de ln 2.

```
NÚMERO DE PONTOS? 8
X( 0 ), y( 0 ) = ? 1,0
X( 1 ), y( 1 ) = ? 4,1.3862944
X( 2 ), y( 2 ) = ? 6,1.7917595
X( 3 ), y( 3 ) = ? 5,1.6094379
X( 4 ), y( 4 ) = ? 3,1.0986123
X( 5 ), y( 5 ) = ? 1.5,0.40546411
X( 6 ), y( 6 ) = ? 2.5,0.91629073
X( 7 ), y( 7 ) = ? 3.5,1.2527630

INTERPOLAÇÃO EM X = 2
GRAU      F(X)          ERRO
0         0.000000      0.462098
1         0.462098      0.103746
2         0.565844      0.062924
3         0.628769      0.046953
4         0.675722      0.021792
5         0.697514      -0.003616
6         0.693898      -0.000459
7         0.693439
```

O significado da posição e da seqüência de dados também pode ser ilustrado usando-se os mesmos dados para obter uma estimativa de ln 2, mas considerando os pontos em uma seqüência diferente. A Figura 18.9 mostra resultados para o caso de se reverter a ordem dos dados originais, isto é, $x_0 = 3{,}5$, $x_1 = 2{,}5$, $x_3 = 1{,}5$ e assim por diante. Como os pontos iniciais, nesse caso, estão mais próximos e espaçados de ambos os lados de ln 2, o erro decresce muito mais rapidamente do que para a situação original. Considerando o termo de 2º grau, o erro já terá sido reduzido para menos do que $\varepsilon_t = 2\%$. Outras combinações poderiam ser usadas para obter diferentes taxas de convergência.

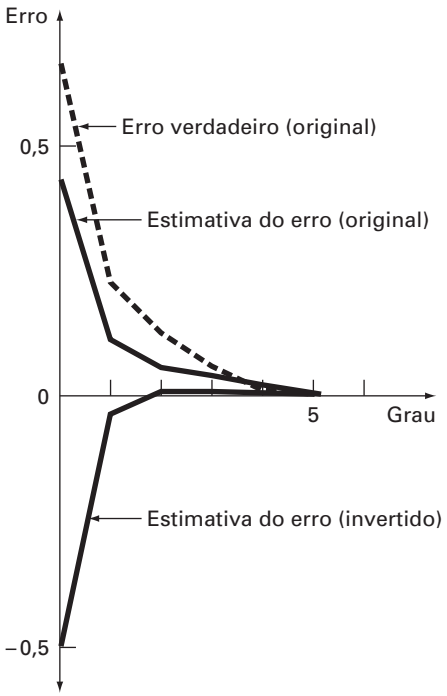


FIGURA 18.9
Erros relativos percentuais para a previsão de ln 2 como função do grau do polinômio interpolador.

O exemplo anterior ilustra a importância da escolha dos pontos base. Como deveria ser intuitivamente óbvio, os pontos teriam de estar centralizados em torno e tão próximo quanto possível da incógnita. Essa observação também é confirmada por um exame direto da equação do erro [Equação (18.17)]. Se supusermos que a diferença dividida finita não varia muito no intervalo dos dados, o erro é proporcional ao produto: $(x - x_0)\,(x - x_1) \cdot \cdot \cdot (x - x_n)$. Obviamente, quanto mais próximos os pontos base estiverem de x, menor será o valor absoluto do produto.

18.2 POLINÔMIOS INTERPOLADORES DE LAGRANGE

O *polinômio interpolador de Lagrange* é simplesmente uma reformulação do polinômio de Newton que evita o cálculo de diferenças divididas. Ele pode ser representado concisamente por:

$$f_n(x) = \sum_{i=0}^{n} L_i(x) f(x_i) \tag{18.20}$$

onde

$$L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j} \tag{18.21}$$

onde Π indica o "produto de". Por exemplo, a versão linear ($n = 1$) seria

$$f_1(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1) \tag{18.22}$$

e a versão de segundo grau seria

$$f_2(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} f(x_0) + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} f(x_1) + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} f(x_2) \tag{18.23}$$

A Equação (18.20) pode ser deduzida diretamente do polinômio de Newton (Quadro 18.1). Entretanto, o raciocínio por trás da formulação de Lagrange pode ser entendido diretamente percebendo-se que cada termo $L_i(x)$ será 1 em $x = x_i$ e 0 em todos os outros pontos da amostra (Figura 18.10). Logo, cada produto $L_i(x) f(x_i)$ assume o valor $f(x_i)$ no ponto x_i da amostra. Conseqüentemente, a somatória de todos os produtos indicados na Equação (18.20) é o único polinômio de grau n que passa exatamente por todos os $n + 1$ pontos dados.

Quadro 18.1 Dedução da Forma de Lagrange Diretamente do Polinômio Interpolador de Newton

O polinômio interpolador de Lagrange pode ser diretamente deduzido da formulação de Newton. Vamos fazer isso apenas para o caso de primeiro grau [Equação (18.2)]. Para deduzir a forma de Lagrange, reformulamos as diferenças divididas. Por exemplo, a primeira diferença dividida,

$$f[x_1, x_0] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} \tag{Q18.1.1}$$

pode ser reformulada como

$$f[x_1, x_0] = \frac{f(x_1)}{x_1 - x_0} + \frac{f(x_0)}{x_0 - x_1} \tag{Q18.1.2}$$

que é chamada de *forma simétrica*. A substituição da Equação (Q18.1.2) na Equação (18.2) fornece

$$f_1(x) = f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1) + \frac{x - x_0}{x_0 - x_1} f(x_0)$$

Finalmente, agrupando os termos similares e simplificando, obtemos a forma de Lagrange,

$$f_1(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1)$$

EXEMPLO 18.6 Polinômios Interpoladores de Lagrange

Enunciado do Problema. Use um polinômio interpolador de Lagrange de primeiro e de segundo graus para calcular ln 2 com base nos dados fornecidos no Exemplo 18.2:

$$x_0 = 1 \qquad f(x_0) = 0$$
$$x_1 = 4 \qquad f(x_1) = 1{,}386294$$
$$x_2 = 6 \qquad f(x_2) = 1{,}791760$$

Solução. O polinômio de primeiro grau [Equação (18.22)] pode ser usado para se obter a estimativa em $x = 2$,

$$f_1(2) = \frac{2 - 4}{1 - 4} 0 + \frac{2 - 1}{4 - 1} 1{,}386294 = 0{,}4620981$$

De modo semelhante, o polinômio de segundo grau é desenvolvido como [Equação (18.23)]

$$f_2(2) = \frac{(2-4)(2-6)}{(1-4)(1-6)}0 + \frac{(2-1)(2-6)}{(4-1)(4-6)}1.386294 + \frac{(2-1)(2-4)}{(6-1)(6-4)}1.791760 = 0.5658444$$

Como esperado, ambos os resultados coincidem com os obtidos anteriormente usando-se o polinômio interpolador de Newton.

Observe que, como no método de Newton, a versão de Lagrange tem uma estimativa do erro de [Equação (18.17)]

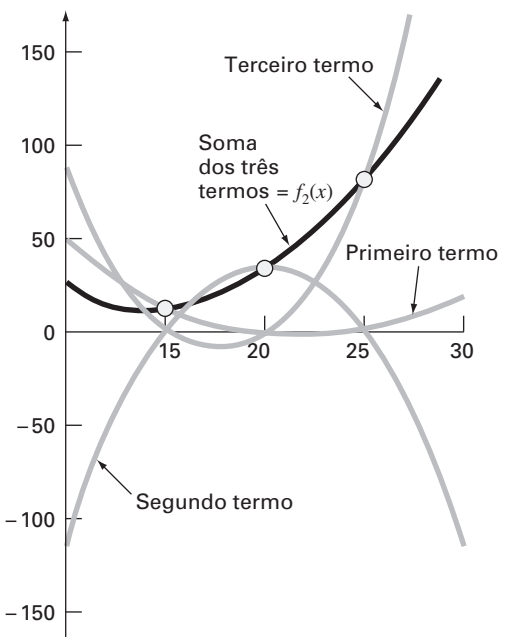
$$R_n = f[x, x_n, x_{n-1}, \ldots, x_0] \prod_{i=0}^{n} (x - x_i)$$

Logo, se um ponto adicional estiver disponível em $x = x_{n+1}$, uma estimativa do erro pode ser obtida. Entretanto, como as diferenças divididas finitas não são usadas como parte do algoritmo de Lagrange, isso raramente é feito.

As Equações (18.20) e (18.21) podem ser programadas de forma muito simples para a implementação em um computador. A Figura 18.11 mostra o pseudocódigo que pode ser usado para tal propósito.

Em resumo, nos casos nos quais o grau do polinômio é desconhecido, o método de Newton tem vantagens por causa da percepção que ele fornece do comportamento das fórmulas para as diferentes ordens. Além disso, a estimativa do erro representada pela Equação (18.18) pode ser facilmente integrada ao algoritmo de Newton, pois a estimativa usa uma diferença finita (Exemplo 18.5). Logo, para cálculos exploratórios, o método de Newton em geral é preferível.

FIGURA 18.10
Uma descrição visual do raciocínio por trás dos polinômios de Lagrange. A figura mostra um caso de segundo grau. Cada um dos três termos na Equação (18.23) passa por um dos pontos dados e é zero nos outros dois. A soma dos três termos deve, portanto, ser o único polinômio de segundo grau $f_2(x)$ que passa pelos três pontos.



```
FUNCTION Lagrng(x, y, n, xx)
  sum = 0
  DOFOR i = 0, n
    product = y_i
    DOFOR j = 0, n
      IF i ≠ j THEN
        product = product*(xx − x_j)/(x_i − x_j)
      ENDIF
    END DO
    sum = sum + product
  END DO
  Lagrng = sum
END Lagrng
```

FIGURA 18.11
Pseudocódigo para implementar a interpolação de Lagrange. Esse algoritmo é planejado para calcular uma única previsão de grau n, onde $n + 1$ é o número de pontos dados.

Quando deve ser feita apenas uma interpolação, as formulações de Newton e de Lagrange exigem esforços computacionais comparáveis. Entretanto, a versão de Lagrange é um pouco mais fácil de programar. Como não exige cálculos nem armazenamento de diferenças divididas, a forma de Lagrange em geral é usada quando o grau do polinômio é conhecido *a priori*.

EXEMPLO 18.7 Interpolação de Lagrange Usando o Computador

Enunciado do Problema. Pode-se usar o algoritmo da Figura 18.11 para estudar o problema de análise de tendência associado com o agora familiar pára-quedista em queda livre. Suponha que tenhamos

Tempo, s	Velocidade medida v, cm/s
1	800
3	2310
5	3090
7	3940
13	4755

Nosso problema é fazer uma estimativa da velocidade do pára-quedista em $t = 10$ s para preencher o grande salto nas medidas entre $t = 7$ e $t = 13$ s. Estamos cientes de que o comportamento dos polinômios interpoladores pode ser inesperado. Logo, construiremos polinômios de quarto, terceiro, segundo e primeiro graus e compararemos os resultados.

Solução. O algoritmo de Lagrange pode ser usado para construir polinômios interpoladores de quarto, terceiro, segundo e primeiro graus.

Pode-se traçar o polinômio de quarto grau e os dados de entrada como mostrado na Figura 18.12*a*. É evidente, desse gráfico, que o valor estimado de y em $x = 10$ é mais alto do que a tendência geral dos dados.

As Figuras 18.2*b* a *d* mostram os gráficos dos resultados dos cálculos para polinômios interpoladores de terceiro, segundo e primeiro graus, respectivamente. Observa-se que, quanto menor o grau, menor será a estimativa do valor da velocidade em $t = 10$ s. Os gráficos dos polinômios interpoladores indicam que os polinômios de grau mais alto tendem a superestimar a tendência dos dados. Isso sugere que as versões de primeiro e segundo graus são mais adequadas para essa análise de tendência particular. Entretanto, deve-se lembrar que, como estamos lidando com dados incertos, a regressão seria, na realidade, mais apropriada.

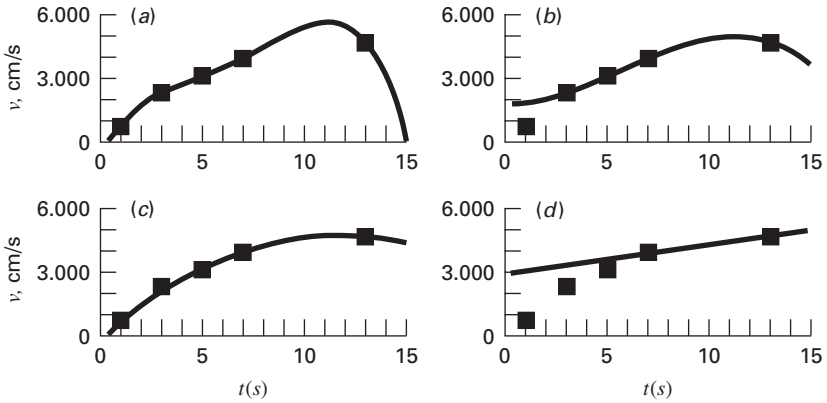


FIGURA 18.12
Gráficos mostrando interpolações de (*a*) quarto grau, (*b*) terceiro grau, (*c*) segundo grau e (*d*) primeiro grau.

O exemplo anterior ilustra que polinômios de grau mais alto tendem a ser mal condicionados, isto é, tendem a ser altamente sensíveis a erros de arredondamento. O mesmo problema se aplica à regressão polinomial de grau mais alto. A aritmética de dupla precisão às vezes ajuda a aliviar o problema. Entretanto, conforme o grau aumenta, existirá um ponto no qual os erros de arredondamento interferirão com a habilidade de interpolar usando as abordagens simples descritas até aqui.

18.3 COEFICIENTES DE UM POLINÔMIO INTERPOLADOR

Embora ambos os polinômios, de Newton e de Lagrange, sejam adequados para a determinação de valores intermediários entre pontos, não fornecem um polinômio conveniente na forma convencional

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \tag{18.24}$$

Um método direto para calcular os coeficientes desses polinômios é baseado no fato de que são necessários $n + 1$ pontos dados para determinar os $n + 1$ coeficientes. Logo, equações algébricas lineares simultâneas podem ser usadas para calcular os a's. Por exemplo, suponha que você queira calcular os coeficientes da parábola

$$f(x) = a_0 + a_1x + a_2x^2 \tag{18.25}$$

São necessários os três pontos dados: $(x_0, f(x_0))$, $(x_1, f(x_1))$ e $(x_2, f(x_2))$. Cada um pode ser substituído na Equação (18.25) para fornecer

$$\begin{aligned} f(x_0) &= a_0 + a_1x_0 + a_2x_0^2 \\ f(x_1) &= a_0 + a_1x_1 + a_2x_1^2 \\ f(x_2) &= a_0 + a_1x_2 + a_2x_2^2 \end{aligned} \tag{18.26}$$

Logo, nesse caso, os x's são conhecidos e os a's são as incógnitas. Como há o mesmo número de equações e de incógnitas, a Equação (18.26) poderia ser resolvida por um método de eliminação da Parte Três.

Deve ser observado que a abordagem anterior não é o método mais eficiente disponível para determinar os coeficientes de um polinômio interpolador. Press et al. (1992) fornecem uma discussão e códigos computacionais para abordagens mais eficientes. Qualquer que seja a tecnologia usada, deve-se tomar cuidado. Os sistemas como a Equação (18.26) são notoriamente mal condicionados. Tanto se forem resolvidos por um método de eliminação quanto se forem resolvidos com um algoritmo mais eficiente, os coeficientes resultantes podem ser altamente não-acurados, particularmente para valores altos de n. Quando usados em uma interpolação subseqüente, em geral eles fornecem resultados errados.

Em resumo, se você estiver interessado em determinar um ponto intermediário, use interpolação de Newton ou de Lagrange. Se precisar determinar uma equação da forma da Equação (18.24), limite-se a polinômios de grau mais baixo e verifique seus resultados cuidadosamente.

18.4 INTERPOLAÇÃO INVERSA

Como a nomenclatura indica, os valores de $f(x)$ e de x são, na maioria dos contextos de interpolação, as variáveis dependentes e independentes, respectivamente. Como conseqüência, os valores dos x's são, de modo típico, uniformemente espaçados. Um exemplo simples é a tabela de valores deduzidos para a função $f(x) = 1/x$,

x	1	2	3	4	5	6	7
$f(x)$	1	0,5	0,3333	0,25	0,2	0,1667	0,1429

Agora, suponha que você precise usar os mesmos dados, mas lhe foi dado um valor para $f(x)$ e você deve determinar o valor correspondente de x. Por exemplo, para os dados da tabela acima, suponha que lhe foi pedido para determinar o valor de x que corresponde a $f(x) = 0{,}3$. Nesse caso, como a função está disponível e é fácil de manipular, a resposta correta pode ser determinada diretamente como $x = 1/0{,}3 = 3{,}3333$.

Tal problema é chamado de *interpolação inversa*. Para um caso mais complicado, você poderia ser tentado a trocar os valores de $f(x)$ e de x [isto é, simplesmente considerar x como função de $f(x)$] e usar uma abordagem como a interpolação de Lagrange para determinar o resultado. Infelizmente, quando as variáveis são invertidas, não há nenhuma garantia de que os valores ao longo da nova abscissa [os $f(x)$'s] serão regularmente espaçados. De fato, em muitos casos, os valores sofrerão um efeito "telescópico", ou seja, terão a aparência de uma escala logarítmica com alguns pontos adjacentes agrupados e outros muito espalhados. Por exemplo, para $f(x) = 1/x$ o resultado é

$f(x)$	0,1429	0,1667	0,2	0,25	0,3333	0,5	1
x	7	6	5	4	3	2	1

Tal espaçamento não-uniforme na abscissa em geral leva a oscilações no polinômio interpolador resultante. Isso pode ocorrer mesmo para polinômios de grau mais baixo.

Uma estratégia alternativa é ajustar um polinômio interpolador de grau n, $f_n(x)$, aos dados originais [isto é, com $f(x)$ em função de x]. Na maioria dos casos, como os x's são igualmente espaçados, os polinômios não serão mal condicionados. A resposta ao problema então se reduz a determinar o valor de x que torna o polinômio igual ao $f(x)$ dado. Logo, o problema de interpolação se reduz ao problema de raízes!

Por exemplo, para o problema de interpolação inversa, uma abordagem simples seria ajustar um polinômio quadrático aos três pontos: (2, 0,5), (3, 0,3333) e (4, 0,25). O resultado seria

$$f_2(x) = 1{,}08333 - 0{,}375x + 0{,}041667x^2$$

A resposta ao problema de interpolação inversa de encontrar o x correspondente a $f(x) = 0{,}3$ envolveria, portanto, a determinação da raiz de

$$0{,}3 = 1{,}08333 - 0{,}375x + 0{,}041667x^2$$

Nesse caso simples, a fórmula quadrática pode ser usada para calcular

$$x = \frac{0{,}375 \pm \sqrt{(-0{,}375)^2 - 4(0{,}041667)0{,}78333}}{2(0{,}041667)} = \begin{matrix} 5{,}704158 \\ 3{,}295842 \end{matrix}$$

Logo, a segunda raiz, 3,296, é uma boa aproximação do valor verdadeiro de 3,333, Se quiséssemos acurácia adicional, poderia ser usado um polinômio de terceiro ou de quarto grau, junto com um dos métodos de localização de raízes da Parte Dois.

18.5 COMENTÁRIOS ADICIONAIS

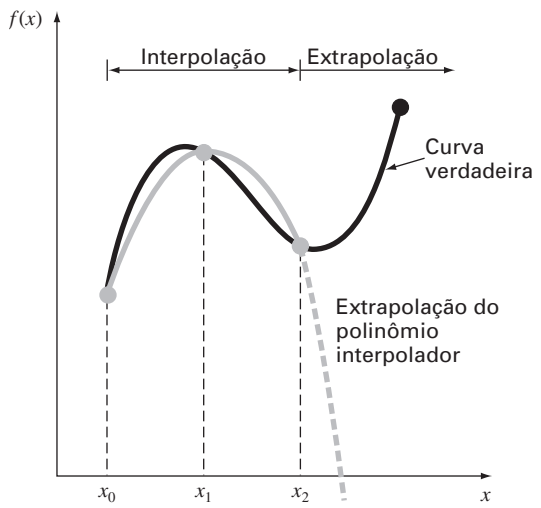
Antes de prosseguir para a próxima seção, deve-se mencionar dois tópicos adicionais: interpolação com dados igualmente espaçados e extrapolação.

Como ambos os polinômios, de Newton e de Lagrange, são compatíveis com dados arbitrariamente espaçados, você poderia imaginar por que tratamos do caso especial de dados igualmente espaçados (Quadro 18.2). Antes do advento dos computadores digitais, essas técnicas tinham grande utilidade para interpolação a partir de tabelas com argumentos igualmente espaçados. Na realidade, um esquema computacional conhecido como tabela de diferenças divididas era desenvolvido para facilitar a implementação de tais técnicas. (A Figura 18.5 é um exemplo dessa tabela.)

Entretanto, como essas fórmulas são subconjuntos dos esquemas de Newton e de Lagrange compatíveis com o computador, e como muitas funções tabuladas estão disponíveis como sub-rotinas de bibliotecas, a necessidade para versões igualmente espaçadas diminuiu. Apesar disso, foram incluídas nesse ponto por causa de sua relevância em partes posteriores deste livro. Em particular, elas são necessárias para deduzir fórmulas de integração numérica que usam tipicamente dados igualmente espaçados (Capítulo 21). Como as fórmulas de integração numérica têm relevância na solução de equações diferenciais ordinárias, o material no Quadro 18.2 também é significativo na Parte Sete.

A *extrapolação* é o processo de fazer uma estimativa de um valor de $f(x)$ que está fora do intervalo dos pontos base conhecidos, $x_0, x_1, \ldots, x_n$ (Figura 18.13). Em uma seção anterior, mencionou-se que a interpolação mais acurada é geralmente obtida quando a incógnita está perto do centro dos pontos base. Obviamente, isso é violado quando a incógnita está fora do intervalo e, conseqüentemente, o erro na extrapolação pode ser muito alto. Como descrito na Figura 18.13, a natureza de extremidade aberta da extrapolação representa um passo no desconhecido, pois o processo estende a curva além da região conhecida. Dessa forma, a curva verdadeira poderia facilmente divergir da previsão. Portanto, deve-se tomar extremo cuidado quando ocorrer um caso no qual seja preciso extrapolar.

FIGURA 18.13
Ilustração da possível divergência de uma previsão extrapolada. A extrapolação é baseada no ajuste de uma parábola pelos primeiros três pontos conhecidos.



Quadro 18.2 Interpolação com Dados Igualmente Espaçados

Se os dados forem igualmente espaçados e estiverem em ordem crescente, então a variável independente assume os valores

$$\begin{aligned} x_1 &= x_0 + h \\ x_2 &= x_0 + 2h \\ &\vdots \\ x_n &= x_0 + nh \end{aligned}$$

onde h é o intervalo, ou o tamanho do passo, entre os dados. Com base nisso, as diferenças divididas finitas podem ser expressas de forma concisa. Por exemplo, a segunda diferença dividida progressiva é

$$f[x_0, x_1, x_2] = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}$$

o que pode ser expresso por

$$f[x_0, x_1, x_2] = \frac{f(x_2) - 2f(x_1) + f(x_0)}{2h^2} \qquad \text{(Q18.2.1)}$$

pois $x_1 - x_0 = x_2 - x_1 = (x_2 - x_0)/2 = h$. Lembre-se agora de que a segunda diferença progressiva é igual a [numerador da Equação (4.24)]

$$\Delta^2 f(x_0) = f(x_2) - 2f(x_1) + f(x_0)$$

Portanto, a Equação (Q18.2.1) pode ser representada por

$$f[x_0, x_1, x_2] = \frac{\Delta^2 f(x_0)}{2!h^2}$$

ou, em geral,

$$f[x_0, x_1, \ldots, x_n] = \frac{\Delta^n f(x_0)}{n!h^n} \qquad \text{(Q18.2.2)}$$

Usando a equação (Q18.2.2), pode-se expressar o polinômio interpolador de Newton [Equação (18.15)] para o caso de dados igualmente espaçados como

$$\begin{aligned} f_n(x) = {} & f(x_0) + \frac{\Delta f(x_0)}{h}(x - x_0) \\ & + \frac{\Delta^2 f(x_0)}{2!h^2}(x - x_0)(x - x_0 - h) \\ & + \cdots + \frac{\Delta^n f(x_0)}{n!h^n}(x - x_0)(x - x_0 - h) \\ & \cdots [x - x_0 - (n-1)h] + R_n \end{aligned} \qquad \text{(Q18.2.3)}$$

onde o resto é o mesmo que a Equação (18.16). Essa equação é conhecida como *fórmula de Newton* ou *fórmula de Newton-Gregory*. Ela pode ser ainda mais simplificada definindo-se uma nova quantidade, α:

$$\alpha = \frac{x - x_0}{h}$$

Essa definição pode ser usada para deduzir as seguintes expressões simplificadas para os termos na Equação (Q18.2.3):

$$\begin{aligned} x - x_0 &= \alpha h \\ x - x_0 - h &= \alpha h - h = h(\alpha - 1) \\ &\vdots \\ x - x_0 - (n-1)h &= \alpha h - (n-1)h = h(\alpha - n + 1) \end{aligned}$$

a qual pode ser substituída na Equação (Q18.2.3) para fornecer

$$\begin{aligned} f_n(x) = {} & f(x_0) + \Delta f(x_0)\alpha + \frac{\Delta^2 f(x_0)}{2!}\alpha(\alpha - 1) \\ & + \cdots + \frac{\Delta^n f(x_0)}{n!}\alpha(\alpha - 1)\cdots(\alpha - n + 1) + R_n \end{aligned} \qquad \text{(Q18.2.4)}$$

onde

$$R_n = \frac{f^{(n+1)}(\xi)}{(n+1)!}h^{n+1}\alpha(\alpha - 1)(\alpha - 2)\cdots(\alpha - n)$$

Essa notação concisa terá utilidade na dedução e análise de erro das fórmulas de integração no Capítulo 21.

Além da fórmula progressiva, fórmulas regressivas e centradas de Newton-Gregory também estão disponíveis. Carnahan, Luther e Wilkes (1969) podem ser consultados para mais informação relativa à interpolação para dados igualmente espaçados.

18.6 INTERPOLAÇÃO POR SPLINES

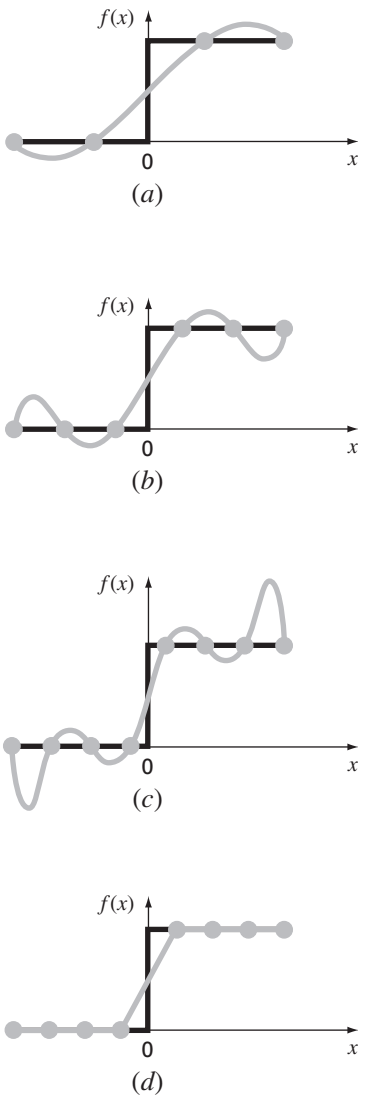
Na seção anterior, polinômios de grau n foram usados para interpolar $n + 1$ pontos dados. Por exemplo, para oito pontos, pode-se determinar exatamente um polinômio de grau sete. Essa curva capturaria todas as oscilações (pelo menos até a sétima derivada, inclusive) sugeridas por esses pontos. Entretanto, há casos em que essas funções podem levar a resultados errôneos por causa de erros de arredondamento e de erros na estimativa. Uma abordagem alternativa é aplicar polinômios de grau mais baixo a subconjuntos dos pontos dados. Tais polinômios conectadores são chamados *funções splines*.

Por exemplo, curvas de terceiro grau usadas para conectar cada par de pontos dados são chamadas de *splines cúbicos*. Essas funções podem ser construídas de modo que as conexões entre equações cúbicas adjacentes sejam visualmente lisas. Superficialmente, pareceria que a aproximação de terceiro grau por splines seria inferior à expressão de grau sete. Você poderia perguntar em quais situações um spline seria preferível.

A Figura 18.14 ilustra uma situação na qual um spline funciona melhor do que um polinômio de grau mais alto. Esse é o caso quando uma função é lisa em geral, mas sofre uma mudança abrupta em algum ponto da região de interesse. O aumento em degrau mostrado na Figura 18.14 é um exemplo extremo de tal mudança e serve para ilustrar esse ponto.

As Figuras 18.14*a* a 18.14*c* ilustram como polinômios de grau mais alto tendem a passar por grandes oscilações na vizinhança de uma mudança abrupta. Em contraste, o spline também liga os pontos, mas como é limitado a variações de grau mais baixo, as oscilações são mantidas em um mínimo. Dessa forma, os splines, em geral, fornecem uma aproximação superior do comportamento de funções que têm variações locais abruptas.

FIGURA 18.14
Uma representação visual da situação na qual os splines são superiores aos polinômios interpoladores de grau mais alto. A função a ser ajustada sofre uma mudança brusca em $x = 0$. As partes (*a*) a (*c*) indicam que a variação abrupta induz oscilações nos polinômios interpoladores. Em contraste, como é limitado a segmentos de reta, um spline linear (*d*) fornece uma aproximação muito mais aceitável.



O conceito de spline originou-se de uma técnica de desenho na qual era usada uma faixa fina e flexível (chamada *spline*) para desenhar uma curva lisa passando por um conjunto de pontos. O processo é descrito na Figura 18.15 para uma série de cinco pinos (pontos dados). Nessa técnica, o desenhista coloca papel sobre uma tábua de madeira e prega tachinhas ou pinos no papel (e na tábua) nas posições dos pontos dados. Uma curva cúbica lisa resulta de intercalar a faixa entre os pinos. Assim, o nome "splines cúbicos" foi adotado para polinômios desse tipo.

Nesta seção, serão usadas primeiro funções lineares simples para introduzir alguns conceitos básicos e problemas associados à interpolação por splines. Então, desenvolveremos um algoritmo para ajustar splines quadráticos aos dados. Finalmente, apresentaremos o material sobre o spline cúbico, que é a versão mais útil e comum na prática da engenharia.

18.6.1 Splines Lineares

A ligação mais simples entre dois pontos é uma reta. Os splines de primeiro grau para um grupo de pontos ordenados podem ser definidos como um conjunto de funções lineares

$$f(x) = f(x_0) + m_0(x - x_0) \qquad x_0 \le x \le x_1$$

$$f(x) = f(x_1) + m_1(x - x_1) \qquad x_1 \le x \le x_2$$

$$\vdots$$

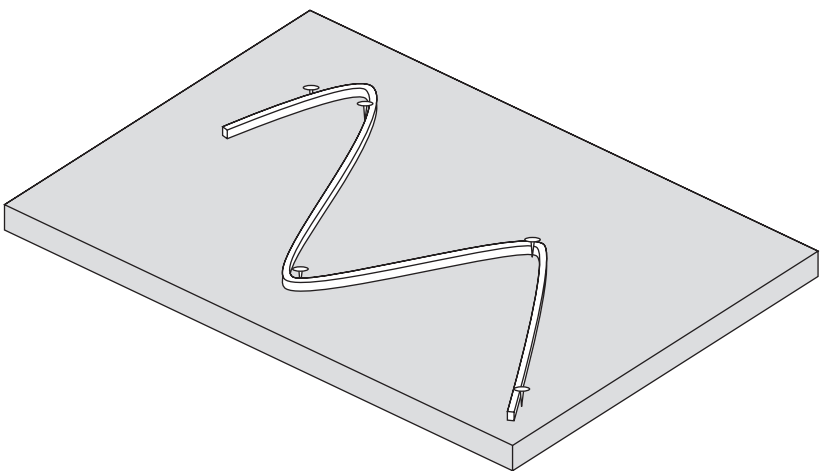
$$f(x) = f(x_{n-1}) + m_{n-1}(x - x_{n-1}) \qquad x_{n-1} \le x \le x_n$$

onde m_i é a inclinação da reta ligando os pontos:

$$m_i = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} \tag{18.27}$$

Essas equações podem ser usadas para calcular a função em qualquer ponto entre x_0 e x_n, primeiro localizando-se o intervalo no qual o ponto se encontra. A seguir, a equação apropriada é usada para determinar o valor da função dentro do intervalo. O método é obviamente idêntico à interpolação linear.

FIGURA 18.15
A técnica de desenho que usa um spline para desenhar curvas lisas por uma série de pontos. Observe como, nas extremidades, o spline se torna menos curvo. Isso é chamado de um spline "natural".



EXEMPLO 18.8 Splines de Primeiro Grau

Enunciado do Problema. Ajuste os dados da Tabela 18.1 com um spline de primeiro grau. Calcule a função em $x = 5$.

Solução. Os dados podem ser usados para determinar as inclinações entre os pontos. Por exemplo, para o intervalo $x = 4{,}5$ a $x = 7$ a inclinação pode ser calculada usando-se a Equação (18.27):

$$m = \frac{2{,}5 - 1}{7 - 4{,}5} = 0{,}60$$

As inclinações para os outros intervalos podem ser calculadas e o spline de primeiro grau resultante está traçado na Figura 18.16*a*. O valor em $x = 5$ é 1,3.

TABELA 18.1
Dados a serem ajustados com funções spline.

x	*f*(*x*)
3,0	2,5
4,5	1,0
7,0	2,5
9,0	0,5

A inspeção visual da Figura 18.16*a* indica que a principal desvantagem dos splines de primeiro grau é que eles não são lisos. Essencialmente, nos pontos dados nos quais dois splines se encontram (chamados nós), a inclinação varia abruptamente. Em termos formais, a primeira derivada da função é descontínua nesses pontos. Essa deficiência é superada usando-se splines polinomiais de grau mais alto, que garantem que eles sejam lisos nos nós, igualando as derivadas em tais pontos, como discutido na próxima seção.

18.6.2 Splines Quadráticos

Para garantir que as *m*-ésimas, derivadas sejam contínuas nos nós, um spline de grau pelo menos $m + 1$ deve ser usado. Polinômios de terceiro grau ou splines cúbicos que garantam continuidade das primeira e segunda derivadas são usados mais freqüentemente na prática. Embora derivadas de terceira ordem ou de ordem mais alta possam ser descontínuas quando usando splines cúbicos, elas não podem ser detectadas visualmente e, conseqüentemente, são ignoradas.

Como a dedução dos splines cúbicos é um pouco complicada, será abordada em uma seção posterior. Primeiro se ilustrará o conceito de interpolação por splines usando polinômios de segundo grau. Esses "splines quadráticos" têm primeira derivada contínua nos nós. Embora os splines quadráticos não garantam segundas derivadas iguais nos nós, eles servem bem para demonstrar o procedimento geral no desenvolvimento de splines de grau mais alto.

O objetivo nos splines quadráticos é determinar um polinômio de segundo grau para cada intervalo entre os pontos dados. Esse polinômio para cada intervalo pode ser representado de forma geral por

$$f_i(x) = a_i x^2 + b_i x + c_i \tag{18.28}$$

A Figura 18.17 foi incluída para ajudar a esclarecer a notação. Para $n + 1$ pontos dados ($i = 0, 1, 2, \ldots, n$), existem n intervalos e, conseqüentemente, $3n$ constantes indeterminadas (os a's, b's e c's) para calcular. Portanto, $3n$ equações ou condições são necessárias para calcular as incógnitas. São elas:

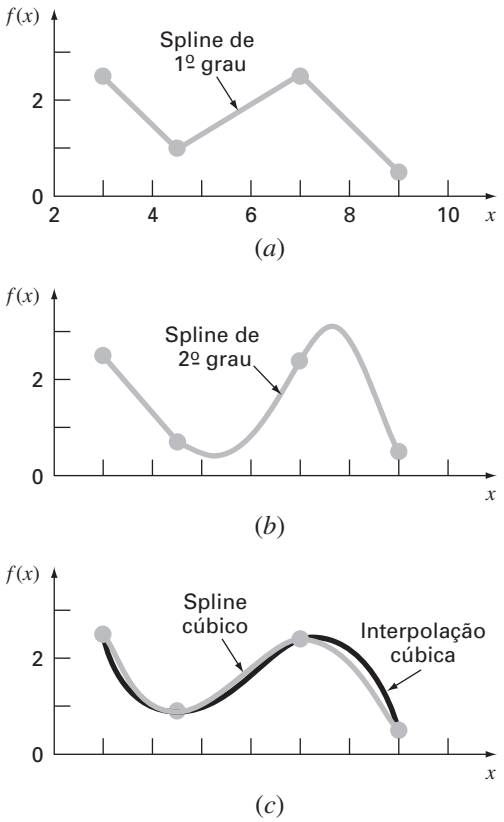


FIGURA 18.16
Ajustes por spline de um conjunto de quatro pontos. (*a*) Spline linear, (*b*) spline quadrático e (*c*) spline cúbico, com um polinômio interpolador cúbico também traçado.

1. *Os valores da função e dos polinômios adjacentes devem ser iguais nos nós interiores*. Essa condição pode ser representada por

$$a_{i-1}x_{i-1}^2 + b_{i-1}x_{i-1} + c_{i-1} = f(x_{i-1}) \tag{18.29}$$

$$a_i x_{i-1}^2 + b_i x_{i-1} + c_i = f(x_{i-1}) \tag{18.30}$$

para $i = 2$ a n. Como apenas os nós interiores foram usados, as Equações (18.29) e (18.30) fornecem cada uma $n - 1$ condições para um total de $2n - 2$ condições.

2. *As primeira e última funções devem passar pelos pontos extremos*. Isso acrescenta duas equações adicionais:

$$a_1 x_0^2 + b_1 x_0 + c_1 = f(x_0) \tag{18.31}$$

$$a_n x_n^2 + b_n x_n + c_n = f(x_n) \tag{18.32}$$

para um total de $2n - 2 + 2 = 2n$ condições.

3. *As primeiras derivadas nos nós interiores devem ser iguais*. A primeira derivada da Equação (18.28) é

$$f'(x) = 2ax + b$$

Portanto, a condição pode ser representada de modo geral por

$$2a_{i-1}x_{i-1} + b_{i-1} = 2a_i x_{i-1} + b_i \tag{18.33}$$

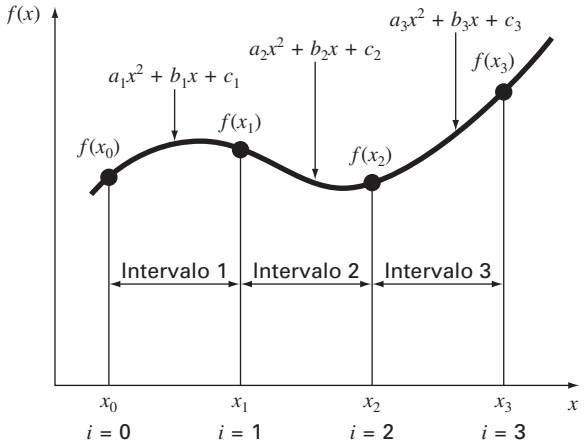


FIGURA 18.17
Notação usada para deduzir splines quadráticos. Observe que há n intervalos e $n + 1$ pontos dados. O exemplo mostrado é para $n = 3$.

para $i = 2$ a n. Isso fornece outras $n - 1$ condições para um total de $2n + n - 1 = 3n - 1$. Como temos $3n$ incógnitas, ainda falta uma condição. A menos que se tenha alguma informação adicional relativa às funções ou suas derivadas, é preciso fazer uma escolha arbitrária para ter sucesso no cálculo das constantes. Embora existam diversas escolhas diferentes possíveis, optou-se pela seguinte:

4. *Suponha que a segunda derivada seja nula no primeiro ponto*. Como a segunda derivada da Equação (18.28) é $2a_i$, essa condição pode ser expressa matematicamente como

$$a_1 = 0 \tag{18.34}$$

A interpretação visual dessa condição é que os primeiros dois pontos serão ligados por uma reta.

EXEMPLO 18.9 Splines Quadráticos

Enunciado do Problema. Ajuste um spline quadrático aos mesmos dados usados no Exemplo 18.8 (Tabela 18.1). Use os resultados para fazer uma estimativa do valor em $x = 5$.

Solução. Nesse problema, há quatro pontos dados e $n = 3$ intervalos. Portanto, $3(3) = 9$ incógnitas devem ser determinadas. As Equações (18.29) e (18.30) fornecem $2(3) - 2 = 4$ condições:

$$\begin{aligned} 20{,}25a_1 + 4{,}5b_1 + c_1 &= 1{,}0 \\ 20{,}25a_2 + 4{,}5b_2 + c_2 &= 1{,}0 \\ 49a_2 + \quad 7b_2 + c_2 &= 2v5 \\ 49a_3 + \quad 7b_3 + c_3 &= 2{,}5 \end{aligned}$$

Fazer com que a primeira e a última função passem pelos valores inicial e final acrescenta mais duas [Equação (18.31)]:

$$9a_1 + 3b_1 + c_1 = 2{,}5$$

e [Equação (18.32)]

$$81a_3 + 9b_3 + c_3 = 0{,}5$$

A continuidade das derivadas cria mais $3 - 1 = 2$ [Equação (18.33)]:

$$9a_1 + b_1 = 9a_2 + b_2$$
$$14a_2 + b_2 = 14a_3 + b_3$$

Finalmente, a Equação (18.34) especifica que $a_1 = 0$. Como essa equação especifica a_1 exatamente, o problema se reduz a resolver oito equações simultâneas. Tais condições podem ser expressas na forma matricial como

$$\begin{bmatrix} 4{,}5 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 20{,}25 & 4{,}5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 49 & 7 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 49 & 7 & 1 \\ 3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 81 & 9 & 1 \\ 1 & 0 & -9 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 14 & 1 & 0 & -14 & -1 & 0 \end{bmatrix} \begin{Bmatrix} b_1 \\ c_1 \\ a_2 \\ b_2 \\ c_2 \\ a_3 \\ b_3 \\ c_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1 \\ 2{,}5 \\ 2{,}5 \\ 2{,}5 \\ 0{,}5 \\ 0 \\ 0 \end{Bmatrix}$$

Essas equações podem ser resolvidas usando-se as técnicas da Parte Três, o que resulta em:

$$a_1 = 0 \qquad b_1 = -1 \qquad c_1 = 5{,}5$$
$$a_2 = 0{,}64 \qquad b_2 = -6{,}76 \qquad c_2 = 18{,}46$$
$$a_3 = -1{,}6 \qquad b_3 = 24{,}6 \qquad c_3 = -91{,}3$$

que, por sua vez, pode ser substituído nas equações quadráticas originais para determinar as seguintes relações para cada intervalo:

$$f_1(x) = -x + 5{,}5 \qquad 3{,}0 \le x \le 4{,}5$$
$$f_2(x) = 0{,}64x^2 - 6{,}76x + 18{,}46 \qquad 4{,}5 \le x \le 7{,}0$$
$$f_3(x) = -1{,}6x^2 + 24{,}6x - 91{,}3 \qquad 7{,}0 \le x \le 9{,}0$$

Quando se usa f_2, a previsão para $x = 5$ é, portanto,

$$f_2(5) = 0{,}64(5)^2 - 6{,}76(5) + 18{,}46 = 0{,}66$$

O ajuste total do spline está mostrado na Figura 18.16*b*. Observe que há duas desvantagens que prejudicam o ajuste: (1) a reta ligando os dois primeiros pontos e (2) o spline para o último intervalo parece ir até muito alto. Os splines cúbicos da próxima seção não exibem essas desvantagens e, como conseqüência, são métodos melhores para interpolação por splines.

18.6.3 Splines Cúbicos

O objetivo nos splines cúbicos é determinar um polinômio de 3º grau para cada intervalo entre os nós, como em

$$f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i \tag{18.35}$$

Logo, para $n + 1$ pontos dados ($i = 0, 1, 2, \ldots, n$), existem n intervalos e, conseqüentemente, $4n$ constantes indeterminadas para calcular. Exatamente como para os splines quadráticos, $4n$ condições são necessárias para calcular as incógnitas. São elas:

1. Os valores da função e dos polinômios adjacentes devem ser iguais nos nós interiores ($2n - 2$ condições).
2. A primeira e a última função devem passar pelos pontos extremos (2 condições).
3. As primeiras derivadas nos nós interiores devem ser iguais ($n - 1$ condições).
4. As segundas derivadas nos nós interiores devem ser iguais ($n - 1$ condições).
5. As segundas derivadas nos nós extremos são nulas (2 condições).

A interpretação visual da condição 5 é que a função se torna uma reta nos nós extremos. A especificação de tais condições nas extremidades leva ao que é chamado de spline "natural", que tem esse nome porque o spline desenhado naturalmente se comporta dessa forma (Figura 18.15). Se o valor da segunda derivada nos nós extremos não for nula (isto é, existe alguma curvatura), essa informação pode ser usada alternativamente para fornecer as duas condições finais.

Esses cinco tipos de condições fornecem o total das $4n$ equações necessárias para determinar os $4n$ coeficientes. Enquanto que certamente é possível desenvolver splines cúbicos dessa forma, vamos apresentar uma técnica que exige a solução de apenas $n - 1$ equações. Embora a dedução desse método (Quadro 18.3) seja um pouco menos simples do que aquela para os splines cúbicos, o ganho em eficiência vale o esforço.

A dedução do Quadro 18.3 resulta na seguinte equação cúbica para cada intervalo:

$$\begin{aligned} f_i(x) = {} & \frac{f_i''(x_{i-1})}{6(x_i - x_{i-1})}(x_i - x)^3 + \frac{f_i''(x_i)}{6(x_i - x_{i-1})}(x - x_{i-1})^3 \\ & + \left[\frac{f(x_{i-1})}{x_i - x_{i-1}} - \frac{f''(x_{i-1})(x_i - x_{i-1})}{6}\right](x_i - x) \\ & + \left[\frac{f(x_i)}{x_i - x_{i-1}} - \frac{f''(x_i)(x_i - x_{i-1})}{6}\right](x - x_{i-1}) \end{aligned} \tag{18.36}$$

Quadro 18.3 Dedução dos Splines Cúbicos

O primeiro passo na dedução (Cheney e Kincaid, 1985) é baseado na observação de que, como cada par de nós é ligado por um polinômio cúbico, a segunda derivada no interior de cada intervalo é uma reta. A Equação (18.35) pode ser derivada duas vezes para verificar essa observação. Com base nisso, as segundas derivadas podem ser representadas por um polinômio interpolador de Lagrange de 1º grau [Equação (18.22)]:

$$f_i''(x) = f_i''(x_{i-1})\frac{x - x_i}{x_{i-1} - x_i} + f_i''(x_i)\frac{x - x_{i-1}}{x_i - x_{i-1}} \tag{Q18.3.1}$$

onde $f_i''(x)$ é o valor da segunda derivada em um ponto qualquer x no i-ésimo intervalo. Logo, essa equação é uma reta ligando a segunda derivada no primeiro nó $f''(x_{i-1})$ com a segunda derivada no segundo nó $f''(x_i)$.

A seguir, a Equação (Q18.3.1) pode ser integrada duas vezes para fornecer uma expressão para $f_i(x)$. Entretanto, essa expressão irá conter duas constantes de integração indeterminadas. Tais constantes podem ser determinadas invocando a condição de igualdade das funções — $f_i(x)$ deve ser igual a $f(x_{i-1})$ em x_{i-1} e $f_i(x)$ deve ser igual a $f_i(x_i)$ em x_i. Fazendo esses cálculos, obtém-se a seguinte equação cúbica:

$$\begin{aligned} f_i(x) = {} & \frac{f_i''(x_{i-1})}{6(x_i - x_{i-1})}(x_i - x)^3 + \frac{f_i''(x_i)}{6(x_i - x_{i-1})}(x - x_{i-1})^3 \\ & + \left[\frac{f(x_{i-1})}{x_i - x_{i-1}} - \frac{f''(x_{i-1})(x_i - x_{i-1})}{6}\right](x_i - x) \\ & + \left[\frac{f(x_i)}{x_i - x_{i-1}} - \frac{f''(x_i)(x_i - x_{i-1})}{6}\right](x - x_{i-1}) \end{aligned} \tag{Q18.3.2}$$

Agora, sem dúvida, essa relação é uma expressão muito mais complicada para o spline cúbico para o i-ésimo intervalo do que, por exemplo, a Equação (18.35). Entretanto, observe que ela contém apenas dois "coeficientes" indeterminados, as segundas derivadas no início e no final do intervalo — $f''(x_{i-1})$ e $f''(x_i)$. Logo, se determinarmos as segundas derivadas adequadas em cada nó, a Equação (Q18.3.2) será um polinômio de 3º grau que pode ser usado para interpolar no intervalo.

As segundas derivadas podem ser calculadas usando-se a condição de que as primeiras derivadas nos nós devem ser contínuas:

$$f_i'(x_i) = f_{i+1}'(x_i) \tag{Q18.3.3}$$

A Equação (Q18.3.2) pode ser derivada para dar uma expressão para a primeira derivada. Se isso for feito para os i-ésimo e $(i - 1)$-ésimo intervalos e se os dois resultados forem igualados de acordo com a Equação (Q18.3.3), resultam as seguintes relações:

$$\begin{aligned} & (x_i - x_{i-1})f''(x_{i-1}) + 2(x_{i+1} - x_{i-1})f''(x_i) \\ & \quad + (x_{i+1} - x_i)f''(x_{i+1}) \\ & \quad = \frac{6}{x_{i+1} - x_i}[f(x_{i+1}) - f(x_i)] \\ & \quad + \frac{6}{x_i - x_{i-1}}[f(x_{i-1}) - f(x_i)] \end{aligned} \tag{Q18.3.4}$$

Se a Equação (Q18.3.4) for escrita para todos os nós interiores, teremos $n - 1$ equações simultâneas com $n + 1$ segundas derivadas. Contudo, como esse é um spline cúbico natural, as segundas derivadas nos nós extremos são nulas e o problema se reduz a $n - 1$ equações com $n - 1$ incógnitas. Além disso, observe que o sistema de equações será tridiagonal. Logo, não apenas reduzimos o número de equações, como também as colocamos em uma forma que é extremamente fácil de resolver (lembre-se da Seção 11.1.1).

Essa equação contém apenas duas incógnitas — as segundas derivadas nas extremidades de cada intervalo. Essas incógnitas podem ser calculadas usando-se a seguinte equação:

$$(x_i - x_{i-1}) f''(x_{i-1}) + 2(x_{i+1} - x_{i-1}) f''(x_i) + (x_{i+1} - x_i) f''(x_{i+1}) = \frac{6}{x_{i+1} - x_i}[f(x_{i+1}) - f(x_i)] + \frac{6}{x_i - x_{i-1}}[f(x_{i-1}) - f(x_i)] \quad (18.37)$$

Se essa equação for escrita para todos os nós interiores, resultarão $n - 1$ equações simultâneas com $n - 1$ incógnitas. (Lembre-se, as segundas derivadas nos nós extremos são nulas.) A aplicação dessas equações é ilustrada no seguinte exemplo.

EXEMPLO 18.10 Splines Cúbicos

Enunciado do Problema. Ajuste splines cúbicos aos mesmos dados usados nos Exemplos 18.8 e 18.9 (Tabela 18.1). Utilize os resultados para fazer uma estimativa do valor em $x = 5$.

Solução. O primeiro passo é usar a Equação (18.37) para gerar um conjunto de equações simultâneas que serão utilizadas para determinar as segundas derivadas nos nós. Por exemplo, para o primeiro nó interior, são usados os seguintes dados:

$$x_0 = 3 \qquad f(x_0) = 2{,}5$$
$$x_1 = 4{,}5 \qquad f(x_1) = 1$$
$$x_2 = 7 \qquad f(x_2) = 2{,}5$$

Esses valores podem ser substituídos na Equação (18.37) para fornecer

$$(4{,}5 - 3)f''(3) + 2(7 - 3)f''(4{,}5) + (7 - 4{,}5)f''(7) = \frac{6}{7 - 4{,}5}(2{,}5 - 1) + \frac{6}{4{,}5 - 3}(2{,}5 - 1)$$

Por causa da condição de spline natural, $f''(3) = 0$, e a equação se reduz a

$$8f''(4{,}5) + 2{,}5f''(7) = 9{,}6$$

De modo análogo, a Equação (18.37) pode ser aplicada ao segundo nó interior para dar

$$2{,}5f''(4{,}5) + 9f''(7) = -9{,}6$$

Essas duas equações podem ser resolvidas simultaneamente por

$$f''(4{,}5) = 1{,}67909$$
$$f''(7) = -1{,}53308$$

Esses valores podem ser substituídos na Equação (18.36), junto com valores para os x's e os $f(x)$'s, para fornecer

$$f_1(x) = \frac{1{,}67909}{6(4{,}5 - 3)}(x - 3)^3 + \frac{2{,}5}{4{,}5 - 3}(4{,}5 - x) + \left[\frac{1}{4{,}5 - 3} - \frac{1{,}67909(4{,}5 - 3)}{6}\right](x - 3)$$

ou

$$f_1(x) = 0{,}186566(x - 3)^3 + 1{,}666667(4{,}5 - x) + 0{,}246894(x - 3)$$

Essa equação é o spline cúbico para o primeiro intervalo. Substituições similares podem ser feitas para determinar equações para o segundo e o terceiro intervalos:

$$f_2(x) = 0{,}111939(7 - x)^3 - 0{,}102205(x - 4{,}5)^3 - 0{,}299621(7 - x) + 1{,}638783(x - 4{,}5)$$

e

$$f_3(x) = -0{,}127757(9 - x)^3 + 1{,}761027(9 - x) + 0{,}25(x - 7)$$

Essas três equações então são usadas para calcular valores em cada intervalo. Por exemplo, o valor em $x = 5$, que está no segundo intervalo, é calculado como sendo

$$f_2(5) = 0{,}111939(7 - 5)^3 - 0{,}102205(5 - 4{,}5)^3 - 0{,}299621(7 - 5) + 1{,}638783(5 - 4{,}5) = 1{,}102886$$

Outros valores foram calculados e os resultados estão mostrados na Figura 18.16*c*.

Os resultados dos Exemplos 18.8 a 18.10 estão resumidos na Figura 18.16. Observe a melhora progressiva no ajuste à medida que nos movemos dos splines lineares para quadráticos para cúbicos. Foi superposto também um polinômio interpolador cúbico na Figura 18.16*c*. Embora o spline cúbico consista em uma série de curvas de terceiro grau, o ajuste resultante é diferente do obtido usando-se um polinômio de terceiro grau, o que se deve ao fato de que o spline natural exige segundas derivadas nulas nos nós extremos, enquanto o polinômio cúbico não tem tal restrição.

FIGURA 18.18
Algoritmo para a interpolação por spline cúbico.

```
SUBROUTINE Spline (x,y,n,xu,yu,dy,d2y)
LOCAL e_n, f_n, g_n, r_n, d2x_n
CALL Tridiag(x,y,n,e,f,g,r)
CALL Decomp(e,f,g,n−1)
CALL Subst(e,f,g,r,n−1,d2x)
CALL Interpol(x,y,n,d2x,xu,yu,dy,d2y)
END Spline

SUBROUTINE Tridiag (x,y,n,e,f,g,r)
f_1 = 2 * (x_2−x_0)
g_1 = (x_2−x_1)
r_1 = 6/(x_2−x_1) * (y_2−y_1)
r_1 = r_1+6/(x_1−x_0) * (y_0−y_1)
DOFOR i = 2, n−2
  e_i = (x_i−x_i−1)
  f_i = 2 * (x_i+1 − x_i−1)
  g_i = (x_i+1 − x_i)
  r_i = 6/(x_i+1 − x_i) * (y_i+1 − y_i)
  r_i = r_i+6/(x_i − x_i−1) * (y_i−1 − y_i)
END DO
e_n−1 = (x_n−1 − x_n−2)
f_n−1 = 2 * (x_n − x_n−2)
r_n−1 = 6/(x_n − x_n−1) * (y_n − y_n−1)
r_n−1 = r_n−1 + 6/(x_n−1 − x_n−2) * (y_n−2 − y_n−1)
END Tridiag

SUBROUTINE Interpol (x,y,n,d2x,xu,yu,dy,d2y)
flag = 0
i = 1
DO
  IF xu ≥ x_i−1 AND xu ≤ x_i THEN
    c1 = d2x_i−1/6/(x_i − x_i−1)
    c2 = d2x_i/6/(x_i − x_i−1)
    c3 = y_i−1/(x_i − x_i−1) − d2x_i−1 * (x_i−x_i−1)/6
    c4 = y_i/(x_i − x_i−1) − d2x_i * (x_i−x_i−1)/6
    t1 = c1 * (x_i − xu)^3
    t2 = c2 * (xu − x_i−1)^3
    t3 = c3 * (x_i − xu)
    t4 = c4 * (xu − x_i−1)
    yu = t1 + t2 + t3 + t4
    t1 = −3 * c1 * (x_i − xu)^2
    t2 = 3 * c2 * (xu − x_i−1)^2
    t3 = −c3
    t4 = c4
    dy = t1 + t2 + t3 + t4
    t1 = 6 * c1 * (x_i − xu)
    t2 = 6 * c2 * (xu − x_i−1)
    d2y = t1 + t2
    flag = 1
  ELSE
    i = i + 1
  END IF
  IF i = n + 1 OR flag = 1 EXIT
END DO
IF flag = 0 THEN
  PRINT "fora do intervalo"
  pause
END IF
END Interpol
```

18.6.4 Algoritmo Computacional para Splines Cúbicos

O método para calcular splines cúbicos delineados na seção anterior é ideal para implementação computacional. Lembre-se de que, por algumas manipulações engenhosas, o método reduziu o problema à resolução de $n - 1$ equações simultâneas. Um benefício adicional da dedução é que, como especificado pela Equação (18.37), o sistema de equações é tridiagonal. Como descrito na Seção 11.1, estão disponíveis algoritmos para resolver tais sistemas de maneira extremamente eficiente. A Figura 18.18 esboça um esquema computacional que incorpora essas características.

Observe que a rotina na Figura 18.18 retorna um único valor interpolado, *yu*, para um dado valor da variável independente *xu*. Essa é apenas uma das formas como a interpolação por splines pode ser implementada. Por exemplo, você poderia querer determinar os coeficientes uma única vez e então fazer muitas interpolações. Além disso, a rotina retorna tanto a primeira (*dy*) quanto a segunda (*dy*2) derivadas em *xu*. Embora não seja necessário calcular essas quantidades, elas se mostram úteis em muitas aplicações de interpolação por splines.

PROBLEMAS

18.1 Faça uma estimativa do logaritmo comum de 10 usando interpolação linear.

(a) Interpole entre log 8 = 0,9030900 e log 12 = 1,0791812.

(b) Interpole entre log 9 = 0,9542425 e log 11 = 1,0413927. Para cada interpolação, calcule o erro relativo porcentual baseado no valor verdadeiro.

18.2 Ajuste um polinômio interpolador de Newton de segundo grau para fazer uma estimativa de log 10 usando os dados do Problema 18.1 em $x = 8$, 9 e 11. Calcule o erro relativo porcentual verdadeiro.

18.3 Ajuste um polinômio interpolador de Newton de terceiro grau usando os dados do Problema 18.1.

18.4 Considere os dados

x	1,6	2	2,5	3,2	4	4,5
$f(x)$	2	8	14	15	8	2

(a) Calcule $f(2.8)$ usando polinômios interpoladores de Newton de primeiro a terceiro graus. Escolha a seqüência de pontos para fazer sua estimativa de modo a atingir a melhor acurácia possível.

(b) Utilize a Equação (18.18) para fazer uma estimativa do erro em cada previsão.

18.5 Considere os dados

x	1	2	3	5	7	8
$f(x)$	3	6	19	99	291	444

Calcule $f(4)$ usando polinômios interpoladores de Newton de primeiro a quarto graus. Escolha seus pontos base para obter uma boa acurácia. O que seus resultados indicam em relação ao grau do polinômio usado para gerar os dados da tabela?

18.6 Repita os Problemas 18.1 a 18.3 usando polinômios Lagrange.

18.7 Repita o Problema 18.5 usando polinômios de Lagrange de primeiro a terceiro graus.

18.8 Use interpolação inversa por um polinômio interpolador cúbico e bissecção para determinar o valor de x que corresponde a $f(x) = 0{,}23$ para os seguintes dados tabulados:

x	2	3	4	5	6	7
$f(x)$	0,5	0,3333	0,25	0,2	0,1667	0,1429

18.9 Use interpolação inversa para determinar o valor de x que corresponde a $f(x) = 0{,}85$ para os seguintes dados tabulados:

x	0	1	2	3	4	5
$f(x)$	0	0,5	0,8	0,9	0,941176	0,961538

Observe que os valores na tabela foram gerados pela função $f(x) = x^3/(2 + x^3)$.

(a) Determine o valor correto analiticamente.

(b) Use interpolação cúbica de x em função de y.

(c) Use interpolação inversa com interpolação quadrática e a fórmula quadrática.

(d) Use interpolação inversa com interpolação cúbica e bissecção. Para as partes (**b**) a (**d**), calcule o erro relativo porcentual verdadeiro.

18.10 Determine splines quadráticos para os primeiros cinco pontos dados no Problema 18.4 e faça previsões de $f(3{,}4)$ e $f(2{,}2)$.

18.11 Determine splines cúbicos para os dados do Problema 18.5 e (**a**) faça previsões de $f(4)$ e $f(2{,}5)$ e (**b**) verifique que $f_2(3)$ e $f_3(3) = 19$.

18.12 Determine os coeficientes da parábola que passa pelos últimos três pontos do Problema 18.4.

18.13 Determine os coeficientes da equação cúbica que passa pelos primeiros quatro pontos no Problema 18.5.

18.14 Desenvolva, debug e teste um programa ou em uma linguagem de alto nível ou em uma macrolinguagem de sua escolha para implementar o polinômio interpolador de Newton com base na Figura 18.7.

18.15 Teste o programa que você desenvolveu no Problema 18.14 repetindo os cálculos do Exemplo 18.5.

18.16 Use o programa que você desenvolveu no Problema 18.14 para resolver os Problemas 18.1 a 18.3.

18.17 Use o programa que você desenvolveu no Problema 18.14 para resolver os Problemas 18.4 e 18.5. No Problema 18.4, use todos os dados para desenvolver polinômios de primeiro a quinto graus. Em ambos os problemas, trace a estimativa de erro em função do grau.

18.18 Desenvolva, debug e teste um programa ou em uma linguagem de alto nível ou em uma macrolinguagem de sua escolha para implementar a interpolação de Lagrange. Use como base o pseudocódigo da Figura 18.11. Teste-o repetindo o Exemplo 18.7.

18.19 Uma aplicação útil da interpolação de Lagrange é chamada *tabela de consulta.* Como o nome indica, ela envolve "consultar" um valor intermediário em uma tabela. Para desenvolver tal algoritmo, a tabela dos valores de x e de $f(x)$ é inicialmente armazenada em um par de matrizes unidimensionais. Esses valores são então passados para uma função junto com o valor de x que você quer calcular. A função então realiza duas tarefas. Primeiro, ela percorre a tabela até encontrar o intervalo no qual a incógnita está. Então, aplica uma técnica como a interpolação de Lagrange para determinar o valor apropriado de $f(x)$. Desenvolva tal função usando um polinômio de Lagrange cúbico para fazer a interpolação. Para intervalos intermediários, essa é uma boa escolha, pois a incógnita estará localizada no intervalo no meio dos quatro pontos necessários para gerar o polinômio cúbico. Para o primeiro e o último intervalos, use um polinômio de Lagrange quadrático. Faça também com que o código detecte quando o usuário pede um valor fora do intervalo dos x's. Em tais casos, a função deveria mostrar uma mensagem de erro. Teste seu programa para $f(x) = \ln x$ usando dados entre $x = 0, 1, 2, \ldots, 10$.

18.20 Desenvolva, debug e teste um programa ou em uma linguagem de alto nível ou em uma macrolinguagem de sua escolha para implementar a interpolação por splines cúbicos com base na Figura 18.18. Teste seu programa repetindo o Exemplo 18.10.

18.21 Use o software desenvolvido no Problema 18.20 para ajustar splines cúbicos aos dados dos Problemas 18.4 e 18.5. Em ambos os casos, faça uma previsão de $f(2{,}25)$.

18.22 Use a parte da tabela de vapor dada para H_2O superaquecida a 200 MPa para (**a**) encontrar a entropia correspondente s para um volume específico v de 0,108 m^3/kg com interpolação linear, (**b**) encontrar a mesma entropia correspondente usando interpolação quadrática, e (**c**) encontrar o volume correspondente a uma entropia de 6,6 usando interpolação inversa.

v (m^3/kg)	0,10377	0,11144	0,1254
s (kJ/kg·K)	6,4147	6,5453	6,7664

CAPÍTULO 19

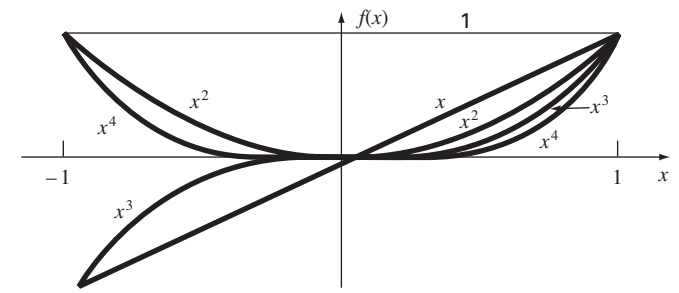
Aproximação de Fourier

Até este ponto, nossa apresentação de interpolação enfatizou polinômios padrão — ou seja, combinações lineares dos monômios 1, x, x^2, . . . , x^m (Figura 19.1*a*). Agora, nos voltamo-nos para uma outra classe de funções que têm imensa importância na engenharia — as funções trigonométricas 1, cos x, cos $2x$, . . . , cos nx, sen x, sen $2x$, . . . , sen nx (Figura 19.1*b*).

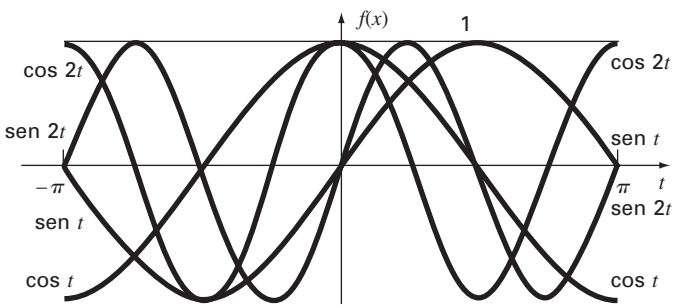
Os engenheiros em geral tratam de sistemas que oscilam ou vibram. Como pode ser esperado, as funções trigonométricas desempenham um papel fundamental na modelagem de tais contextos de problemas. A *aproximação de Fourier* representa um esquema sistemático para usar séries trigonométricas com esse propósito.

Uma das características da análise de Fourier é que ela lida tanto com o domínio de freqüência quanto com o de tempo. Como alguns engenheiros não se sentem confortáveis com o último, uma grande parte do material subseqüente foi dedicada a uma visão geral da aproximação de Fourier. Um aspecto importante dessa visão geral será familiarizá-lo com o domínio de freqüência. Essa orientação é então seguida por uma introdução aos métodos numéricos para o cálculo de transformadas de Fourier discretas.

FIGURA 19.1
Os primeiros cinco (*a*) monômios e (*b*) funções trigonométricas. Observe que, no intervalo mostrado, ambos os tipos de funções assumem valores que variam entre −1 e 1. Entretanto, os valores de pico dos monômios todos ocorrem nas extremidades, enquanto que, para as funções trigonométricas, os picos são distribuídos mais uniformemente pelo intervalo.



(*a*)



(*b*)

19.1 AJUSTE DE CURVAS COM FUNÇÕES SENOIDAIS

Uma função periódica $f(t)$ é uma para a qual

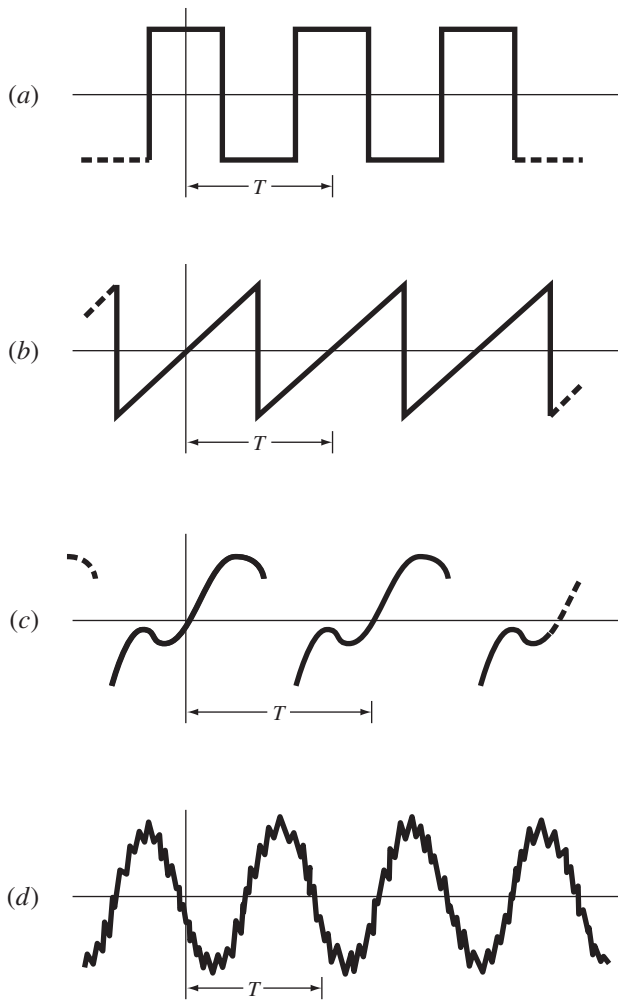
$$f(t) = f(t + T) \tag{19.1}$$

onde T é uma constante chamada *período*, que é o menor valor para o qual a Equação (19.1) vale. Exemplos comuns incluem formas ondulatórias como ondas quadradas e dente de serra (Figura 19.2). As mais fundamentais são as funções senoidais.

Na presente discussão, será usado o termo *senóide* para representar qualquer forma ondulatória que possa ser descrita por um seno ou um cosseno. Não existe nenhuma convenção bem estabelecida para escolher qualquer uma das duas funções, e, de todo modo, os resultados seriam idênticos. Neste capítulo, será usado o cosseno, que pode ser representado, em geral, por

$$f(t) = A_0 + C_1 \cos(\omega_0 t + \theta) \tag{19.2}$$

FIGURA 19.2
Além das funções trigonométricas como o seno e o cosseno, as funções periódicas incluem formas ondulatórias como (*a*) a onda quadrada e (*b*) onda dente de serra. Além dessas formas idealizadas, sinais periódicos por natureza podem ser (*c*) não-ideais e (*d*) contaminados por ruídos. As funções trigonométricas podem ser usadas para representar e analisar todos esses casos.



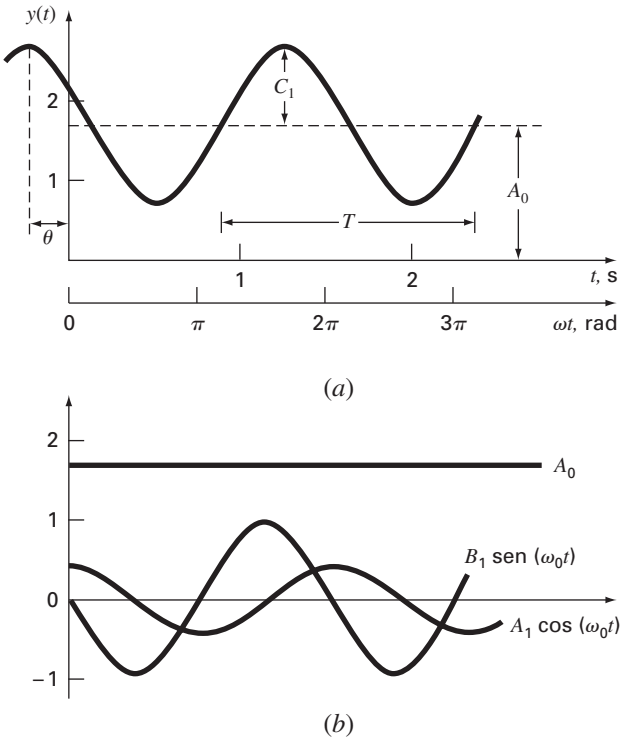


FIGURA 19.3
(*a*) Um gráfico da função senoidal $y(t) = A_0 + C_1 \cos(\omega_0 t + \theta)$. Nesse caso, $A_0 = 1{,}7$, $C_1 = 1$, $\omega_0 = 2\pi/T = 2\pi/(1{,}5\text{ s})$ e $\theta = \pi/3$ radianos = 1,0472 (= 0,25 s). Os outros parâmetros usados para descrever a curva são a freqüência $f = \omega_0/(2\pi)$, que nesse caso é 1 ciclo/(1,5 s) e o período $T = 1{,}5$ s. (*b*) Uma expressão alternativa da mesma curva é $y(t) = A_0 + A_1 \cos(\omega_0 t) + B_1 \operatorname{sen}(\omega_0 t)$. As três componentes dessa função são mostradas em (*b*), onde $A_1 = 0{,}5$ e $B_1 = -0{,}866$. A soma das três curvas em (*b*) fornece a curva única em (*a*).

Logo, quatro parâmetros servem para caracterizar a senóide (Figura 19.3). O *valor médio* A_0 determina a altura média acima da abscissa. A *amplitude* C_1 especifica a altura da oscilação. A *freqüência angular* ω_0 caracteriza com que freqüência o ciclo ocorre. Finalmente, o ângulo de fase, ou *deslocamento angular*, θ parametriza a extensão pela qual a curva senoidal está deslocada horizontalmente. Ele pode ser medido como a distância em radianos de $t = 0$ ao ponto no qual a função cosseno começa um novo ciclo. Como descrito na Figura 19.4*a*, um valor negativo é chamado de *ângulo de fase de retardamento* porque a curva $\cos(\omega_0 t - \theta)$ começa um novo ciclo θ radianos depois de $\cos(\omega_0 t)$. Logo, diz-se que $\cos(\omega_0 t - \theta)$ tem um retardamento em relação a $\cos(\omega_0 t)$. Reciprocamente, como na Figura 19.4*b*, um valor positivo é chamado de ângulo de fase de avanço.

Observe que a freqüência angular (em radianos/tempo) está relacionada com a freqüência f (em ciclos/tempo) por

$$\omega_0 = 2\pi f \tag{19.3}$$

e a freqüência, por sua vez, está relacionada ao período T (em unidades de tempo) por

$$f = \frac{1}{T} \tag{19.4}$$

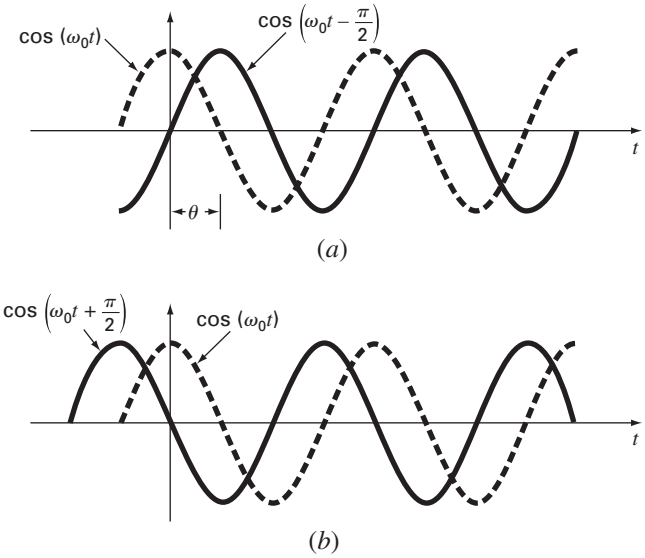


FIGURA 19.4
Descrição gráfica de ângulo de fase de (*a*) retardamento e (*b*) avanço. Observe que a curva com retardamento em (*a*) pode ser descrita alternativamente como $\cos(\omega_0 t + 3\pi/2)$. Em outras palavras, se uma curva tem um retardamento por um ângulo α, ela também pode ser representada como tendo um ângulo de avanço $2\pi - \alpha$.

Embora a Equação (19.2) seja uma caracterização matematicamente adequada de uma função senoidal, é complicado trabalhar com ela do ponto de vista de ajuste de curvas por causa do deslocamento angular incluído no argumento da função cosseno. Essa deficiência pode ser superada com o uso da identidade trigonométrica

$$C_1 \cos(\omega_0 t + \theta) = C_1[\cos(\omega_0 t)\cos(\theta) - \text{sen}\ (\omega_0 \text{t})\text{sen}\ (\theta)]] \tag{19.5}$$

Substituindo a Equação (19.5) na Equação (19.2) e agrupando os termos, obtém-se (Figura (19.3*b*)

$$f(t) = A_0 + A_1 \cos(\omega_0 t) + B_1 \text{sen}\ (\omega_0 \text{t}) \tag{19.6}$$

onde

$$A_1 = C_1 \cos(\theta) \qquad B_1 = -C_1 \text{sen}\ (\theta) \tag{19.7}$$

Dividindo-se as duas partes da Equação (19.7), obtém-se

$$\theta = \text{arctg}\left(-\frac{B_1}{A_1}\right) \tag{19.8}$$

onde, se $A_1 < 0$, some π a θ. Elevando ao quadrado e somando, a Equação (19.7) leva a

$$C_1 = \sqrt{A_1^2 + B_1^2} \tag{19.9}$$

Logo, a Equação (19.6) representa uma formulação alternativa da Equação (19.2) que ainda usa quatro parâmetros, mas que está na forma de um modelo linear geral [lembre-se da Equação (17.23)]. Como se discutirá na próxima seção, ela pode ser usada de modo simples como a base para um ajuste por mínimos quadrados.

Antes de prosseguir para a próxima seção, entretanto, deve-se enfatizar que poderíamos ter usado uma função seno em vez do cosseno no modelo fundamental da Equação (19.2). Por exemplo,

$$f(t) = A_0 + C_1 \text{sen}\ (\omega_0 \text{t} + \delta)$$

poderia ter sido usada. Relações simples podem ser aplicadas para converter entre as duas formas

$$\text{sen}\,(\omega_0 t + \delta) = \cos\left(\omega_0 t + \delta - \frac{\pi}{2}\right)$$

e

$$\cos(\omega_0 t + \theta) = \text{sen}\left(\omega_0 t + \theta + \frac{\pi}{2}\right) \tag{19.10}$$

Em outras palavras, $\theta = \delta - \pi/2$. A única consideração importante é que uma ou outra forma deve ser usada consistentemente. Assim, será usada a versão cosseno em toda a discussão.

19.1.1 Ajuste por Mínimos Quadrados de uma Curva Senoidal

A Equação (19.6) pode ser pensada como um modelo linear de mínimos quadrados

$$y = A_0 + A_1 \cos(\omega_0 t) + B_1 \text{sen}\,(\omega_0 t) + e \tag{19.11}$$

o qual é simplesmente um outro exemplo do modelo geral [lembre-se da Equação (17.23)]

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots + a_m z_m + e \tag{17.23}$$

onde $z_0 = 1$, $z_1 = \cos(\omega_0 t)$, $z_2 = \text{sen}(\omega_0 t)$ e todos os outros z's são nulos. Logo, o objetivo é determinar valores dos coeficientes que minimizem

$$S_r = \sum_{i=1}^{N} \left\{ y_i - [A_0 + A_1 \cos(\omega_0 t_i) + B_1 \text{sen}\,(\omega_0 t_i)] \right\}^2$$

As equações normais para conseguir essa minimização podem ser expressas na forma matricial como [lembre-se da Equação (17.25)]

$$\begin{bmatrix} N & \Sigma \cos(\omega_0 t) & \Sigma \text{sen}\,(\omega_0 t) \\ \Sigma \cos(\omega_0 t) & \Sigma \cos^2(\omega_0 t) & \Sigma \cos(\omega_0 t)\text{sen}\,(\omega_0 t) \\ \Sigma \text{sen}\,(\omega_0 t) & \Sigma \cos(\omega_0 t)\text{sen}\,(\omega_0 t) & \Sigma \text{sen}^2(\omega_0 t) \end{bmatrix} \begin{Bmatrix} A_0 \\ A_1 \\ B_1 \end{Bmatrix} = \begin{Bmatrix} \Sigma y \\ \Sigma y \cos(\omega_0 t) \\ \Sigma y\ \text{sen}\,(\omega_0 t) \end{Bmatrix} \tag{19.12}$$

Essas equações podem ser usadas para determinar os coeficientes desconhecidos. Entretanto, em vez de fazer isso, pode-se examinar o caso especial em que existem N observações igualmente espaçadas em intervalos Δt e com um intervalo total de $T = (N - 1)\ \Delta t$. Nessa situação, podem ser determinados os seguintes valores médios (ver Problema 19.3):

$$\frac{\Sigma \text{sen}\,(\omega_0 t)}{N} = 0 \qquad \frac{\Sigma \cos(\omega_0 t)}{N} = 0$$

$$\frac{\Sigma \text{sen}^2(\omega_0 t)}{N} = \frac{1}{2} \qquad \frac{\Sigma \cos^2(\omega_0 t)}{N} = \frac{1}{2} \tag{19.13}$$

$$\frac{\Sigma \cos(\omega_0 t)\text{sen}\,(\omega_0 t)}{N} = 0$$

Então, para pontos uniformemente espaçados, as equações normais se tornam

$$\begin{bmatrix} N & 0 & 0 \\ 0 & N/2 & 0 \\ 0 & 0 & N/2 \end{bmatrix} \begin{Bmatrix} A_0 \\ A_1 \\ B_1 \end{Bmatrix} = \begin{Bmatrix} \Sigma y \\ \Sigma y \cos(\omega_0 t) \\ \Sigma y\ \text{sen}(\omega_0 t) \end{Bmatrix}$$

A inversa da matriz diagonal é simplesmente uma outra matriz diagonal cujos elementos são os inversos da matriz original. Logo, os coeficientes podem ser determinados por

$$\begin{Bmatrix} A_0 \\ A_1 \\ B_1 \end{Bmatrix} = \begin{bmatrix} 1/N & 0 & 0 \\ 0 & 2/N & 0 \\ 0 & 0 & 2/N \end{bmatrix} \begin{Bmatrix} \Sigma y \\ \Sigma y \cos(\omega_0 t) \\ \Sigma y \text{ sen } (\omega_0 \text{t}) \end{Bmatrix}$$

ou

$$A_0 = \frac{\Sigma y}{N} \tag{19.14}$$

$$A_1 = \frac{2}{N} \Sigma y \cos(\omega_0 t) \tag{19.15}$$

$$B_1 = \frac{2}{N} \Sigma y \text{ sen } (\omega_0 \text{t}) \tag{19.16}$$

EXEMPLO 19.1

Ajuste por Mínimos Quadrados de uma Curva Senoidal

Enunciado do Problema. A curva na Figura 19.3 é descrita por $y = 1{,}7 + \cos(4{,}189t + 1{,}0472)$. Gere 10 valores discretos para essa curva em intervalos de $\Delta t = 0{,}15$ para uma variação de $t = 0$ a 1,35. Use essa informação para calcular os coeficientes da Equação (19.11) por um ajuste por mínimos quadrados.

Solução. Os dados necessários para o cálculo dos coeficientes com $\omega = 4{,}189$ são

t	y	$y \cos(\omega_0 t)$	$y \text{ sen}(\omega_0 t)$
0	2,200	2,200	0,000
0,15	1,595	1,291	0,938
0,30	1,031	0,319	0,980
0,45	0,722	−0,223	0,687
0,60	0,786	−0,636	0,462
0,75	1,200	−1,200	0,000
0,90	1,805	−1,460	−1,061
1,05	2,369	−0,732	−2,253
1,20	2,678	0,829	−2,547
1,35	2,614	2,114	−1,536
$\Sigma =$	17,000	2,502	−4,330

Esses resultados podem ser usados para determinar [Equações (19.14) a (19.16)]

$$A_0 = \frac{17{,}000}{10} = 1{,}7 \qquad A_1 = \frac{2}{10} 2{,}502 = 0{,}500 \qquad B_1 = \frac{2}{10}(-4{,}330) = -0{,}866$$

Logo, o ajuste por mínimos quadrados é

$$y = 1{,}7 + 0{,}500 \cos(\omega_0 t) - 0{,}866 \text{sen } (\omega_0 \text{t})$$

O modelo também pode ser expresso na forma da Equação (19.2) calculando-se [Equação (19.8)]

$$\theta = \text{arctg} \left(-\frac{-0{,}866}{0{,}500} \right) = 1{,}0472$$

e [Equação (19.9)]

$$C_1 = \sqrt{(0{,}5)^2 + (-0{,}866)^2} = 1{,}00$$

para obter

$$y = 1{,}7 + \cos(\omega_0 t + 1{,}0472)$$

ou, alternativamente, como uma função seno usando-se a Equação (19.10)

$$y = 1{,}7 + \text{sen}\ (\omega_0 t + 2{,}618)$$

A análise anterior pode ser estendida ao modelo geral

$$f(t) = A_0 + A_1 \cos(\omega_0 t) + B_1\ \text{sen}\ (\omega_0 \text{t}) + \text{A}_2 \cos(2\omega_0 \text{t}) + \text{B}_2\ \text{sen}\ (2\omega_0 \text{t}) + \cdots + A_m \cos(m\omega_0 t) + B_m\ \text{sen}\ (\text{m}\omega_0 \text{t})$$

onde, para dados igualmente espaçados, os coeficientes podem ser calculados por

$$A_0 = \frac{\Sigma y}{N}$$

$$\left.\begin{aligned} A_j &= \frac{2}{N}\Sigma y \cos(j\omega_0 t) \\ B_j &= \frac{2}{N}\Sigma y\ \text{sen}\ (\text{j}\omega_0 \text{t}) \end{aligned}\right\} \qquad j = 1, 2, \ldots, m$$

Embora essas relações possam ser usadas para ajustar dados no sentido de regressão (isto é, $N > 2m + 1$), uma aplicação alternativa seria usá-las para interpolação ou colocação — ou seja, usá-las para o caso em que o número de incógnitas, $2m + 1$, é igual ao número de pontos dados, N. Essa é a abordagem utilizada na série de Fourier contínua, como descrito a seguir.

19.2 SÉRIE DE FOURIER NO CONTÍNUO

No decorrer do estudo de problemas de fluxo de calor, Fourier mostrou que uma função periódica arbitrária pode ser representada por uma série infinita de funções senoidais com freqüências harmonicamente relacionadas. Para uma função de período T, uma série de Fourier no contínuo pode ser escrita[1]

$$f(t) = a_0 + a_1 \cos(\omega_0 t) + b_1 \text{sen}\ (\omega_0 \text{t}) + \text{a}_2 \cos(2\omega_0 \text{t}) + \text{b}_2 \text{sen}\ (2\omega_0 \text{t}) + \cdots$$

ou, mais concisamente,

$$f(t) = a_0 + \sum_{k=1}^{\infty} [a_k \cos(k\omega_0 t) + b_k \text{sen}\ (\text{k}\omega_0 \text{t})] \tag{19.17}$$

onde $\omega_0 = 2\pi/T$ é chamada *freqüência fundamental* e os seus múltiplos constantes $2\omega_0$, $3\omega_0$ etc. são chamados *harmônicos*. Logo, a Equação (19.17) expressa $f(t)$ como uma combinação linear das funções base: 1, $\cos(\omega_0 t)$, $\text{sen}(\omega_0 t)$, $\cos(2\omega_0 t)$, $\text{sen}(2\omega_0 t)$, ...

Como descrito no Quadro 19.1, os coeficientes da Equação (19.17) podem ser calculados por

$$a_k = \frac{2}{T}\int_0^T f(t) \cos(k\omega_0 t)\, dt \tag{19.18}$$

e

$$b_k = \frac{2}{T}\int_0^T f(t) \text{sen}\ (\text{k}\omega_0 \text{t})\ \text{dt} \tag{19.19}$$

para $k = 1, 2, \ldots$ e

$$a_0 = \frac{1}{T}\int_0^T f(t)\, dt \tag{19.20}$$

[1] A existência da série de Fourier é garantida nas condições de Dirichlet, as quais especificam que a função periódica tenha um número finito de máximos e de mínimos e que exista um número finito de descontinuidades do tipo salto. Em geral, todas as funções periódicas que aparecem fisicamente satisfazem essas condições.

Quadro 19.1 Determinação dos Coeficientes da Série de Fourier no Contínuo

Como foi feito para os dados discretos da Seção 19.1.1, as seguintes relações podem ser estabelecidas:

$$\int_0^T \text{sen}\,(\text{k}\omega_0\text{t})\,\text{dt} = \int_0^T \cos(k\omega_0 t)\,dt = 0 \qquad \text{(Q19.1.1)}$$

$$\int_0^T \cos(k\omega_0 t)\text{sen}\,(\text{g}\omega_0\text{t})\,\text{dt} = 0 \qquad \text{(Q19.1.2)}$$

$$\int_0^T \text{sen}\,(\text{k}\omega_0\text{t})\text{sen}\,(\text{g}\omega_0\text{t})\,\text{dt} = 0 \qquad \text{(Q19.1.3)}$$

$$\int_0^T \cos(k\omega_0 t)\cos(g\omega_0 t)\,dt = 0 \qquad \text{(Q19.1.4)}$$

$$\int_0^T \text{sen}^2(\text{k}\omega_0\text{t})\,\text{dt} = \int_0^T \cos^2(k\omega_0 t)\,dt = \frac{T}{2} \qquad \text{(Q19.1.5)}$$

Para calcular seus coeficientes, cada lado da Equação (19.17) pode ser integrado para fornecer

$$\int_0^T f(t)\,dt = \int_0^T a_0\,dt + \int_0^T \sum_{k=1}^{\infty}[a_k \cos(k\omega_0 t) + b_k \text{sen}(\text{k}\omega_0\text{t})]\,\text{dt}$$

Como todos os termos da somatória são da forma da Equação (Q19.1.1), a equação se torna

$$\int_0^T f(t)\,dt = a_0 T$$

a qual pode ser resolvida por

$$a_0 = \frac{\int_0^T f(t)\,dt}{T}$$

Logo, a_0 é simplesmente o valor médio da função sobre o período.

Para calcular um dos coeficientes do cosseno, por exemplo, a_m, a Equação (19.17) pode ser multiplicada por $\cos(m\omega_0 t)$ e integrada para fornecer

$$\int_0^T f(t)\cos(m\omega_0 t)\,dt = \int_0^T a_0 \cos(m\omega_0 t)\,dt + \int_0^T \sum_{k=1}^{\infty} a_k \cos(k\omega_0 t)\cos(m\omega_0 t)\,dt + \int_0^T \sum_{k=1}^{\infty} b_k \text{sen}\,(\text{k}\omega_0\text{t})\cos(\text{m}\omega_0\text{t})\,dt \qquad \text{(Q19.1.6)}$$

Das Equações (Q19.1.1), (Q19.1.2) e (Q19.1.4), vemos que todos os termos do lado direito são nulos, com exceção do caso $k = m$. Esse último caso pode ser calculado pela Equação (Q19.1.5) e, portanto, a Equação (Q19.1.6) pode ser usada para determinar a_m, ou, mais geralmente [Equação (19.18)],

$$a_k = \frac{2}{T}\int_0^T f(t)\cos(k\omega_0 t)\,dt$$

para $k = 1, 2, \ldots$

De modo análogo, a Equação (19.17) pode ser multiplicada por $\text{sen}(m\omega_0 t)$, integrada e manipulada para fornecer a Equação (19.19).

EXEMPLO 19.2

Aproximação por Série de Fourier no Contínuo

Enunciado do Problema. Use a série de Fourier no contínuo para aproximar a função de onda quadrada ou retangular (Figura 19.5)

$$f(t) = \begin{cases} -1 & -T/2 < t < -T/4 \\ 1 & -T/4 < t < T/4 \\ -1 & T/4 < t < T/2 \end{cases}$$

Solução. Como a altura média da onda é zero, o valor de $a_0 = 0$ pode ser obtido diretamente. Os coeficientes restantes podem ser calculados por [Equação (19.18)]

$$a_k = \frac{2}{T}\int_{-T/2}^{T/2} f(t)\cos(k\omega_0 t)\,dt$$

$$= \frac{2}{T}\left[-\int_{-T/2}^{-T/4}\cos(k\omega_0 t)\,dt + \int_{-T/4}^{T/4}\cos(k\omega_0 t)\,dt - \int_{T/4}^{T/2}\cos(k\omega_0 t)\,dt\right]$$

As integrais podem ser calculadas, fornecendo

$$a_k = \begin{cases} 4/(k\pi) & \text{para } k = 1, 5, 9, \ldots \\ -4/(k\pi) & \text{para } k = 3, 7, 11, \ldots \\ 0 & \text{para } k = \text{inteiros pares} \end{cases}$$

Analogamente, pode ser determinado que todos os b's são nulos. Portanto, a aproximação por série de Fourier é

$$f(t) = \frac{4}{\pi}\cos(\omega_0 t) - \frac{4}{3\pi}\cos(3\omega_0 t) + \frac{4}{5\pi}\cos(5\omega_0 t) - \frac{4}{7\pi}\cos(7\omega_0 t) + \cdots$$

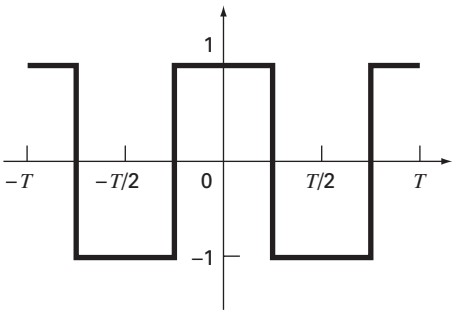
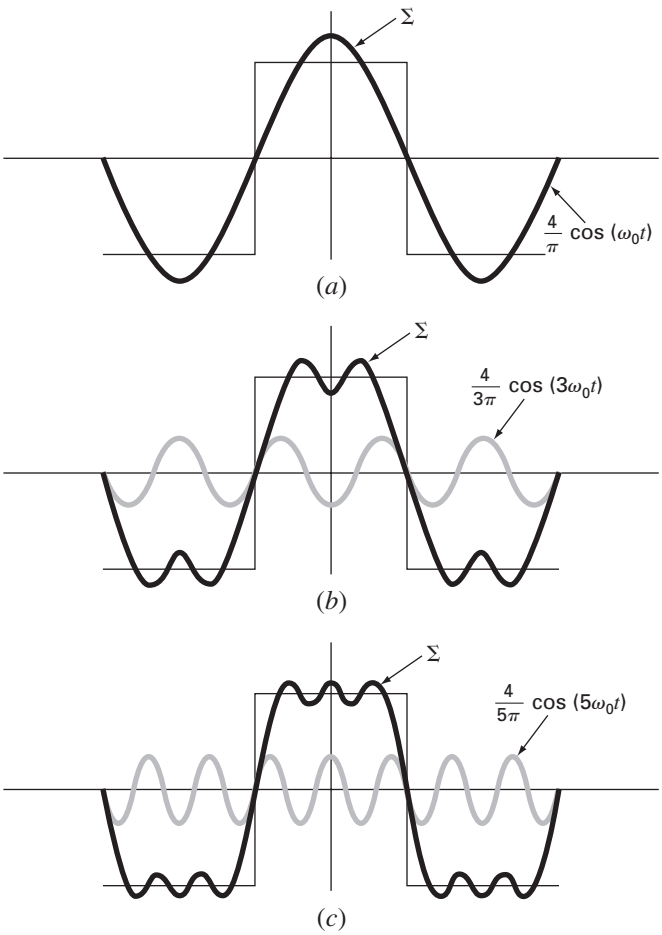


FIGURA 19.5
Uma onda de forma quadrada ou retangular de altura 2 e período $T = 2\pi/\omega_0$.

FIGURA 19.6
A aproximação por série de Fourier da onda quadrada da Figura 19.5. A série de gráficos mostra a somatória até, e incluindo, o (*a*) primeiro, (*b*) segundo e (*c*) terceiro termo. Os termos individuais que foram somados em cada estágio também são mostrados.



Os resultados até os três primeiros termos são mostrados na Figura 19.6.

Deve ser mencionado que a onda quadrada na Figura 19.5 é chamada uma função *par* porque $f(t) = f(-t)$. Outro exemplo de uma função par é cos(t). Pode ser mostrado (Van Valkenburg, 1974) que os b's em uma série de Fourier sempre são iguais a zero para funções pares. Observe também que *funções ímpares* são aquelas para as quais $f(t) = -f(-t)$. A função sen(t) é uma função ímpar. Nesse caso, os a's serão iguais a zero.

Além da forma trigonométrica da Equação (19.17), a série de Fourier pode ser expressa em termos das funções exponenciais como (ver Quadro 19.2 e Apêndice A)

$$f(t) = \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{ik\omega_0 t} \tag{19.21}$$

onde $i = \sqrt{-1}$ e

$$\tilde{c}_k = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{-ik\omega_0 t}\,dt \tag{19.22}$$

Essa formulação alternativa terá utilidade em todo o restante do capítulo.

19.3 DOMÍNIOS DE FREQÜÊNCIA E TEMPO

Até este ponto, a discussão da aproximação de Fourier se limitou ao *domínio do tempo*. Isso foi feito porque a maioria de nós fica bastante confortável idealizando o comportamento de uma função nessa dimensão. Embora não seja tão familiar, o *domínio de freqüência* fornece uma perspectiva alternativa para caracterizar o comportamento de funções oscilantes.

Logo, do mesmo modo como a amplitude pode ser traçada em função do tempo, também pode ser traçada em função da freqüência. Ambos os tipos de expressão estão

Quadro 19.2 Forma Complexa das Séries de Fourier

A forma trigonométrica da série de Fourier contínua é

$$f(t) = a_0 + \sum_{k=1}^{\infty}[a_k \cos(k\omega_0 t) + b_k \text{sen}(\text{k}\omega_0\text{t})] \tag{Q19.2.1}$$

Da identidade de Euler, o seno e o cosseno podem ser expressos na forma exponencial como

$$\text{sen } x = \frac{e^{ix} - e^{-ix}}{2i} \tag{Q19.2.2}$$

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \tag{Q19.2.3}$$

as quais podem ser substituídas na Equação (Q19.2.1) para fornecer

$$f(t) = a_0 + \sum_{k=1}^{\infty}\left(e^{ik\omega_0 t}\frac{a_k - ib_k}{2} + e^{-ik\omega_0 t}\frac{a_k + ib_k}{2}\right) \tag{Q19.2.4}$$

porque $1/i = -i$. Pode-se definir as diversas constantes

$$\tilde{c}_0 = a_0$$

$$\tilde{c}_k = \frac{a_k - ib_k}{2}$$

$$\tilde{c}_{-k} = \frac{a_{-k} - ib_{-k}}{2} = \frac{a_k + ib_k}{2} \tag{Q19.2.5}$$

onde, por causa das propriedades ímpar e par do seno e do cosseno, $a_k = a_{-k}$ e $b_k = -b_{-k}$. A Equação (Q19.2.4) pode, portanto, ser reescrita como

$$f(t) = \tilde{c}_0 + \sum_{k=1}^{\infty}\tilde{c}_k e^{ik\omega_0 t} + \sum_{k=1}^{\infty}\tilde{c}_{-k}e^{-ik\omega_0 t}$$

ou

$$f(t) = \sum_{k=0}^{\infty}\tilde{c}_k e^{ik\omega_0 t} + \sum_{k=1}^{\infty}\tilde{c}_{-k}e^{-ik\omega_0 t}$$

Para simplificar ainda mais, em vez de somar a segunda série de 1 a ∞, faz-se a soma de –1 a $-\infty$,

$$f(t) = \sum_{k=0}^{\infty}\tilde{c}_k e^{ik\omega_0 t} + \sum_{k=-1}^{-\infty}\tilde{c}_{-k}e^{ik\omega_0 t}$$

ou

$$f(t) = \sum_{k=-\infty}^{\infty}\tilde{c}_k e^{ik\omega_0 t} \tag{Q19.2.6}$$

onde a soma inclui o termo $k = 0$.

Para calcular os $\tilde{c}_k$'s, as Equações (19.18) e (19.19) são substituídas na Equação (Q19.2.5) para fornecer

$$\tilde{c}_k = \frac{1}{T}\int_{-T/2}^{T/2} f(t)\cos(k\omega_0 t)\,dt - i\frac{1}{T}\int_{-T/2}^{T/2} f(t)\text{sen}\,(k\omega_0 t)\,dt$$

Usando as Equações (Q19.2.2) e (Q19.2.3) e simplificando, obtém-se

$$\tilde{c}_k = \frac{1}{T}\int_{-T/2}^{T/2} f(t)e^{-ik\omega_0 t}\,dt \tag{Q19.2.7}$$

Portanto, as Equações (Q19.2.6) e (Q19.2.7) são as versões complexas da Equações (19.17) a (19.20). Observe que o Apêndice A inclui um resumo das relações entre todas as formas da série de Fourier incluídas neste capítulo.

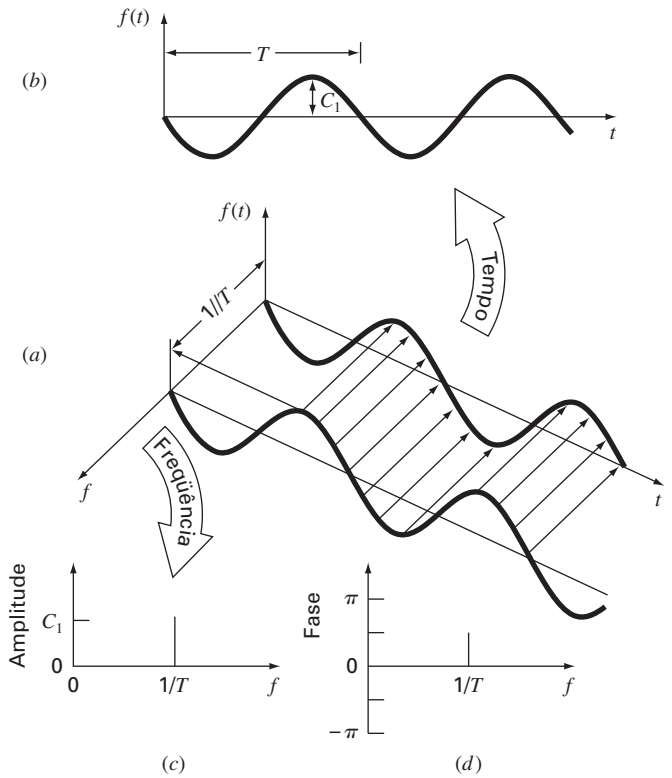
descritos na Figura 19.7*a*, na qual foi desenhado um gráfico tridimensional da função senoidal,

$$f(t) = C_1 \cos\left(t + \frac{\pi}{2}\right)$$

Nesse gráfico, a amplitude da curva, $f(t)$, é a variável dependente e o tempo t e a freqüência $f = \omega_0/2\pi$ são as variáveis independentes. Logo, os eixos da amplitude e do tempo formam um *plano temporal* e a amplitude e a freqüência formam um *plano de freqüências*. A função senoidal pode, portanto, ser concebida como existindo a uma distância $1/T$ ao longo do eixo de freqüência e correndo paralela ao eixo do tempo. Conseqüentemente, quando falamos sobre o comportamento de uma função senoidal no domínio do tempo, queremos dizer a projeção da curva sobre o plano temporal (Figura 19.7*b*). Analogamente, o comportamento no domínio da freqüência é simplesmente a projeção no plano da freqüência.

Como na Figura 19.7*c*, essa projeção é uma medida da amplitude positiva máxima da função senoidal C_1. A variação total pico a pico é desnecessária por causa da simetria. Juntamente com a posição $1/T$ ao longo do eixo de freqüência, a Figura 19.7*c* agora define a amplitude e a freqüência da função senoidal. Isso é informação suficiente para reproduzir a forma da curva no domínio do tempo. Contudo, um parâmetro a mais, ou seja, o ângulo de fase, é necessário para posicionar a curva em relação a $t = 0$. Conseqüentemente, um diagrama de fase, como o mostrado na Figura 19.7*d*, também deve ser incluído. O ângulo de fase é determinado como a distância (em radianos) do zero ao ponto no qual um pico positivo ocorre. Se o pico ocorrer depois do zero, ele é dito atrasado

FIGURA 19.7
(*a*) Uma descrição de como uma função senoidal pode ser retratada nos domínios do tempo e da freqüência. A projeção temporal é reproduzida em (*b*), enquanto que a projeção na amplitude-freqüência é reproduzida em (*c*). A projeção fase-freqüência é mostrada em (*d*).

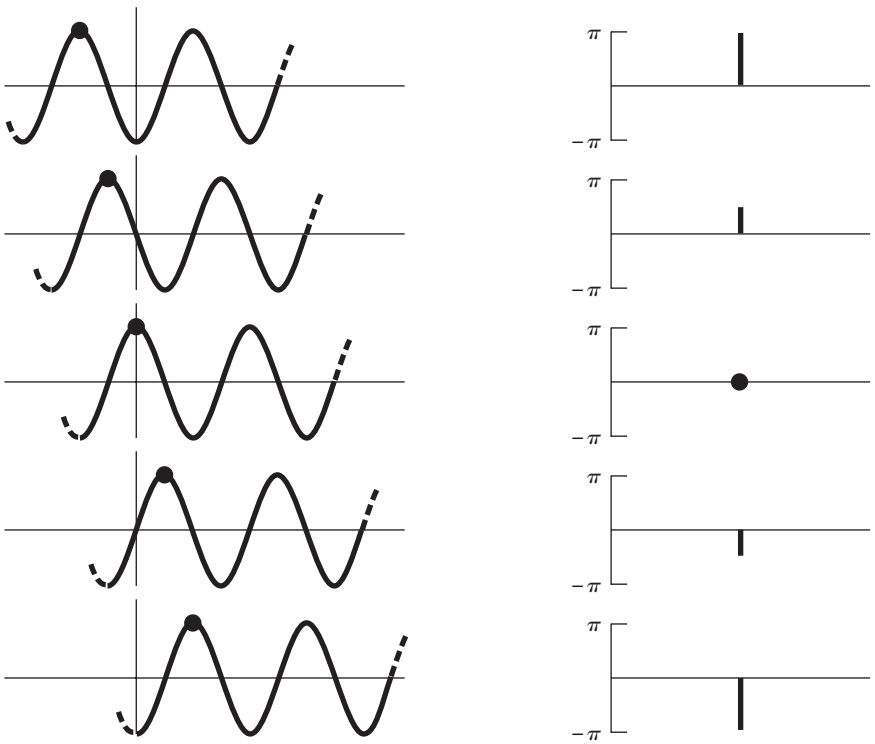


(lembre-se da discussão de retardamentos ou de avanço na Seção 19.1) e, por convenção, o ângulo de fase é dado com um sinal negativo. Reciprocamente, um pico antes do zero é dito avançado e o ângulo de fase é positivo. Logo, para a Figura 19.7, o pico está avançado com relação a zero e o ângulo de fase é marcado como $+\pi/2$. A Figura 19.8 descreve algumas outras possibilidades.

Pode-se ver agora que as Figuras 19.7*c* e *d* fornecem uma forma alternativa de apresentar ou resumir as características pertinentes da curva senoidal da Figura 19.7*a*. Elas são chamadas de *linha espectral*. É claro que, para uma única função senoidal, elas não são muito interessantes. Entretanto, quando aplicadas a situações mais complicadas, como uma série de Fourier, seus verdadeiros poder e valor são revelados. Por exemplo, a Figura 19.9 mostra as linhas espectrais de amplitude e de fase para a função onda quadrada do Exemplo 19.2.

Tais espectros fornecem informações que não seriam aparentes a partir do domínio do tempo. Isso pode ser visto comparando-se as Figuras 19.6 e 19.9. A Figura 19.6 apresenta duas perspectivas alternativas do domínio do tempo. A primeira, a onda quadrada original, não diz nada sobre as curvas senoidais que a compõem. A alternativa é mostrar essas funções senoidais — isto é, $(4/\pi)\cos(\omega_0 t)$, $-(4/3\pi)\cos(3\omega_0 t)$, $(4/5\pi)\cos(5\omega_0 t)$ etc. Tal alternativa não fornece uma visualização adequada da estrutura desses harmônicos. Em contraste, as Figuras 19.9*a* e *b* fornecem uma descrição gráfica dessa estrutura. Como tal, a linha espectral representa "impressões digitais" que ajudam a caracterizar e a entender uma forma de onda complicada. Elas são particularmente valiosas nos casos não idealizados, em que, algumas vezes, permitem discernir a estrutura em sinais que de outra forma seriam obscuros. Na próxima seção, será descrita a transformada de Fourier que permitirá estender tais análises a ondas com forma não-periódica.

FIGURA 19.8
Várias fases de uma função senoidal mostrando as linhas espectrais de fase associadas.



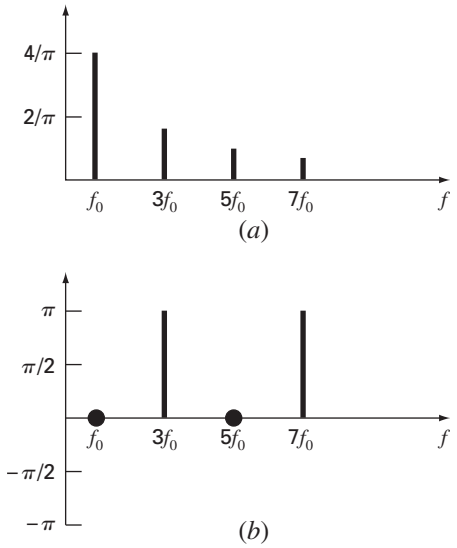


FIGURA 19.9
Linhas espectrais de (*a*) amplitude e (*b*) fase para a onda quadrada da Figura 19.5.

19.4 INTEGRAL E TRANSFORMADA DE FOURIER

Embora a série de Fourier seja uma ferramenta útil na investigação do espectro de uma função periódica, há muitas onda que não se repetem regularmente. Por exemplo, um relâmpago ocorre apenas uma vez (ou, pelo menos, passará um longo tempo antes que ocorra novamente), mas causará interferência nos receptores operando em uma larga escala de freqüências — como TVs, rádios e receptores de onda curta. Tal evidência sugere que um sinal não-recorrente como o produzido por um relâmpago exibe um espectro de freqüência contínuo. Como tais fenômenos são de grande interesse para os engenheiros, uma alternativa para a série de Fourier seria valiosa para analisar essas ondas não-periódicas.

A *integral de Fourier* é a ferramenta principal disponível para tal propósito. Ela pode ser deduzida da forma exponencial da série de Fourier

$$f(t) = \sum_{k=-\infty}^{\infty} \tilde{c}_k e^{ik\omega_0 t} \tag{19.23}$$

onde

$$\tilde{c}_k = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-ik\omega_0 t}\, dt \tag{19.24}$$

onde $\omega_0 = 2\pi/T$ e $k = 0, 1, 2, \ldots$

A transição de uma função periódica para uma função não-periódica pode ser feita permitindo-se que o período tenda a infinito. Em outras palavras, quando T se torna infinito, a função nunca se repete, e, assim, torna-se não periódica. Se for permitido que isso ocorra, pode ser demonstrado (por exemplo, Van Valkenburg, 1974; Hayt e Kemmerly,1986) que a série de Fourier se reduz a

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\omega_0) e^{i\omega_0 t}\, d\omega_0 \tag{19.25}$$

e os coeficientes se tornam uma função no contínuo da variável freqüência ω, como em

$$F(i\omega_0) = \int_{-\infty}^{\infty} f(t) e^{-i\omega_0 t}\, dt \tag{19.26}$$

A função $F(i\omega_0)$, como definida pela Equação (19.26), é chamada *integral de Fourier* de $f(t)$. Além disso, as Equações (19.25) e (19.26) são conhecidas coletivamente como *par da transformada de Fourier*. Logo, além de ser chamada de integral de Fourier, $F(i\omega_0)$ também é chamada de *transformada de Fourier* de $f(t)$. No mesmo espírito, $f(t)$, como definida pela Equação (19.25), é conhecida como a *transformada de Fourier inversa* de $F(i\omega_0)$. Então, o par permite transformar para a frente e para trás entre os domínios de tempo e freqüência para um sinal não-periódico.

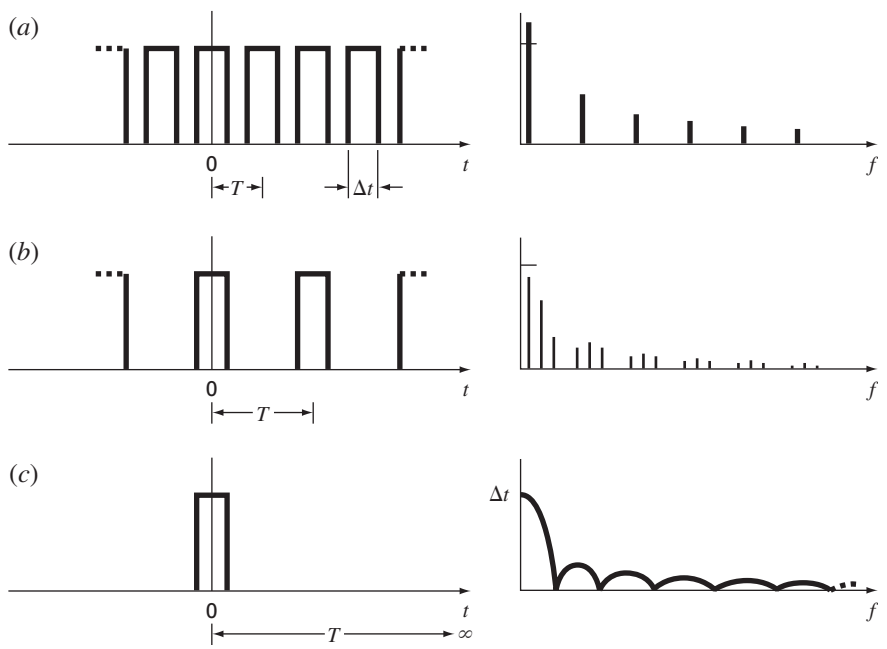
A distinção entre a série e a transformada de Fourier deveria agora estar bem clara. A principal diferença é que cada uma se aplica a uma classe diferente de funções — a série, a formas de ondas periódicas, e a transformada, a formas não-periódicas. Além dessa diferença principal, as duas abordagens diferem na forma como se movem entre os domínios de tempo e de freqüência. A série de Fourier converte uma função definida no contínuo, periódica no domínio do tempo, para amplitudes no domínio de freqüência, em freqüências discretas. Em contraste, a transformada de Fourier converte uma função no contínuo no domínio do tempo em uma função no contínuo no domínio da freqüência. Assim, o espectro de freqüência discreto gerado pela série de Fourier é análogo ao espectro de freqüência contínuo gerado pela transformada de Fourier.

O deslocamento de um espectro discreto para um espectro contínuo pode ser ilustrado graficamente. Na Figura 19.10*a*, vê-se um trem de pulso de ondas retangulares, com larguras do pulso iguais à metade do período junto com seu espectro discreto associado. Essa é a mesma função que foi investigada anteriormente no Exemplo 9.2, com a exceção de que ela foi deslocada verticalmente.

Na Figura 19.10*b*, o fato de se dobrar o período do trem de pulso tem dois efeitos no espectro. Primeiro, duas linhas a mais de freqüência são adicionadas de cada lado das componentes originais. Segundo, a amplitude das componentes é reduzida.

Conforme se permita que o período tenda a infinito, esses efeitos continuam, enquanto mais e mais linhas espectrais são agrupadas até que o espaçamento entre as linhas vai a zero. No limite, a série converge para a integral de Fourier no contínuo, descrita na Figura 19.10*c*.

FIGURA 19.10
Ilustração de como o espectro de freqüência discreta de uma série de Fourier para um trem de pulso (*a*) se aproxima de um espectro de freqüência no contínuo de uma integral de Fourier (*c*) quando for permitido que o período tenda a infinito.



Agora que foi introduzida uma forma de analisar um sinal não-periódico, vai ser dado o passo final em nossa dedução. Na próxima seção, será reconhecido o fato de que um sinal é raramente caracterizado como uma função definida no contínuo do tipo necessário para implementar a Equação (19.26). Em vez disso, os dados estão invariavelmente na forma discreta. Assim, agora vai ser mostrado como calcular uma transformada de Fourier para tais medidas discretas.

19.5 TRANSFORMADA DE FOURIER DISCRETA (TFD)

Em engenharia, as funções são, em geral, representadas por conjuntos finitos de valores discretos. Além disso, em geral, os dados são coletados ou convertidos para tal forma discreta. Como descrito na Figura 19.11, um intervalo de 0 a t pode ser dividido em N subintervalos com largura $\Delta t = T/N$. O subscrito n é usado para designar os instantes discretos nos quais as medidas foram tomadas. Logo, f_n designa o valor da função definida no contínuo $f(t)$ no instante t_n.

Observe que os pontos dados estão especificados em $n = 0, 1, 2, \ldots, N-1$. Não está incluído um valor em $n = N$. (Ver Ramirez, 1985, para as razões para a exclusão de f_N.)

Para o sistema na Figura 19.11, uma transformada de Fourier discreta pode ser escrita como

$$F_k = \sum_{n=0}^{N-1} f_n e^{-ik\omega_0 n} \qquad \text{para } k = 0 \text{ a } N-1 \tag{19.27}$$

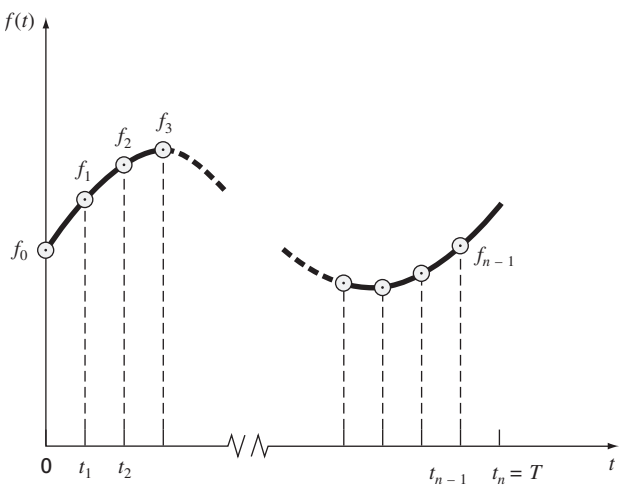
e a transformada inversa de Fourier como

$$f_n = \frac{1}{N} \sum_{k=0}^{N-1} F_k e^{ik\omega_0 n} \qquad \text{para } n = 0 \text{ a } N-1 \tag{19.28}$$

onde $\omega_0 = 2\pi/N$.

As Equações (19.27) e (19.28) representam o análogo discreto das Equações (19.26) e (19.25), respectivamente. Como tal, elas podem ser usadas para calcular tanto a transformada de Fourier direta quanto a inversa para dados discretos. Embora esses cálculos possam ser feitos à mão, são extremamente árduos. Como expresso pela Equação (19.27), a TFD requer N^2 operações complexas. Então, será desenvolvido um algoritmo computacional para implementar a TFD.

FIGURA 19.11
Os pontos amostrais na série de Fourier discreta.



Algoritmo Computacional para a TFD. Observe que o fator $1/N$ na Equação (19.28) é simplesmente um fator de escala que pode ser incluído tanto na Equação (19.27) quanto na Equação (19.28), mas não em ambas. Para o nosso algoritmo computacional, vamos mudá-lo para a Equação (19.27), de modo que o primeiro coeficiente F_0 (o qual é o análogo do coeficiente a_0) seja igual à média aritmética das amostras. Além disso, para desenvolver um algoritmo que possa ser implementado em linguagens que não aceitem variáveis complexas, podemos usar a identidade de Euler,

$$e^{\pm ia} = \cos a \pm i \text{ sen } a$$

para reescrever as Equações (19.27) e (19.28) como

$$F_k = \frac{1}{N}\sum_{n=0}^{N}[f_n \cos(k\omega_0 n) - if_n \text{ sen}(k\omega_0 n)] \tag{19.29}$$

e

$$f_n = \sum_{k=0}^{N-1}[F_k \cos(k\omega_0 n) + iF_k \text{ sen}(k\omega_0 n)] \tag{19.30}$$

O pseudocódigo para implementar a Equação (19.19) está listado na Figura 19.12. Esse algoritmo pode ser desenvolvido em um programa de computador para calcular a TFD. A saída de tal programa está listada na Figura 19.13 para a análise de uma função cosseno.

FIGURA 19.12
Pseudocódigo para calcular a TFD.

```
DOFOR k = 0, N − 1
  DOFOR n = 0, N − 1
    angulo = kω0n
    real_k = real_k + f_n cos(angle)/N
    imaginário_k = imaginário_k − f_n sin(angulo)/N
  END DO
END DO
```

FIGURA 19.13
Saída de um programa baseado no algoritmo da Figura 19.12 para a TFD de dados gerados por uma função cosseno $f(t) = \cos[2\pi(12{,}5)t]$ em 16 pontos com $\Delta t = 0{,}01$ s.

ÍNDICE	f(t)	REAL	IMAGINARIO
0	1,000	0,000	0,000
1	0,707	0,000	0,000
2	0,000	0,500	0,000
3	-0,707	0,000	0,000
4	-1,000	0,000	0,000
5	-0,707	0,000	0,000
6	0,000	0,000	0,000
7	0,707	0,000	0,000
8	1,000	0,000	0,000
9	0,707	0,000	0,000
10	0,000	0,000	0,000
11	-0,707	0,000	0,000
12	-1,000	0,000	0,000
13	-0,707	0,000	0,000
14	0,000	0,500	0,000
15	0,707	0,000	0,000

19.6 TRANSFORMADA RÁPIDA DE FOURIER (FFT)

Embora o algoritmo descrito na seção anterior calcule adequadamente a TFD, ele é computacionalmente pesado, pois são necessárias N^2 operações. Conseqüentemente, para amostra de dados mesmo de tamanho moderado, a determinação direta da TFD pode gastar muito tempo.

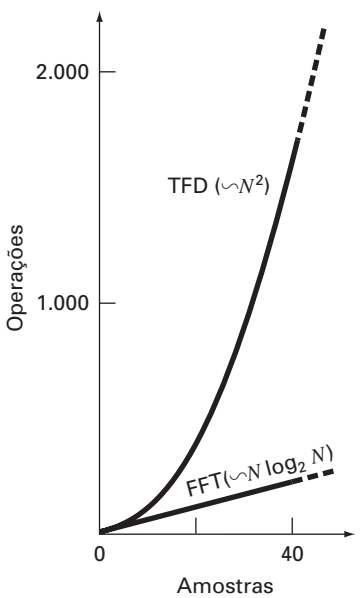
A *transformada rápida de Fourier*, ou *FFT*, é um algoritmo que foi desenvolvido para calcular a TFD de uma forma extremamente econômica. Sua velocidade vem do fato de que ela utiliza os resultados dos cálculos anteriores para reduzir o número de operações. Em particular, ela explora a periodicidade e a simetria das funções trigonométricas para calcular a transformada com aproximadamente $N \log_2 N$ operações (Figura 19.14). Logo, para $N = 50$ amostras, a FFT é da ordem de 10 vezes mais rápida do que a TFD padrão. Para $N = 1.1000$, é cerca de 100 vezes mais rápida.

O primeiro algoritmo do tipo FFT foi desenvolvido por Gauss no início do século XIX (Heideman et al., 1984). Outras importantes contribuições foram feitas por Runge, Danielson, Lanczos e outros no início do século XX. No entanto, como a transformada discreta em geral levava várias semanas para ser calculada à mão, não atraiu um vasto interesse antes do desenvolvimento do computador digital moderno.

Em 1965, J. W. Cooley e J. W. Tukey publicaram um artigo chave no qual esboçaram um algoritmo para calcular a FFT. Esse esquema, que é parecido com aqueles de Gauss e outros investigadores, é chamado de *algoritmo de Cooley-Tukey*. Hoje em dia, há muitas outras abordagens que são desdobramentos desse método.

A idéia básica por trás de cada um desses algoritmos é que uma TFD de comprimento N é decomposta ou "dizimada" em TFDs sucessivamente menores. Existe uma variedade de maneiras diferentes de implementar esse princípio. Por exemplo, o algoritmo de Cooley-Tukey é um membro do que é chamado de técnicas de *dizimação no tempo.* Na seção presente, será descrita uma abordagem alternativa chamada de *algoritmo de Sande-Tukey*, método que é um membro de outra classe de algoritmos chamados de técnicas de *dizimação na freqüência*. A distinção entre as duas classes será estudada depois de discutido o método.

FIGURA 19.14
Gráfico do número de operações em função do tamanho da amostra para a TFD padrão e para a FFT.



19.6.1 Algoritmo de Sande-Tukey

No caso presente, será suposto que N é uma potência inteira de 2,

$$N = 2^M \tag{19.31}$$

onde M é um inteiro. Essa restrição é introduzida para simplificar o algoritmo resultante. Agora, lembre-se de que a TFD pode ser representada, em geral, por

$$F_k = \sum_{n=0}^{N-1} f_n e^{-i(2\pi/N)nk} \qquad \text{para } k = 0 \text{ a } N-1 \tag{19.32}$$

onde $2\pi/N = \omega_0$. A Equação (19.32) pode também ser representada por

$$F_k = \sum_{n=0}^{N-1} f_n W^{nk}$$

onde W é uma função-peso com valores complexos definida por

$$W = e^{-i(2\pi/N)} \tag{19.33}$$

Suponha agora que a amostra seja dividida na metade e que a Equação (19.32) seja expressa em termos dos primeiros e dos últimos $N/2$ pontos:

$$F_k = \sum_{n=0}^{(N/2)-1} f_n e^{-i(2\pi/N)kn} + \sum_{n=N/2}^{N-1} f_n e^{-i(2\pi/N)kn}$$

onde $k = 0, 1, 2, \ldots, N-1$. Uma nova variável, $m = n - N/2$, pode ser criada de modo que a variação na segunda somatória seja consistente com a variação na primeira,

$$F_k = \sum_{n=0}^{(N/2)-1} f_n e^{-i(2\pi/N)kn} + \sum_{m=0}^{(N/2)-1} f_{m+N/2} e^{-i(2\pi/N)k(m+N/2)}$$

ou

$$F_k = \sum_{n=0}^{(N/2)-1} (f_n + e^{-i\pi k} f_{n+N/2}) e^{-i2\pi kn/N} \tag{19.34}$$

A seguir, perceba que o fator $e^{-i\pi k} = (-1)^k$. Logo, para pontos pares, ele é igual a 1, e para pontos ímpares, é igual a -1. Portanto, o próximo passo no método é separar a Equação (19.34) de acordo com valores pares e ímpares de k. Para os valores pares,

$$F_{2k} = \sum_{n=0}^{(N/2)-1} (f_n + f_{n+N/2}) e^{-i2\pi(2k)n/N} = \sum_{n=0}^{(N/2)-1} (f_n + f_{n+N/2}) e^{-i2\pi kn/(N/2)}$$

e para os valores ímpares,

$$\begin{aligned} F_{2k+1} &= \sum_{n=0}^{(N/2)-1} (f_n - f_{n+N/2}) e^{-i2\pi(2k+1)n/N} \\ &= \sum_{n=0}^{(N/2)-1} (f_n - f_{n+N/2}) e^{-i2\pi n/N} e^{-i2\pi kn/(N/2)} \end{aligned}$$

para $k = 0, 1, 2, \ldots, (N/2) - 1$.

Essas equações também podem ser expressas em termos da Equação (19.33). Para os valores pares,

$$F_{2k} = \sum_{n=0}^{(N/2)-1} (f_n + f_{n+N/2}) W^{2kn}$$

e para os valores ímpares,

$$F_{2k+1} = \sum_{n=0}^{(N/2)-1} (f_n - f_{n+N/2}) W^n W^{2kn}$$

Agora, pode-se perceber um fato chave. Essas expressões pares e ímpares podem ser interpretadas como sendo iguais às transformadas das seqüências de comprimento $(N/2)$

$$g_n = f_n + f_{n+N/2} \tag{19.35}$$

e

$$h_n = (f_n - f_{n+N/2})W^n \qquad \text{para } n = 0, 1, 2, \ldots, (N/2) - 1 \tag{19.36}$$

Logo, segue diretamente que

$$\left.\begin{aligned} F_{2k} &= G_k \\ F_{2k+1} &= H_k \end{aligned}\right\} \qquad \text{para } k = 0, 1, 2, \ldots, (N/2) - 1$$

Em outras palavras, um cálculo com N pontos pode ser substituído por dois cálculos com $(N/2)$ pontos. Como cada um dos últimos exige aproximadamente $(N/2)^2$ multiplicações e adições complexas, a abordagem produz uma economia de um fator 2 — isto é, N^2 *versus* $2(N/2)^2 = N^2/2$.

O esquema é descrito na Figura 19.15 para $N = 8$. A TFD é calculada formando-se inicialmente as seqüências g_n e h_n e então calculando-se as TFDs para $N/2$ pontos para se obter as transformadas com numeração par e ímpar. Os pesos W^n são, algumas vezes, chamados de *twiddle factors*.

Agora, é claro que essa abordagem "divida e conquiste" pode ser repetida no segundo estágio. Assim, pode-se calcular as TFDs para $(N/4)$ pontos das quatro seqüências de comprimento $N/4$ compostas dos primeiros e dos últimos $N/4$ pontos das Equações (19.35) e (19.36).

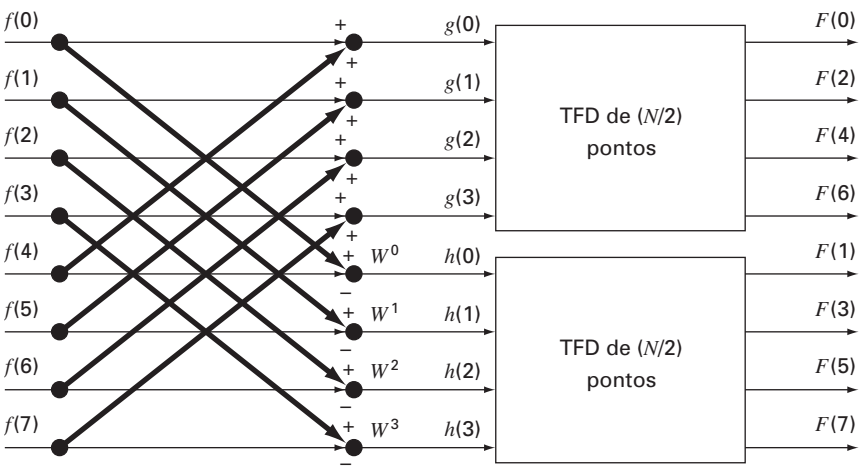
A estratégia é continuar até a conclusão inevitável, quando serão calculadas $N/2$ TFD de dois pontos (Figura 19.16). O número total de operações para o cálculo total é da ordem de $N \log_2 N$. O contraste entre esse nível de esforço e aquele da TFD padrão (Figura 19.14) ilustra por que a *FFT* é tão importante.

Algoritmo Computacional. É uma tarefa relativamente simples expressar a Figura 19.16 como um algoritmo. Como aconteceu no caso do algoritmo para a TFD da Figura 19.12, será usada a identidade de Euler,

$$e^{\pm ia} = \cos a \pm i \text{ sen } a$$

para permitir que o algoritmo seja implementado em linguagens que não acomodem explicitamente variáveis complexas.

FIGURA 19.15
Fluxograma do primeiro estágio em uma decomposição do tipo dizimação em freqüência de uma TFD de N pontos em duas TFDs de $(N/2)$ pontos para $N = 8$.

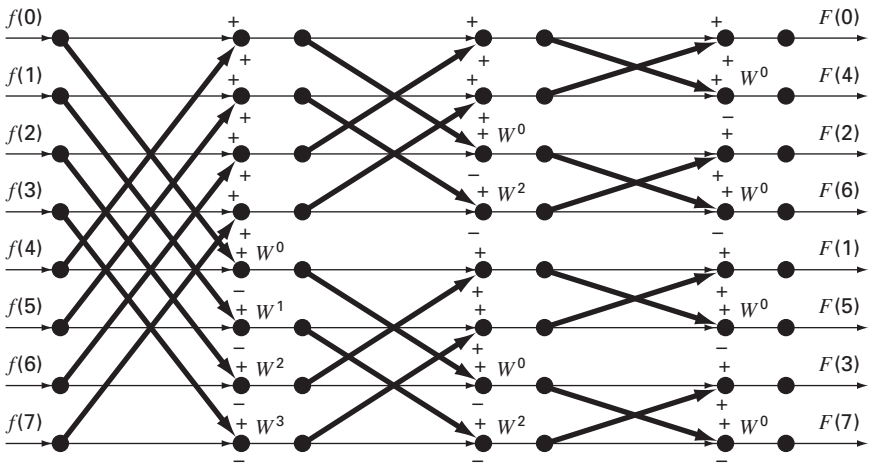


Uma inspeção mais detalhada da Figura 19.16 indica que sua célula computacional fundamental é a assim chamada *rede de borboleta* descrita na Figura 19.17*a*. O pseudocódigo para implementar uma dessas células é mostrado na Figura 19.17*b*.

O pseudocódigo para a FFT está listado na Figura 19.18. A primeira parte consiste essencialmente em três laços encaixados para implementar os cálculos incorporados na Figura 19.16. Observe que os dados com valores reais são originalmente armazenados no vetor *x*. Observe também que o laço externo percorre os *M* estágios [lembre-se da Equação (19.31)] do fluxograma.

Depois que a primeira parte for executada, a TFD terá sido calculada em uma ordem embaralhada (ver lado direito da Figura 19.16). Esses coeficientes de Fourier podem ser desembaralhados por um procedimento chamado *inversão de bits*. Se os subscritos de 0 a 7 forem expressos em binários, a ordenação correta pode ser obtida invertendo-se esses bits (Figura 19.19). A segunda parte do algoritmo implementa esse procedimento.

FIGURA 19.16
Fluxograma da decomposição completa do tipo dizimação em freqüência de uma TFD de oito pontos.



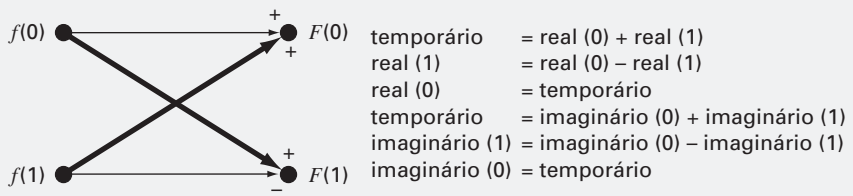


FIGURA 19.17
(*a*) Uma rede de borboleta que representa o cálculo fundamental da Figura 19.16. (*b*) Pseudocódigo para implementar (*a*).

(*a*)

```
m = LOG(N)/LOG(2)
N2 = N
DOFOR k = 1, m
  N1 = N2
  N2 = N2/2
  angulo = 0
  arg = 2π/N1
  DOFOR j = 0, N2 − 1
    c = cos(angle)
    s = −sin(angulo)
    DOFOR i = j, N − 1, N1
      kk = i + N2
      xt = x(i) − x(kk)
      x(i) = x(i) + x(kk)
      y_t = y(i) − y(kk)
      y(i) = y(i) + y(kk)
      x(kk) = xt * c − yt * s
      y(kk) = yt * c + xt * s
    END DO
    angulo = (j + 1) * arg
  END DO
END DO
```

(*b*)

```
j = 0
DOFOR i = 0, N − 2
  IF (i < J) THEN
     xt = x_j
     x_j = x_i
     x_i = xt
     yt = y_j
     y_j = y_i
     y_i = yt
  END IF
  k = N/2
  DO
     IF (k ≥ j + 1) EXIT
     j = j − k
     k = k/2
  END DO
  j = j + k
END DO
DOFOR i = 0, N − 1
  x(i) = x(i)/N
  y(i) = y(i)/N
END DO
```

FIGURA 19.18
Pseudocódigo para implementar uma FFT por dizimação em freqüência. Observe que o pseudocódigo é composto de duas partes: (*a*) a própria FFT e (*b*) uma rotina de inversão de bits para desembaralhar a ordem dos coeficientes de Fourier resultantes.

Ordem Embaralhada (Decimal)		Ordem Embaralhada (Binário)		Ordem com Bits Invertidos (Binário)		Resultado Final (Decimal)
F(0)		F(000)		F(000)		F(0)
F(4)		F(100)		F(001)		F(1)
F(2)		F(010)		F(010)		F(2)
F(6)	⇒	F(110)	⇒	F(011)	⇒	F(3)
F(1)		F(001)		F(100)		F(4)
F(5)		F(101)		F(101)		F(5)
F(3)		F(011)		F(110)		F(6)
F(7)		F(111)		F(111)		F(7)

FIGURA 19.19
Descrição do processo de inversão dos bits.

19.6.2 Algoritmo de Cooley-Tukey

A Figura 19.20 mostra um fluxograma para implementar o algoritmo de Cooley-Tukey. Nesse caso, a amostra é inicialmente dividida em pontos com numeração par e ímpar, e os resultados finais estão na ordem correta.

Essa abordagem é chamada de *dizimação no tempo* e é o inverso do algoritmo de Sande-Tukey descrito na seção anterior. Embora as duas classes de métodos sejam diferentes na organização, ambas exibem $N \log_2 N$ operações, as quais são o ponto forte da abordagem FFT.

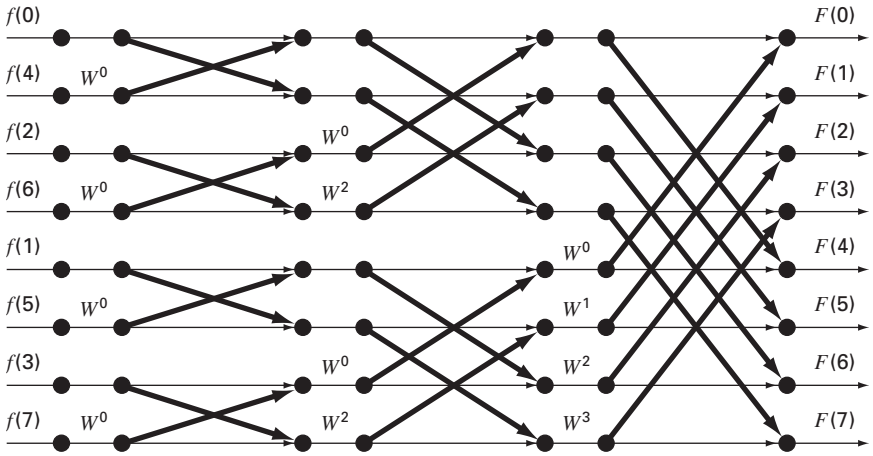


FIGURA 19.20
Fluxograma de uma FFT por dizimação no tempo de uma TFD de oito pontos.

19.7 O ESPECTRO DE POTÊNCIA

A FFT tem muitas aplicações na engenharia, variando da análise de vibrações de estruturas e mecanismos a processamento de sinais. Como descrito anteriormente, os espectros de amplitude e fase fornecem um meio de discernir a estrutura subjacente de sinais aparentemente aleatórios. Analogamente, uma análise útil chamada de espectro de potência pode ser desenvolvida a partir da transformada de Fourier.

Como o nome indica, o espectro de potência deriva da análise da saída de potência de sistemas elétricos. Em termos matemáticos, a potência de um sinal periódico no domínio do tempo pode ser definida por

$$P = \frac{1}{T}\int_{-T/2}^{T/2} f^2(t)\,dt \tag{19.37}$$

Agora, uma outra forma de olhar para essa informação é expressá-la no domínio da freqüência calculando a potência associada a cada componente da freqüência. Essa informação pode então ser exibida como um *espectro de potência*, um gráfico da potência em função da freqüência.

Se a série de Fourier para $f(t)$ for

$$f(t) = \sum_{k=-\infty}^{\infty} F_k e^{ik\omega_0 t} \tag{19.38}$$

vale a seguinte relação (ver Gabel e Roberts, 1987, para detalhes):

$$\frac{1}{T}\int_{-T/2}^{T/2} f^2(t)\,dt = \sum_{k=-\infty}^{\infty} |F_k|^2 \tag{19.39}$$

Logo, a potência em $f(t)$ pode ser determinada somando-se os quadrados dos coeficientes de Fourier, ou seja, as potências associadas às componentes de freqüências individuais.

Agora, lembre-se de que, nessa representação, o único harmônico real consiste em ambas as componentes de freqüência em $\pm k\omega_0$. Sabemos também que os coeficientes positivo e negativo são iguais. Portanto, a potência em $f_k(t)$, o k-ésimo harmônico real de $f(t)$, é

$$p_k = 2\,|F_k|^2 \tag{19.40}$$

O espectro de potência é o gráfico de p_k como uma função da freqüência $k\omega_0$. Dedicaremos a Seção 20.3 a uma aplicação em engenharia envolvendo a FFT e o espectro de potência gerado com pacotes de software.

Informação Adicional. A discussão anterior foi uma breve introdução à aproximação de Fourier e à FFT. Informações adicionais sobre a primeira podem ser encontradas em Van Valkenburg (1974), Chirlian (1969), e Hayt e Kemmerly (1986). Referências sobre a FFT incluem Davis e Rabinowitz (1975); Cooley, Lewis e Welch (1977); e Brigham (1974). Uma boa introdução a ambas pode ser encontrada em Ramirez (1985), Oppenheim e Schafer (1975), e Gabel e Roberts (1987).

19.8 AJUSTE DE CURVAS COM BIBLIOTECAS E PACOTES

As bibliotecas e os pacotes de software têm grandes recursos para o ajuste de curvas. Nesta seção, será dada uma pequena amostra dos mais úteis.

19.8.1 Excel

No presente contexto, a aplicação mais útil do Excel é para a análise de regressão e, em uma extensão menor, para a interpolação polinomial. Além de umas poucas funções de biblioteca (ver Tabela 19.1), há duas formas primárias pelas quais esses recursos podem ser implementados: o comando Trendline e o Data Analysis Toolpack (pacote de ferramentas para análise de dados).

O Comando Trendline (Menu Insert). Esse comando permite que diversos modelos de tendência diferentes sejam adicionados a um gráfico. Os modelos incluem ajustes linear, polinomial, logarítmico, exponencial, por potência e por média móvel. Os seguintes exemplos ilustram como o comando Trendline é usado.

TABELA 19.1 Funções de biblioteca do Excel relacionadas ao ajuste de dados por regressão.

Função	Descrição
FORECAST	Retorna um valor em uma tendência linear
GROWTH	Retorna valores em uma tendência exponencial
INTERCEPT	Retorna a intersecção com o eixo *y* da reta de regressão linear
LINEST	Retorna os parâmetros de uma tendência linear
LOGEST	Retorna os parâmetros de uma tendência exponencial
SLOPE	Retorna a inclinação da reta de regressão linear
TREND	Retorna valores em uma tendência linear

EXEMPLO 19.3 Usando o Comando Trendline do Excel

Enunciado do Problema. Você pode ter observado que diversos dos ajustes disponíveis no Trendline foram discutidos anteriormente no Capítulo 17 (por exemplo, linear, polinomial, exponencial e por potência). Um recurso adicional é o modelo logarítmico

$$y = a_0 + a_1 \log x$$

Ajuste os seguintes dados com esse modelo usando o comando Trendline do Excel:

x	0,5	1	1,5	2	2,5	3	3,5	4	4.5	5	5,5
y	0,53	0,69	1,5	1,5	2	2,06	2,28	2,23	2,73	2,42	2,79

Solução. Para chamar o comando Trendline, deve ser criado um gráfico relacionando uma série de variáveis dependentes e independentes. Para o caso presente, usamos a ferramenta Chart Wizard do Excel para criar um gráfico dos dados no plano XY.

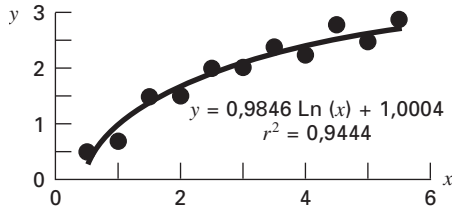


FIGURA 19.21
Ajuste de um modelo logarítmico aos dados do Exemplo 19.3.

A seguir, pode-se escolher o gráfico (clicando duas vezes sobre ele) e a série (posicionando o cursor do mouse sobre um dos valores e clicando uma vez). Os comandos Insert e Trendline são então chamados com o mouse ou digitando

/ **I**nsert T**r**endline

Nesse ponto, uma janela de diálogo se abrirá com duas etiquetas: Options e Type. A etiqueta Options fornece uma forma de adaptar o ajuste ao gosto do usuário. O mais importante no contexto presente é exibir tanto a equação quanto o valor do coeficiente de determinação (r^2) no gráfico. A principal escolha na etiqueta Type é especificar o tipo de tendência. No caso presente, escolha **Logarithmic**. O ajuste resultante junto com r^2 está mostrado na Figura 19.21.

O comando Trendline fornece uma maneira prática de ajustar diversos modelos comumente usados aos dados. Além disso, a inclusão da opção **Polinomial** significa que ele também pode ser usado para interpolação polinomial. Entretanto, o fato de seu conteúdo estatístico ser limitado a r^2 significa que ele não permite que sejam feitas inferências estatísticas com relação ao ajuste do modelo. O Data Analysis Toolpack descrito a seguir fornece uma boa alternativa quando tais inferências forem necessárias.

O Data Analysis Toolpack. Esse Excel Add-In Package contém recursos abrangentes para o ajuste de curvas com o método dos mínimos quadrados linear geral. Como descrito anteriormente na Seção 17.4, tais modelos são da forma geral

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots + a_m z_m + e \tag{17.23}$$

onde $z_0, z_1, \ldots, z_m$ são $m + 1$ funções diferentes. O próximo exemplo ilustra como esses modelos podem ser ajustados com o Excel.

EXEMPLO 19.4 Usando o Data Analysis Toolpack do Excel

Enunciado do Problema. Os seguintes dados foram coletados para a inclinação, o raio hidráulico e a velocidade da água escoando em um canal:

S, m/m	0,0002	0,0002	0,0005	0,0005	0,001	0,001
R, m	0,2	0,5	0,2	0,5	0,2	0,5
U, m/s	0,25	0,5	0,4	0,75	0,5	1

Há razões teóricas (lembre-se da Seção 8.2) para acreditar que esses dados possam ser ajustados por um modelo de potência da forma

$$U = \alpha S^{\sigma} R^{\rho}$$

onde α, σ e ρ são coeficientes obtidos empiricamente. Há razões teóricas (novamente, ver Seção 8.2) para acreditar que σ e ρ deveriam ter valores de aproximadamente 0,5 e 0,667, respectivamente. Ajuste esses dados com o Excel e avalie se sua estimativa de regressão contradiz os valores esperados para os coeficientes do modelo.

Solução. O logaritmo do modelo de potência é usado primeiro para convertê-lo para uma forma linear da Equação (17.23),

$$U = \log \alpha + \sigma \log S + \rho \log R$$

Uma planilha em Excel pode ser desenvolvida com ambos os dados originais junto com os logaritmos comuns, como na seguinte:

	A	B	C	D	E	F
1	S	R	U	log(S)	log(R)	log(U)
2	0.0002	0.2	0.25	**-3.69897**	-0.69897	-0.60206
3	0.0002	0.5	0.5	-3.69897	-0.30103	-0.30103
4	0.0005	0.2	0.4	-3.30103	-0.69897	-0.39794
5	0.0005	0.5	0.75	-3.30103	-0.30103	-0.12494
6	0.001	0.2	0.5	-3	-0.69897	-0.30103
7	0.001	0.5	1	-3	-0.30103	0

=log(A2)

Como mostrado, uma forma eficiente de gerar o logaritmo é digitar a fórmula para calcular o primeiro log(S). Essa fórmula pode ser copiada à direita e para baixo para gerar os outros logaritmos.

Por causa de sua situação como "add-in" (programa de assistência) na versão do Excel disponível no momento da impressão do livro, o Data Analysis Toolpack às vezes precisa ser carregado no Excel. Para fazer isso, simplesmente use o mouse ou digite

/**T**ools Add-I**n**s

A seguir, escolha **Analysis Toolpack** e OK. Se o "add-in" tiver sucesso, a escolha Data Analysis será adicionada ao menu Tools.

Depois de escolher **Data Analysis** do menu Tools, um menu Data Analysis aparecerá na tela contendo um grande número de rotinas estatisticamente orientadas. Escolha **Regression** e uma janela de diálogo aparecerá, pedindo a você informações sobre a regressão. Depois de se certificar de que a escolha padrão **New Worksheet Ply** foi escolhida, preencha F2:F7 com o intervalo para y e D2:E7 com o intervalo para x e selecione OK. A seguinte planilha será criada:

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	*Regression Statistics*						
4	Multiple R	0.998353					
5	R Square	0.996708					
6	Adjusted R Square	0.994513					
7	Standard Error	0.015559					
8	Observations	6					
9							
10	ANOVA						
11		*df*	*SS*	*MS*	*F*	*Significance F*	
12	Regression	2	0.219867	0.10993	454.1106	0.0001889	
13	Residual	3	0.000726	0.00024			
14	Total	5	0.220593				
15							
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
17	Intercept	**1.522452**	0.075932	20.05010	0.000271	1.2808009	1.7641028
18	X Variable 1	**0.433137**	0.022189	19.52030	0.000294	**0.362521**	**0.503752**
19	X Variable 2	**0.732993**	0.031924	22.96038	0.000181	**0.631395**	**0.834590**

Logo, o ajuste resultante será

$$\log U = 1{,}522 + 0{,}433 \log S + 0{,}733 \log R$$

ou, calculando a exponencial na base 10,

$$U = 33{,}3S^{0{,}433}R^{0{,}733}$$

Observe que intervalos de confiança de 95% serão gerados para os coeficientes. Então, há uma probabilidade de 95% de que o expoente da inclinação verdadeira caia entre 0,363 e 0,504 e que o coeficiente do raio hidráulico verdadeiro caia entre 0,631 e 0,835. Logo, o ajuste não contradiz os expoentes teóricos.

Finalmente, deve ser observado que a ferramenta Excel Solver pode ser usada para fazer uma *regressão não-linear* minimizando diretamente a soma dos quadrados dos resíduos entre a previsão do modelo não-linear e os dados. A Seção 20.1 é dedicada a um exemplo de como isso pode ser feito.

19.8.2 MATLAB

Como resumido na Tabela 19.2, o software MATLAB tem diversas funções de biblioteca que geram os recursos totais descritos nesta parte do livro. O exemplo a seguir ilustra como uns poucos deles podem ser usados.

TABELA 19.2 Algumas das funções MATLAB para implementar interpolação, regressão, splines e FFT.

Função	Descrição
polyfit	Ajusta polinômios aos dados
interp1	Interpolação em 1 D
interp2	Interpolação em 2 D
spline	Interpolação dos dados com spline cúbico
fft	Transformada de Fourier discreta

EXEMPLO 19.5 Usando o MATLAB para Ajuste de Curvas

Enunciado do Problema. Explore como o MATLAB pode ser usado para ajustar curvas aos dados. Para fazê-lo, use a função seno para gerar valores de $f(x)$ igualmente espaçados de 0 a 10. Use um tamanho de passo 1 de modo que a caracterização resultante da onda seno seja esparsa (Figura 19.22). A seguir, ajuste-a com (*a*) interpolação linear, (*b*) um polinômio de 5º grau e (*c*) um spline cúbico.

Solução.

(a) Os valores para as variáveis independente e dependente podem ser colocados em vetores por

```
>> x=0:10;
>> y=sen(x);
```

Um novo vetor, espaçado de modo mais fino nas variáveis independentes, pode ser gerado e armazenado no vetor **xi,**

```
>> xi=0:.25:10;
```

A função MATLAB **interp1** pode então ser utilizada para gerar valores da variável dependente *yi* para todos os valores *xi* usando interpolação linear. Ambos os dados originais (*x*, *y*) e os dados interpolados linearmente podem ser marcados juntos, como mostrado no seguinte gráfico:

```
>> yi=interp1(x,y,xi);
>> plot(x,y,'o',xi,yi)
```

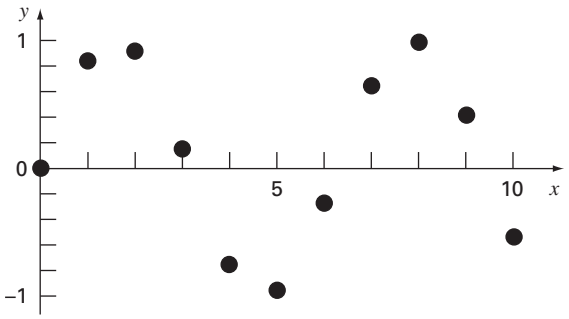
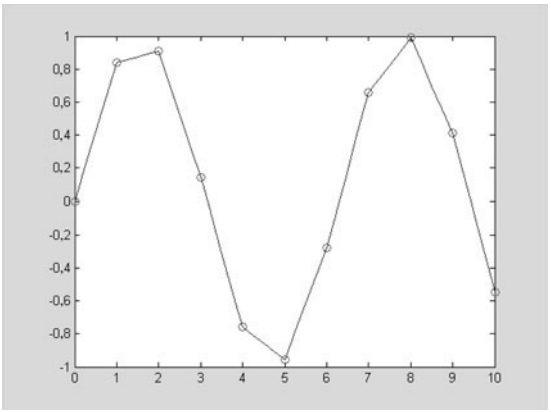


FIGURA 19.22
A mostra de 11 pontos de uma curva senoidal.

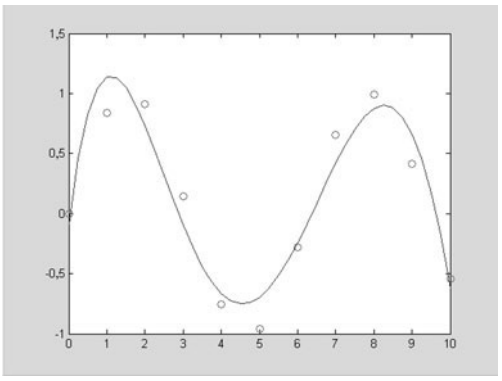


(b) A seguir, a função **polyfit** do MATLAB pode ser utilizada para gerar os coeficientes de um polinômio de 5º grau que ajuste os dados originais esparsos.

```
>> p=polyfit(x,y,5)
p=
   0.0008  -0.0290   0.3542  -1.6854   2.5860  -0.0915
```

onde o vetor **p** guarda os coeficientes do polinômio, os quais podem, por sua vez, ser usados para gerar um novo conjunto de valores *yi*, que podem novamente ser marcados junto com a amostra esparsa original,

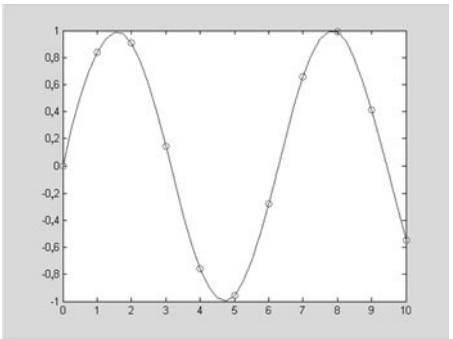
```
>> yi = polyval(p,xi);
>> plot(x,y,'o',xi,yi)
```



Assim, o polinômio captura o padrão geral dos dados, mas não passa pela maioria dos pontos.

(c) Finalmente, a função **spline** do MATLAB pode ser usada para ajustar um spline cúbico aos dados originais esparsos na forma de um novo conjunto de valores *yi*, os quais podem novamente ser marcados junto com a amostra esparsa original.

```
>> yi=spline(x,y,xi);
>> plot(x,y,'o',xi,yi)
```



Deve ser observado que o MATLAB também tem excelentes recursos para fazer a análise de Fourier. A Seção 20.3 será dedicada a um exemplo de como isso pode ser realizado.

19.8.3 IMSL

O IMSL tem diversas rotinas para ajuste de curvas que geram todos os recursos cobertos neste livro e mais alguns. Uma amostra é apresentada na Tabela 19.3. Na presente discussão, o enfoque será na rotina RCURV. Essa rotina ajusta um polinômio por mínimos quadrados aos dados.

A RCURV é implementada pelo seguinte comando CALL:

```
CALL RCURV (NOBS, XDATA, YDATA, NDEG, B, SSPOLY, STAT)
```

onde NOBS = Número de observações. (Entrada)

XDATA = Vetor de comprimento NOBS contendo os valores de *x*. (Entrada)

YDATA = Vetor de comprimento NOBS contendo os valores de *y*. (Entrada)

NDEG = Grau do polinômio. (Entrada)

B = Vetor de comprimento NDEG + 1 contendo os coeficientes.

SSPOLY = Vetor de comprimento NDEG + 1 contendo a seqüência das somas dos quadrados. (Saída) SSPOLY(1) contém a soma dos quadrados devida à média. Para i = 1, 2, . . . NDEG, SSPOLY(i + 1) contém a soma dos quadrados devida a x^i ajustado para a média, x, x^2, . . . e x^{i-1}.

STAT = Vetor de comprimento 10 contendo estatísticas. (Saída)

onde 1 = Média de *x*
2 = Média de *y*
3 = Variância da amostra de *x*
4 = Variância da amostra de *y*
5 = Quadrado de *R* (em porcentagem)
6 = Graus de liberdade para a regressão
7 = Soma dos quadrados da regressão
8 = Graus de liberdade do erro
9 = Soma dos quadrados dos erros
10 = Número de pontos dados (*x*, *y*) contendo NAN (não um número) como valor de *x* ou de *y*.

TABELA 19.3 Rotinas IMSL para ajuste de curvas.

Categoria	Rotina	Descrição
• Interpolação por spline cúbico	CSIEZ	Rotina para spline cúbico fácil de usar
	CSINT	Not-a-knot
	CSDEC	Condições nas derivadas nas extremidades
• Cálculo e integração de spline cúbico	CSVAL	Cálculo
	CSDER	Cálculo da derivada
	CS1GD	Cálculo em um reticulado
	CSITG	Integração
• Interpolação por B-spline		
• Polinomial por partes		
• Rotinas para interpolação polinomial quadrática para dados em reticulados		
• Interpolação de dados dispersos		
• Aproximação por mínimos quadrados	RLINE	Polinomial linear
	RCURV	Polinomial geral
	FNLSQ	Funções gerais
• Suavização por spline cúbico		
• Aproximação de Chebyshev com pesos racionais		Aproximação de Chebyshev com pesos racionais
• FFT trigonométrica real	FFTRF	Transformada progressiva
	FFTRB	Transformada regressiva ou inversa
	FFTRI	Rotina de inicialização para a FFTR
• FFT exponencial complexa	FFTCF	Transformada progressiva
	FFTCB	Transformada regressiva ou inversa
	FFTCI	Rotina de inicialização para a FFTC
• FFTs com senos e cossenos reais		
• Real quarter sine and quarter cosine FFT		
• FFTs complexas bi e tridimensionais		
• Convoluções e correlações		
• Transformada de Laplace		

EXEMPLO 19.6 Usando o IMSL para Regressão Polinomial

Enunciado do Problema. Use o RCURV para determinar um polinômio cúbico que forneça um ajuste por mínimos quadrados para os seguintes dados:

x	0,05	0,12	0,15	0,30	0,45	0,70	0,84	1,05
y	0,957	0,851	0,832	0,720	0,583	0,378	0,295	0,156

Solução. Um exemplo de um programa principal em Fortran 90 e função usando RCURV para resolver esse problema pode ser escrito como

```
PROGRAM Fitpoly
use msimsl
IMPLICIT NONE
INTEGER::ndeg,nobs,i,j
PARAMETER (ndeg=3, nobs=8)
REAL::b(ndeg+1),sspoly(ndeg+1),stat(10),x(nobs),y(nobs),
  ycalc(nobs)
DATA x/0.05,0.12,0.15,0.30,0.45,0.70,0.84,1.05/
DATA y/0.957,0.851,0.832,0.720,0.583,0.378,0.295,0.156/
CALL RCURV(nobs,x,y,ndeg,B,sspoly,stat)
PRINT *, 'Fitted polynomial is'
DO i = 1,ndeg+1
  PRINT '(1X, "X^",I1," TERM: ",F8.4)', i-1, b(i)
END DO
PRINT *
PRINT '(1X,"R^2: ",F5.2,"%")',stat(5)
PRINT *
PRINT *, 'NO. X  Y  YCALC'
DO i = 1,nobs
  ycalc=0.
  DO j = 1,ndeg+1
    ycalc(i)=ycalc(i)+b(j)*x(i)**(j-1)
  END DO
  PRINT '(1X,I8,3(5X,F8.4))', i, x(i), y(i), ycalc(i)
END DO
END
```

Um exemplo de execução é

```
Fitted polynomial is
X^0 TERM:    .9909
X^1 TERM:  -1.0312
X^2 TERM:    .2785
X^3 TERM:   -.0513
R^2: 99.81%
      NO.         X          Y        YCALC
       1         .0500      .9570      .9401
       2         .1200      .8510      .8711
       3         .1500      .8320      .8423
       4         .3000      .7200      .7053
       5         .4500      .5830      .5786
       6         .7000      .3780      .3880
       7         .8400      .2950      .2908
       8        1.0500      .1560      .1558
```

PROBLEMAS

19.1 O pH de um reator varia senoidalmente no decorrer de um dia. Use regressão por mínimos quadrados para ajustar a Equação (19.11) aos seguintes dados. Use seu ajuste para determinar a média, a amplitude e o instante de máximo pH.

Tempo, h	0	2	4	5	7	9	12	15	20	22	24
pH	7,6	7,2	7	6,5	7,5	7,2	8,9	9,1	8,9	7,9	7

19.2 A radiação solar em Tucson, no Arizona, foi tabulada como

Tempo, meses	J	F	M	A	M	J	J	A	S	O	N	D
Radiação, W/m^2	144	188	245	311	351	359	308	287	260	211	159	131

Supondo que cada mês dure 30 dias, ajuste uma curva senoidal a esses dados. Use a equação resultante para prever a radiação em meados de agosto.

19.3 Os valores médios de uma função podem ser determinados por

$$\overline{f(x)} = \frac{\int_0^x f(x)dx}{x}$$

Use essa relação para verificar os resultados da Equação (19.13).

19.4 Use uma série de Fourier no contínuo para aproximar a onda dente de serra na Figura P19.4. Trace os primeiros três termos junto com a soma.

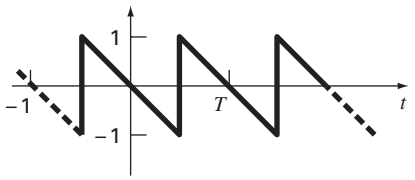


Figura P19.4
Uma onda dente de serra.

19.5 Use uma série de Fourier no contínuo para aproximar a forma de onda na Figura P19.5. Trace os primeiros três termos junto com a soma.

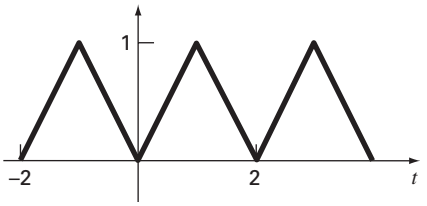


Figura P19.5
Uma onda triangular.

19.6 Construa as linhas espectrais de amplitude e de fase para o Problema 19.4.
19.7 Construa as linhas espectrais de amplitude e de fase para o Problema 19.5.
19.8 Um retificador de meia onda pode ser caracterizado por

$$C_1 = \left[\frac{1}{\pi} + \frac{1}{2}\operatorname{sen} t - \frac{2}{3\pi}\cos 2t - \frac{2}{15\pi}\cos 4t - \frac{2}{35\pi}\cos 6t - \cdots\right]$$

onde C_1 é a amplitude da onda. Trace os primeiros quatro termos junto com a soma.
19.9 Construa as linhas espectrais de amplitude e de fase para o Problema 19.8.
19.10 Desenvolva um programa amigável ao usuário para a TFD baseado no algoritmo da Figura 19.12. Teste-o reproduzindo a Figura 19.13.
19.11 Use o programa do Problema 19.10 para calcular a TFD para a onda triangular do Problema 19.8. Faça uma amostra da onda de $t = 0$ a $4T$. Use 32, 64 e 128 pontos amostrais. Cronometre cada execução e trace o tempo de execução em função de N para verificar a Figura 19.14.
19.12 Desenvolva um programa amigável ao usuário para a FFT baseado no algoritmo da Figura 19.18. Teste-o reproduzindo a Figura 19.13.
19.13 Repita o Problema 19.11 usando o software que você desenvolveu no Problema 19.12.
19.14 Um objeto é suspenso em um túnel de vento e a força é medida para diversos níveis de velocidade do vento. Os resultados estão tabulados a seguir. Use o comando Trendline do Excel para ajustar uma equação de potência aos dados. Trace F em função de v junto com a equação de potência e r^2.

v, m/s	10	20	30	40	50	60	70	80
F, N	25	70	380	550	610	1.220	830	1.450

19.15 Use o Data Analysis Toolpack do Excel para desenvolver uma regressão polinomial para os seguintes dados de concentração de oxigênio dissolvido na água fresca em função da temperatura no nível do mar. Determine o grau necessário do polinômio para ser compatível com a precisão dos dados.

T, °C	0	8	16	24	32	40
o, mg/l	14,62	11,84	9,87	8,42	7,31	6,41

19.16 Use o Data Analysis Toolpack do Excel para ajustar uma reta aos seguintes dados. Determine o intervalo de confiança de 90% para a intersecção com o eixo y. Se ele englobar o zero, refaça a regressão, mas a intersecção com o eixo y forçando a ser zero (essa é uma opção na janela de diálogo da **Regression**).

x	2	4	6	8	10	12	14
y	6,5	7	13	17,8	19	25,8	26,9

19.17 (**a**) Use o MATLAB para ajustar um spline cúbico aos seguintes dados:

x	0	2	4	7	10	12
y	20	20	12	7	6	6

Determine o valor de y em $x = 1{,}5$. (**b**) Repita (**a**), mas com primeiras derivadas nulas nos nós extremos. Observe que o instrumento "help" do MATLAB descreve como prescrever as derivadas nas extremidades.
19.18 Use o MATLAB para gerar 64 pontos para a função

$$f(t) = \cos(10t) + \operatorname{sen}(3t)$$

de $t = 0$ a 2π. Some uma componente aleatória ao sinal com a função `randn`. Tome a FFT desses valores e trace os resultados.
19.19 De modo análogo à Seção 19.8.2, use o MATLAB para ajustar os dados do Problema 19.15 usando (**a**) interpolação linear, (**b**) uma regressão polinomial de 3º grau e (**c**) um spline. Use cada abordagem para prever a concentração de oxigênio em $T = 10$.
19.20 A função de Runge é escrita como

$$f(x) = \frac{1}{1 + 25x^2}$$

Gere 9 valores igualmente espaçados dessa função no intervalo $[-1, 1]$. Ajuste esses dados com (**a**) um polinômio de 8º grau, (**b**) um spline linear e (**c**) um spline cúbico. Apresente seus resultados graficamente.
19.21 Repita o Problema 19.15, mas use a rotina RCURV do IMSL.
19.22 Um corante é injetado em um volume de sangue circulante para medir a saída cardíaca de um paciente, que é a vazão em volume do sangue saindo do ventrículo esquerdo do coração. Em outras palavras, a saída cardíaca é o número de litros de sangue que seu coração bombeia em um minuto. Para uma pessoa em repouso, a taxa poderia ser de 5 ou 6 litros por minuto. Se você for um maratonista treinado correndo uma maratona, sua saída cardíaca pode

ser tão alta quanto 30 L/min. Os dados a seguir mostram a resposta de um indivíduo quando 5 mg de corante foi injetado em seu sistema venoso.

Tempo (s)	2	6	9	12	15	18	20	24
Concentração (mg/L)	0	1,5	3,2	4,1	3,4	2	1	0

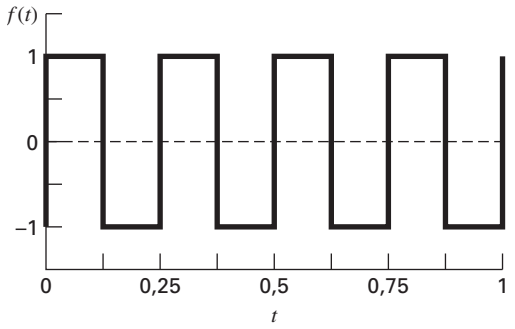


Figura P19.23

Ajuste uma curva polinomial a esses pontos dados e use a função para aproximar a saída cardíaca do paciente, que pode ser calculada por

$$\text{Saída cardíaca} = \frac{\text{quantidade de corante}}{\text{área sob a curva}}\left(\frac{\text{L}}{\text{min}}\right)$$

19.23 Em circuitos elétricos, é comum observar comportamento de corrente na forma de uma onda quadrada como a mostrada na Figura P19.23. Determinando a série de Fourier de

$$f(t) = \begin{cases} A_0 & 0 \le t \le T/2 \\ -A_0 & T/2 \le t \le T \end{cases}$$

obtemos a série de Fourier

$$f(t) = \sum_{n=1}^{\infty}\left(\frac{4A_0}{(2n-1)\pi}\right)\text{sen}\left(\frac{2\pi(2n-1)t}{T}\right)$$

Sejam $A_0 = 1$ e $T = 0{,}25$ s. Trace os primeiros seis termos da série de Fourier individualmente, bem como a soma desses seis termos. Use um pacote como o Excel ou o MATLAB se possível.

19.24 Faça um gráfico dos seguintes dados com (**a**) um polinômio interpolador de 6º grau, (**b**) um spline cúbico e (**c**) um spline cúbico com derivadas nulas nas extremidades.

x	0	100	200	400	600	800	1.000
$f(x)$	0	0,82436	1,00000	0,73576	0,40601	0,19915	0,09158

Em cada caso, compare seu gráfico com a seguinte equação, que foi usada para gerar os dados

$$f(x) = \frac{x}{200}e^{-\frac{x}{200}+1}$$

CAPÍTULO 20

Estudo de Casos: Ajuste de Curvas

O propósito deste capítulo é usar os métodos numéricos para ajuste de curvas para resolver alguns problemas de engenharia. A primeira aplicação, tirada da engenharia química, ilustra como um modelo não-linear pode ser linearizado e ajustado aos dados usando regressão linear. A segunda aplicação usa splines para estudar um problema relevante para a área ambiental da engenharia civil: aquecimento e transporte de massa em um lago estratificado.

A terceira aplicação ilustra como a transformada rápida de Fourier pode ser usada em engenharia elétrica para analisar um sinal determinando seus harmônicos principais. A aplicação final ilustra como a regressão linear múltipla é usada para analisar dados experimentais para problemas de fluidos tirados das engenharias mecânica e aeroespacial.

20.1 REGRESSÃO LINEAR E MODELOS POPULACIONAIS (ENGENHARIA QUÍMICA/BIOENGENHARIA)

Fundamentos. Os modelos de crescimento populacional são importantes em muitos campos da engenharia. É fundamental, para muitos dos modelos, a hipótese de que a taxa de variação da população (dp/dt) é proporcional à própria população (p) em qualquer instante (t), ou, de forma equivalente,

$$\frac{dp}{dt} = kp \tag{20.1}$$

onde k é o fator de proporcionalidade chamado de taxa de crescimento específica e tem unidade de tempo^{-1}. Se k for uma constante, então a solução da Equação (20.1) pode ser obtida da teoria de equações diferenciais:

$$p(t) = p_0 e^{kt} \tag{20.2}$$

onde p_0 é a população quando $t = 0$. Observa-se que $p(t)$ na Equação (20.2) tende a infinito quando t se torna grande. Esse comportamento é claramente impossível em sistemas reais. Portanto, o modelo precisa ser modificado para se tornar mais realista.

Solução. Primeiro, é preciso reconhecer que a taxa de crescimento específica k não pode ser constante quando a população se torna grande. Esse é o caso porque, quando p tende a infinito, o organismo que está sendo modelado passará a ser limitado por fatores como falta de alimento e produção de dejetos tóxicos. Uma forma de expressar isso matematicamente é usar um modelo de taxa de crescimento saturado

$$k = k_{\text{máx}} \frac{f}{K + f} \tag{20.3}$$

onde k_{max} é a *taxa máxima de crescimento atingível* para grandes valores de alimentos (f) e K é a *constante de meia saturação*. O gráfico da Equação (20.3) na Figura 20.1 mostra que, quando $f = K$, $k = k_{\text{máx}}/2$. Portanto, K é a quantidade de alimento disponível que sustenta uma taxa de crescimento populacional igual à metade da taxa máxima.

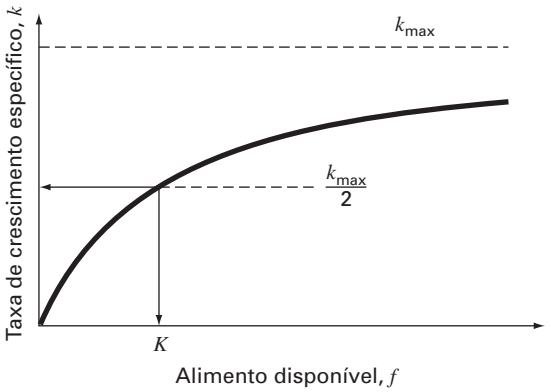


FIGURA 20.1
Gráfico da taxa de crescimento específico em função do alimento disponível para o modelo de taxa de crescimento saturado usado para caracterizar a cinética microbial. O valor K é chamado de constante de meia saturação porque corresponde à concentração na qual a taxa de crescimento específico é a metade de seu valor máximo.

As constantes K e k_{max} são valores empíricos baseados nas medidas experimentais de k para diversos valores de f. Como um exemplo, suponha que a população p represente o fermento empregado na produção comercial de cerveja e que f seja a concentração da fonte de carbono a ser fermentada. As medidas de k em função de f para o fermento estão mostradas na Tabela 20.1. É necessário calcular k_{max} e K a partir desses dados empíricos. Consegue-se isso invertendo a Equação (20.3) de maneira análoga à Equação (17.17) para obter

$$\frac{1}{k} = \frac{K + f}{k_{max} f} = \frac{K}{k_{max}} \frac{1}{f} + \frac{1}{k_{max}} \tag{20.4}$$

Com essa manipulação, transformamos a Equação (20.3) em uma forma linear, isto é, $1/k$ é uma função linear de $1/f$, com inclinação K/k_{max} e intersecção com o eixo y $1/k_{max}$. Esses valores estão traçados na Figura 20.2.

Por causa dessa transformação, os procedimentos dos mínimos quadrados linear descritos no Capítulo 17 podem ser usados para determinar $k_{max} = 1{,}23$ dia^{-1} e $K = 22{,}18$ mg/L. Esses resultados combinados com a Equação (20.3) são comparados aos dados não transformados na Figura 20.3 e quando substituídos no modelo da Equação (20.1) dão

$$\frac{dp}{dt} = 1{,}23 \frac{f}{22{,}18 + f} p \tag{20.5}$$

TABELA 20.1 Dados usados para calcular as constantes para um modelo de taxa de crescimento saturado para caracterizar a cinética microbial.

f, mg/L	k, dia^{-1}	$1/f$, L/mg	$1/k$, dia
7	0,29	0,14286	3,448
9	0,37	0,11111	2,703
15	0,48	0,06666	2,083
25	0,65	0,04000	1,538
40	0,80	0,02500	1,250
75	0,97	0,01333	1,031
100	0,99	0,01000	1,010
150	1,07	0,00666	0,935

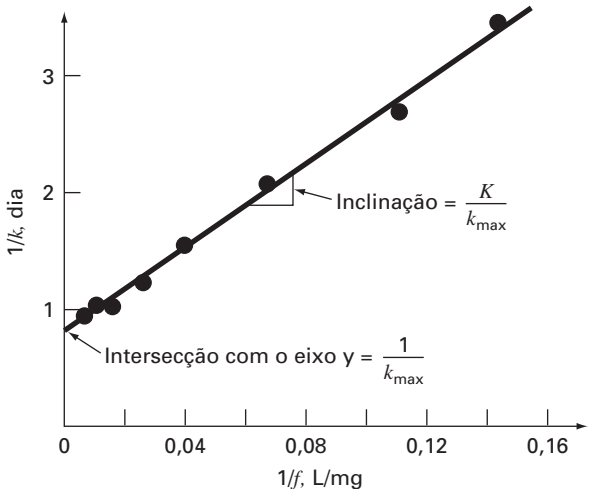


FIGURA 20.2
Versão linearizada do modelo de taxa de crescimento saturado. A reta é um ajuste por mínimos quadrados que é usado para calcular os coeficientes do modelo $k_{max} = 1{,}23$ dia^{-1} e $K = 22{,}18$ mg/L para um fermento usado na produção de cerveja.

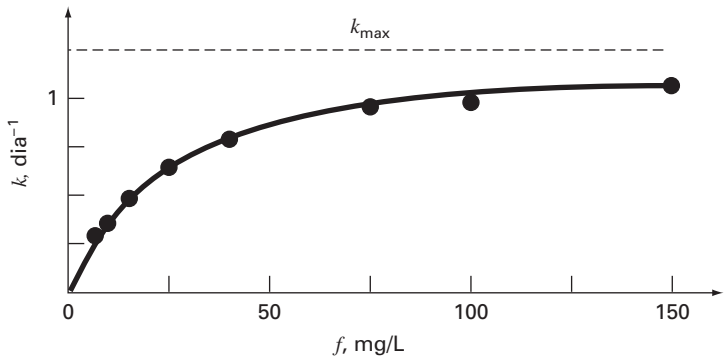


FIGURA 20.3
Ajuste do modelo de taxa de crescimento saturado para um fermento usado na produção comercial de cerveja.

Observe que o ajuste fornece uma soma dos quadrados dos resíduos (como calculado com os dados não transformados) de 0,001305.

A Equação (20.5) pode ser resolvida usando a teoria das equações diferenciais ou os métodos numéricos discutidos no Capítulo 25 quando $f(t)$ for conhecida. Se f tender a zero quando p se tornar grande, então dp/dt tende a zero e a população se estabiliza.

A linearização da Equação (20.3) é uma das maneiras de calcular as constantes k_{max} e K. Uma abordagem alternativa, que ajusta a relação em sua forma original, é a regressão não-linear descrita na Seção 17.5. A Figura 20.4 mostra como a ferramenta Solver do Excel pode ser usada para fazer uma estimativa dos parâmetros com regressão não-linear. Como pode ser visto, uma coluna de valores previstos foi desenvolvida com base no modelo e nas estimativas dos parâmetros. Estes são usados para gerar uma coluna de quadrado de resíduos que são somados e o resultado é colocado na célula D14. O Solver do Excel é então chamado para minimizar a célula D14 ajustando as células B1:B2. O resultado, como mostrado na Figura 20.4, fornece estimativas de $k_{max} = 1{,}23$ e $K = 22{,}14$, com um $S_r = 0{,}001302$. Logo, e — como esperado — a regressão não-linear forneça um ajuste ligeiramente melhor, os resultados são quase idênticos. Em outras aplicações, issto pode não ser verdadeiro (ou, a função pode não ser compatível com a linearização) e a regressão não-linear poderia representar a única opção possível para obter um ajuste por mínimos quadrados.

	A	B	C	D
1	kmax	1.2301		
2	K	22.1386		
3				
4	f	k	k-previsto	Res^2
5	7	0.29	**0.295508**	**0.000030**
6	9	0.37	0.355536	0.000209
7	15	0.48	0.496828	0.000283
8	25	0.65	0.652385	0.000006
9	40	0.8	0.791843	0.000067
10	75	0.97	0.949751	0.000410
11	100	0.99	1.007135	0.000294
12	150	1.07	1.071898	0.000004
13				
14			SSR	**0.001302**

=$B$1*A5/($B$2+A5)

=(B5-C5)^2

=SUM(D5..D12)

FIGURA 20.4
Regressão não-linear para ajustar o modelo de taxa de crescimento saturado para um fermento usado na produção comercial de cerveja.

20.2 USO DE SPLINES PARA ESTIMAR TRANSFERÊNCIA DE CALOR (ENGENHARIA CIVIL/AMBIENTAL)

Fundamentos. Os lagos na zona temperada podem tornar-se termicamente estratificados durante o verão. Como descrito na Figura 20.5, água quente flutuante perto da superfície está sobre a água do fundo, mais fria e mais densa. Tal estratificação divide de forma efetiva o lago verticalmente em duas camadas: o *epilímnio* e o *hipolímnio*, separadas por um plano chamado *termóclina*.

A estratificação térmica tem grande importância para os engenheiros ambientais que estudam a poluição de tais sistemas. Em particular, as termóclinas diminuem muito a mistura entre as duas camadas. Como resultado, a decomposição de matéria orgânica pode levar a uma diminuição severa de oxigênio na água isolada do fundo.

A posição da termóclina pode ser definida como o ponto de inflexão da curva temperatura-profundidade — isto é, o ponto no qual $d^2T/dx^2 = 0$. Ele também é o ponto no qual o valor absoluto da primeira derivada ou do gradiente é um máximo. Use splines cúbicos para determinar a profundidade da termóclina para o lago Platte (Tabela 20.2). Use também os splines para determinar o valor do gradiente na termóclina.

FIGURA 20.5
Temperatura em função da profundidade durante o verão no lago Platte, em Michigan.

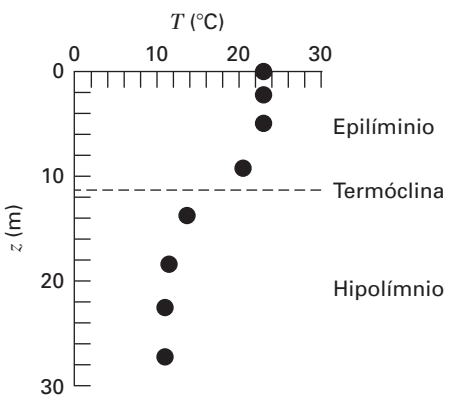


TABELA 20.2 Temperatura em função da profundidade durante o verão no lago Platte, em Michigan.

T, °C	22,8	22,8	22,8	20,6	13,9	11,7	11,1	11,1
z, m	0	2,3	4,9	9,1	13,7	18,3	22,9	27,2

Solução. Os dados são analisados com um programa que foi desenvolvido com base no pseudocódigo da Figura 18.18. Os resultados estão mostrados na Tabela 20.3, que lista as previsões do spline junto com as primeira e segunda derivadas em intervalos de 1 m por toda a coluna de água.

Os resultados estão traçados na Figura 20.6. Observe como a termóclina está claramente localizada em uma profundidade em que o gradiente é maior (isto é, o valor absoluto da derivada é maior) e a segunda derivada é nula. A profundidade é de 11,35 m e o gradiente nesse ponto é −1,61 ºC/m.

TABELA 20.3 Saída do programa para splines com base no pseudocódigo na Figura 18.18.

Profundidade (m)	T(C)	dT/dz	d2T/dz2	Profundidade (m)	T(C)	dT/dz	d2T/dz2
0.	22.8000	−.0115	.0000	15.	12.7652	−.6518	.3004
1.	22.7907	−.0050	.0130	16.	12.2483	−.3973	.2086
2.	22.7944	.0146	.0261	17.	11.9400	−.2346	.1167
3.	22.8203	.0305	−.0085	18.	11.7484	−.1638	.0248
4.	22.8374	−.0055	−.0635	19.	11.5876	−.1599	.0045
5.	22.7909	−.0966	−.1199	20.	11.4316	−.1502	.0148
6.	22.6229	−.2508	−.1884	21.	11.2905	−.1303	.0251
7.	22.2665	−.4735	−.2569	22.	11.1745	−.1001	.0354
8.	21.6531	−.7646	−.3254	23.	11.0938	−.0596	.0436
9.	20.7144	−1.1242	−.3939	24.	11.0543	−.0212	.0332
10.	19.4118	−1.4524	−.2402	25.	11.0480	.0069	.0229
11.	17.8691	−1.6034	−.0618	26.	11.0646	.0245	.0125
12.	16.2646	−1.5759	.1166	27.	11.0936	.0318	.0021
13.	14.7766	−1.3702	.2950	28.	11.1000	.0000	.0000
14.	13.5825	−.9981	.3923				

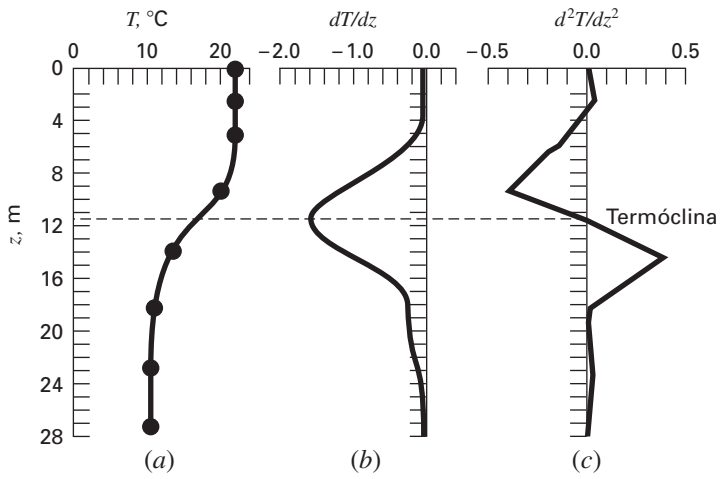


FIGURA 20.6
Gráficos (*a*) da temperatura, (*b*) do gradiente e (*c*) da segunda derivada em função da profundidade (*m*) gerados pelo programa para splines cúbicos. A termóclina está localizada no ponto de inflexão da curva de temperatura-profundidade.

20.3 ANÁLISE DE FOURIER (ENGENHARIA ELÉTRICA)

Fundamentos. A análise de Fourier é usada em muitas áreas da engenharia. Entretanto, é usada extensivamente nas aplicações de engenharia elétrica, como processamento de sinais.

Em 1848, Johann Rudolph Wolf inventou um método para quantificar a atividade solar contando o número de manchas individuais ou grupos de manchas na superfície do Sol. Ele calculou uma quantidade, agora chamada de *número de manchas solares de Wolf,* somando 10 vezes o número de grupos com a contagem individual de manchas. Como na Figura 20.7, os registros desse número remontam ao ano 1700. Com base nos primeiros registros históricos, Wolf determinou o comprimento do ciclo como sendo de 11,1 anos.

Use uma análise de Fourier para confirmar esse resultado aplicando uma FFT aos dados da Figura 20.7. Determine o período desenvolvendo um gráfico de potência em função do período.

Solução. Os dados para o ano e o número de manchas foram baixados da Web[1] e armazenados em um arquivo formatado por tabela: sunspot.dat. O arquivo pode ser carregado no software MATLAB, e a informação do ano e do número, inserida em vetores com o mesmo nome,

```
>> load sunspot.dat
>> year=sunspot(:,1);number=sunspot(:,2);
```

A seguir, uma FFT pode ser aplicada aos números de manchas

```
>> y=fft(number);
```

Depois de se livrar do primeiro harmônico, o comprimento da FFT está determinado (n) e a potência e a freqüência são calculadas,

```
>> y(1)=[ ];
>> n=length(y);
>> power=abs(y(1:n/2)).^2;
>> nyquist=1/2;
>> freq=(1:n/2)/(n/2)*nyquist;
```

Nesse ponto, o espectro de potência é um gráfico da potência em função da freqüência. Entretanto, como o período é mais significativo no presente contexto, pode-se determinar o período e um gráfico potência-período,

```
>> period=1./freq;
>> plot(period,power)
```

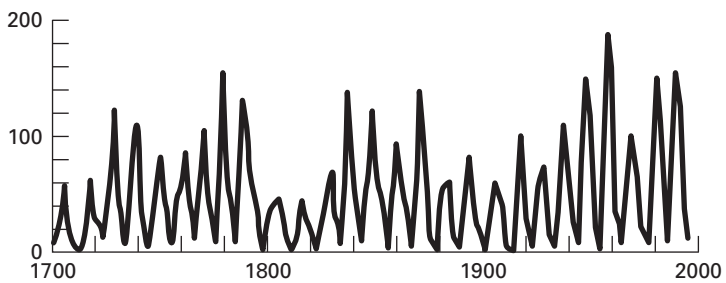


FIGURA 20.7
Gráfico do número de manchas solares de Wolf em função do ano.

[1]Na época da impressão deste livro, o html era http://www.ngdc.noaa.gov//stp/SOLAR/SSN/ssn.html.

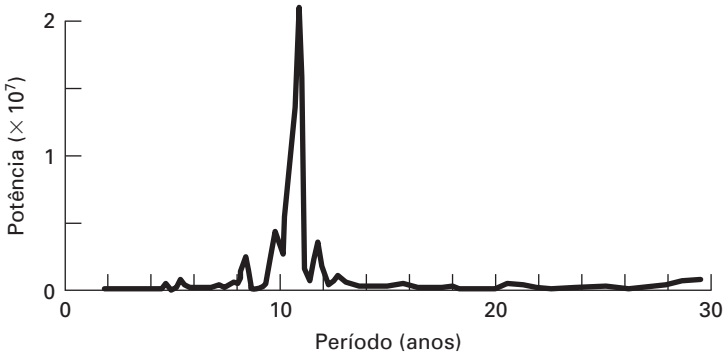


FIGURA 20.8
Espectro de potência do número de manchas solares de Wolf.

O resultado, como mostrado na Figura 20.8, indica um pico em cerca de 11 anos. O valor exato pode ser calculado com

```
>> index=find(power==max(power));
>> period(index)

ans=
    10.9259
```

20.4 ANÁLISE DE DADOS EXPERIMENTAIS (ENGENHARIA MECÂNICA/AEROESPACIAL)

Fundamentos. As variáveis de projeto em engenharia em geral dependem de diversas variáveis independentes. Com freqüência, essa dependência funcional é mais bem caracterizada por uma equação de potência em diversas variáveis. Como discutido na Seção 17.3, uma regressão linear múltipla dos dados transformados pelo log fornece um meio de calcular tais relações.

Por exemplo, um estudo de engenharia mecânica indica que o escoamento de um fluido através de um tubo está relacionado ao diâmetro do tubo e à inclinação (Tabela 20.4). Use regressão linear múltipla para analisar esses dados. Então, use o modelo resultante para prever o escoamento em um tubo com um diâmetro de 2,5 pés e uma inclinação de 0,025 pé/pé.

Solução. A equação de potência a ser calculada é

$$Q = a_0 D^{a_1} S^{a_2} \tag{20.6}$$

TABELA 20.4 Dados experimentais para o diâmetro, a inclinação e o escoamento em tubos circulares de concreto.

Experiência	Diâmetro, pés	Inclinação, pés/pé	Escoamento, pés^3/s
1	1	0,001	1,4
2	2	0,001	8,3
3	3	0,001	24,2
4	1	0,01	4,7
5	2	0,01	28,9
6	3	0,01	84,0
7	1	0,05	11,1
8	2	0,05	69,0
9	3	0,05	200,0

onde Q é o escoamento (pés^3/s), S é a inclinação (pés/pé), D é o diâmetro do tubo (pés), e a_0, a_1 e a_2 são coeficientes. Tomando o logaritmo dessa expressão, obtemos

$$\log Q = \log a_0 + a_1 \log D + a_2 \log S$$

Dessa forma, a equação é adequada à regressão linear múltipla porque log Q é uma função linear de log S e log D. Usando o logaritmo (base 10) dos dados na Tabela 20.4, pode-se gerar as seguintes equações normais expressas na forma matricial [lembre-se da Equação (17.22)]:

$$\begin{bmatrix} 9 & 2{,}334 & -18{,}903 \\ 2{,}334 & 0{,}954 & -4{,}903 \\ -18{,}903 & -4{,}903 & 44{,}079 \end{bmatrix} \begin{Bmatrix} \log a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 11{,}691 \\ 3{,}945 \\ -22{,}207 \end{Bmatrix}$$

Esse sistema pode ser resolvido usando-se eliminação de Gauss, o que dá

$$\log a_0 = 1{,}7475$$
$$a_1 = 2{,}62$$
$$a_2 = 0{,}54$$

Como log $a_0 = 1{,}7475$, então $a_0 = 10^{1{,}7475} = 55{,}9$ e a Equação (20.6) é

$$Q = 55{,}9 D^{2{,}62} S^{0{,}54} \tag{20.7}$$

A Equação (20.7) pode ser usada para prever o escoamento no caso em que $D = 2{,}5$ pés e $S = 0{,}025$ pé/pé, como em

$$Q = 55{,}9(2.5)^{2{,}62}(0{,}025) S^{0{,}54} = 84{,}1 \text{ pés}^3/\text{s}$$

Deve ser observado que a Equação (20.7) pode ser usada para outros propósitos além do cálculo do escoamento. Por exemplo, a inclinação está relacionada à perda de carga h_L e ao comprimento do tubo L por $S = h_L/L$. Se essa relação for substituída na Equação (20.7) e h_L for isolada na fórmula resultante, pode-se deduzir a seguinte equação:

$$h_L = \frac{L}{1721} Q^{1{,}85} D^{4{,}85}$$

Essa relação é chamada de *equação de Hazen-Williams*.

PROBLEMAS

Engenharia Química/Bioengenharia

20.1 Faça os mesmos cálculos que na Seção 20.1, mas use regressão linear e transformações para ajustar os dados com uma equação de potência. Avalie o resultado.

20.2 Você fez experiências e determinou os seguintes valores para a capacidade térmica c em várias temperaturas T para um gás:

T	−50	−30	0	60	90	110
c	1.270	1.280	1.350	1.480	1.580	1.700

Use regressão para determinar um modelo para prever c como função de T.

20.3 A concentração de saturação do oxigênio dissolvido na água como uma função da temperatura e da concentração de cloreto está listada na Tabela P20.3. Use interpolação para fazer uma estimativa do nível de oxigênio dissolvido para $T = 18$ °C com concentração de cloreto de 10 g/L.

20.4 Para os dados na Tabela P20.3, use regressão polinomial para obter uma equação de previsão de 3º grau para a concentração de oxigênio dissolvido como uma função da temperatura para o caso em que a concentração de cloreto é igual a 10 g/L. Use essa equação para fazer uma estimativa da concentração de oxigênio dissolvido para $T = 8$ °C.

20.5 Use regressão linear múltipla para deduzir uma equação de previsão para a concentração de oxigênio dissolvido como uma função da temperatura e do cloreto com base nos dados da Tabela P20.3. Use a equação para fazer uma estimativa da concentração de oxigênio dissolvido para uma concentração de cloreto de 5 g/L a $T = 17$ °C.

20.6 Quando comparado aos modelos dos Problemas 20.4 e 20.5, um modelo um pouco mais sofisticado que leva em conta os efeitos tanto da temperatura quanto do cloreto na saturação do oxigênio dissolvido pode ser considerado como sendo da forma

$$o_s = a_0 + f_3(T) + f_1(c)$$

Ou seja, é suposto que uma constante mais um polinômio de 3º grau na temperatura e uma relação linear no cloreto forneçam resultados superiores. Use a abordagem por mínimos quadrados linear geral para ajustar esse modelo aos dados da Tabela P20.3. Use a equação resultante para fazer uma estimativa da concentração de oxigênio dissolvido para uma concentração de clorito de 10 g/L a $T = 20$ °C.

TABELA P20.3 Concentração de oxigênio dissolvido na água como função da temperatura (°C) e da concentração de cloreto (g/L).

	Oxigênio Dissolvido (mg/L) para Temperatura (°C) e Concentração de Cloreto (g/L)		
T, °C	**$c = 0$ g/L**	**$c = 10$ g/L**	**$c = 20$ g/L**
0	14,6	12,9	11,4
5	12,8	11,3	10,3
10	11,3	10,1	8,96
15	10,1	9,03	8,08
20	9,09	8,17	7,35
25	8,26	7,46	6,73
30	7,56	6,85	6,20

20.7 É conhecido que a força de tensão do plástico aumenta em função do tempo quando ele é tratado com calor. São coletados os seguintes dados:

Tempo	10	15	20	25	40	50	55	60	75
Força de tensão	5	20	18	40	33	54	70	60	78

(a) Ajuste uma reta a esses dados e use a equação para determinar a força de tensão para o instante 32 minutos.
(b) Repita a análise, mas para uma reta com intersecção com o eixo y no zero.

20.8 Os seguintes dados foram reunidos para determinar a relação entre a pressão e a temperatura de um volume fixo de 1 kg de nitrogênio. O volume é 10 m^3.

T, °C	−40	0	40	80	120	160
p, N/m^2	6.900	8.100	9.300	10.500	11.700	12.900

Use a lei dos gases ideais $pV = nRT$ para determinar R com base nesses dados. Observe que, para a lei, T deve ser expressa em kelvins.

20.9 O volume específico de um vapor superaquecido está listado na tabelas de vapor para diversas temperaturas. Por exemplo, na pressão absoluta de 3.000 lb/pol^2:

T, °F	700	720	740	760	780
v, $pés^3/lb_m$	0,0977	0,12184	0,14060	0,15509	0,16643

Determine v para $T = 750$ °F.

20.10 Um reator está termicamente estratificado como na seguinte tabela:

Profundidade, m	0	0,5	1,0	1,5	2,0	2,5	3,0
Temperatura, °C	70	68	55	22	13	11	10

Como descrito na Figura P20.10, o tanque pode ser idealizado como duas zonas separadas por um forte gradiente de temperatura ou uma termóclina. A profundidade desse gradiente pode ser definida como o ponto de inflexão da curva temperatura-profundidade — isto é, o ponto no qual $d^2T/dz^2 = 0$. Nessa profundidade, o fluxo de calor da superfície para a camada do fundo pode ser calculado com a lei de Fourier

$$J = -k\frac{dT}{dz}$$

Use um ajuste por spline cúbico desses dados para determinar a profundidade da termóclina. Se $k = 0{,}02$ cal/(s · cm · °C), calcule o fluxo através da interface.

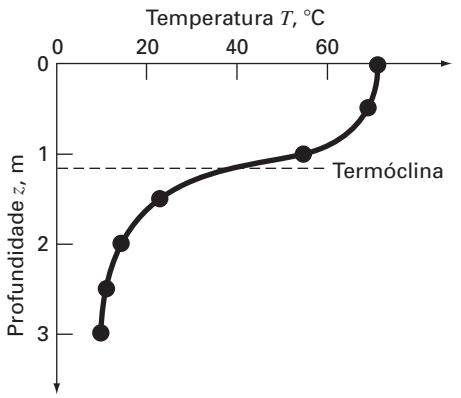


Figura P20.10

20.11 Na doença de Alzheimer, o número de neurônios no córtex diminui conforme a enfermidade progride. Os seguintes dados foram coletados para determinar o número de receptores neurotransmissores que restaram em um cérebro doente. Neurotransmissores livres ([F]) foram incubados com tecido e a concentração que se ligava especificamente a um receptor ([B]) foi medida. Quando a ligação é específica com um receptor, a concentração de ligação está relacionada com a concentração livre pela seguinte expressão:

$$[B] = \frac{B_{max}[F]}{K + [F]}$$

Usando os dados a seguir, determine os parâmetros que minimizam a soma dos quadrados dos resíduos. Além disso, calcule r^2.

[F], nM	0,1	0,5	1	5	10	20	50
[B], nM	10,57	36,61	52,93	82,65	89,46	94,35	101,00

20.12 Os seguintes dados foram tirados de um reator de mistura para a reação $A \rightarrow B$. Use os dados para determinar a melhor estimativa possível para k_{01} e E_1 para o seguinte modelo cinético,

$$-\frac{dA}{dt} = k_{01}e^{-\frac{E_1}{RT}}A$$

onde R é a constante do gás e é igual a 0,00198 Kcal/mol/K

$-dA/dt$ (moles/L/s)	460	960	2.485	1.600	1.245
A (moles/L)	200	150	50	20	10
T (K)	280	320	450	500	550

20.13 Use o seguinte conjunto de dados de pressão-volume para encontrar as melhores constantes possíveis (A_1 e A_2) para a equação de estado mostrada a seguir. $R = 82{,}05$ ml atm/gmol K e $T = 303$ K.

$$\frac{PV}{RT} = 1 + \frac{A_1}{V} + \frac{A_2}{V^2}$$

P (atm)	0,985	1,108	1,363	1,631
V (ml)	25.000	22.200	18.000	15.000

20.14 Dados de concentração foram tomados em 15 instantes de tempo para a reação de polimerização

$$xA + yB \rightarrow A_xB_y$$

Supomos que a reação ocorra via um mecanismo complexo que consiste em muitos passos. Diversos modelos foram considerados e as somas dos quadrados dos resíduos foram calculadas para os ajustes dos modelos aos dados. Os resultados estão mostrados a seguir. Qual modelo descreve melhor os dados (estatisticamente)? Explique sua escolha.

	Modelo A	**Modelo B**	**Modelo C**
S_r	135	105	100
Número de parâmetros do modelo no ajuste	2	3	5

20.15 A seguir estão os dados tomados do crescimento bacteriano em um reator em batelada (depois da fase de atraso terminar). Foi permitido que as bactérias crescessem tanto quanto possível nas primeiras 2,5 horas e então elas foram induzidas a produzir uma proteína recombinante, e essa produção diminui significativamente o crescimento bacteriano. O crescimento teórico das bactérias pode ser descrito por

$$\frac{dX}{dt} = \mu X$$

onde X é o número de bactérias e μ é a taxa de crescimento específico da bactéria durante o crescimento exponencial. Com base nos dados, qual a estimativa da taxa de crescimento específica da bactéria durante as primeiras 2 horas de crescimento? E durante as próximas quatro horas de crescimento?

Tempo, h	0	1	2	3	4	5	6
[Células], g/L	0,100	0,332	1,102	1,644	2,453	3,660	5,460

20.16 O peso molecular de um polímero pode ser determinado a partir de sua viscosidade pela seguinte relação:

$$[\eta] = KM_v^a$$

onde $[\eta]$ é a viscosidade intrínseca do polímero, M_v é o peso molecular médio e K e a são constantes específicas do polímero. A viscosidade intrínseca é determinada experimentalmente pela determinação do tempo de efluxo, ou o tempo que leva para a solução do polímero escoar entre duas linhas gravadas em um viscosímetro capilar, para diversas concentrações diferentes do polímero diluído e extrapolando para diluição infinita. Um gráfico de

$$\frac{\dfrac{t}{t_0} - 1}{c} \quad versus\ c$$

deveria fornecer uma reta, com a intersecção com o eixo y igual a $[\eta]$. A concentração da solução de polímero é c, t é o tempo de efluxo da solução de polímero e t_0 é o tempo de efluxo do solvente sem o polímero. Usando os dados a seguir para os tempos de efluxo de soluções diluídas de poliestireno em metil etil cetona a 25 °C e constantes $K = 3{,}9 \times 10^{-4}$ e $a = 0{,}58$, encontre o peso molecular da amostra de poliestireno.

Concentração de Polímero, g/dL	**Tempo de Efluxo, s**
0 (solvente puro)	83
0,04	89
0,06	95
0,08	104
0,10	114
0,12	126
0,14	139
0,16	155
0,20	191

20.17 Em média, a área A da superfície dos seres humanos está relacionada com o peso W e a altura H. Medidas em diversos indivíduos deram os valores de A na tabela a seguir:

H (cm)	182	180	179	187	189	194	195	193	200
W (kg)	74	88	94	78	84	98	76	86	96
A (m^2)	1,92	2,11	2,15	2,02	2,09	2,31	2,02	2,16	2,31

Determine uma equação para prever a área como uma função da altura e do peso. Use-a para fazer uma estimativa da área da superfície de uma pessoa com 187 cm e 78 kg.

20.18 Determine uma equação para prever a taxa de metabolismo como uma função da massa com base nos seguintes dados:

Animal	**Massa, kg**	**Metabolismo, watts**
Vaca	400	270
Humano	70	82
Ovelha	45	50
Galinha	2	4,8
Rato	0,3	1,45
Pombo	0,16	0,97

20.19 O sangue humano se comporta como um fluido newtoniano (ver Problema 20.51) na região de taxa de cisalhamento alta onde $\dot{\gamma} > 100$. Na região de taxa de cisalhamento baixa onde $\dot{\gamma} < 50$, as células vermelhas tendem a se agregar no que são chamados *roleaux,* o que faz o comportamento do fluido se afastar do newtoniano. Essa região de taxa de cisalhamento baixa é chamada de região de Casson, e existe uma região de transição entre as duas regiões distintas do escoamento. Na região de Casson, quando a taxa de cisalhamento tende a zero, a tensão de cisalhamento tende a um valor finito, análogo ao do plástico de Bingham, que é chamado tensão de cisalhamento inicial, τ_y, e essa tensão deve ser superada para iniciar o escoamento no sangue estagnado. O gráfico do escoamento na região de Casson é usualmente traçado como a raiz quadrada da taxa de cisalhamento em função da raiz quadrada da tensão de cisalhamento, e segue uma relação linear quando traçada dessa maneira. A relação de Casson é

$$\sqrt{\tau} = \sqrt{\tau_y} + K_c\sqrt{\dot{\gamma}}$$

onde K_c é o índice de consistência. Na tabela a seguir estão os valores medidos experimentalmente de $\dot{\gamma}$ e τ_y a partir de uma única amostra de sangue nas regiões de escoamento de Casson e newtoniana.

$\dot{\gamma}$, 1/s	0,91	3,3	4,1	6,3	9,6	23	36	49	65	105	126	215	315	402
τ, N/m^2	0,059	0,15	0,19	0,27	0,39	0,87	1,33	1,65	2,11	3,44	4,12	7,02	10,21	13,01
Região	Casson							Transição		Newtoniana				

Encontre os valores de K_c e τ_y usando regressão linear na região de Casson e determine μ usando regressão linear na região newtoniana. Encontre também o coeficiente de correlação para cada análise de regressão. Trace o gráfico das duas retas de regressão em um gráfico de Casson ($\sqrt{\gamma}$ *versus* $\sqrt{\tau}$) e estenda as retas de regressão como linhas tracejadas nas regiões vizinhas. Inclua também os pontos dados no gráfico. Limite a região de taxa de cisalhamento a $0 < \sqrt{\gamma} < 15$.

20.20 Tecidos moles seguem um comportamento de deformação exponencial sob tensão uniaxial enquanto estão em uma elongação fisiológica ou no intervalo normal. Isso pode ser expresso como

$$\sigma = \frac{E_0}{a}(e^{a\varepsilon} - 1)$$

onde σ é a tensão, ε é a deformação e E_o e a são constantes do material que são determinadas experimentalmente. Para determinar as duas constantes do material, a equação anterior é derivada em relação a ε. Usando essa equação, é estabelecida a relação fundamental para o tecido mole

$$\frac{d\sigma}{d\varepsilon} = E_0 + a\sigma$$

Para determinar E_o e a, os dados de tensão-deformação foram traçados como $d\sigma/d\varepsilon$ *versus* σ e a intersecção com o eixo y e a inclinação desse gráfico são as duas constantes do material, respectivamente.

Na tabela a seguir, estão os dados de tensão-deformação para o *chordae tendineae* do coração (pequenos tendões usados para manter as válvulas do coração fechadas durante a contração do músculo cardíaco; esses dados são da aplicação de carga ao tecido, enquanto curvas diferentes são produzidas quando são retiradas as cargas).

σ, 10^3 N/m^2	87,8	96,6	176	263	351	571	834	1.229	1.624	2.107	2.678	3.380	4.258
ε, 10^{-3} m/m	153	204	255	306	357	408	459	510	561	612	663	714	765

Calcule a derivada $d\sigma/d\varepsilon$ usando diferenças finitas. Faça um gráfico dos dados e elimine os pontos dados próximos do zero, que parecem não seguir a relação linear. O erro nesses dados vem da inabilidade dos instrumentos de ler os pequenos valores nessa região. Faça uma análise de regressão dos pontos restantes para determinar os valores de E_o e a.

Faça um gráfico da tensão *versus* dados de deformação juntamente com a curva analítica expressa pela primeira equação. Isso indicará quão bem a curva analítica ajusta os dados.

Muitas vezes isso não funciona bem porque o valor de E_o é difícil de determinar usando essa técnica. Para resolver o problema, E_o não é usado. Um ponto dado $(\bar{\sigma}, \bar{\varepsilon})$ que está no meio do intervalo da análise de regressão é escolhido. Esses valores são substituídos na primeira equação e um valor para E_o/a é determinado e substituído na primeira equação, que se torna

$$\sigma = \left(\frac{\bar{\sigma}}{e^{a\bar{\varepsilon}} - 1}\right)(e^{a\varepsilon} - 1)$$

Usando essa abordagem, dados experimentais que forem bem definidos produzirão uma boa concordância com os pontos dados e com a curva analítica. Use essa nova relação e trace novamente a tensão em função dos dados de deformação e essa nova curva analítica.

20.21 A espessura da retina varia em conseqüência de certas doenças do olho. Uma maneira de medir a espessura da retina é emitindo um *laser* de baixa energia na retina e registrando as reflexões em um filme. Por causa das propriedades óticas do olho, as reflexões da superfície frontal da retina e da superfície posterior da retina aparecerão como duas retas no filme, separadas por uma certa distância. A distância entre as retas no filme é proporcional à espessura da retina. A seguir estão os dados tomados de um filme examinado. Ajuste duas curvas com a forma gaussiana de altura e posição arbitrárias aos dados e determine a distância entre os centros dos dois picos. Uma curva gaussiana tem a forma

$$f(x) = \frac{ke^{-k^2(x-a)^2}}{\sqrt{\pi}}$$

onde k e a são constantes que se relacionam à altura do pico e ao centro do pico, respectivamente.

Posição	Intensidade da Luz	Posição	Intensidade da Luz	Posição	Intensidade da Luz	Posição	Intensidade da Luz
0,17	5,10	0,24	31,63	0,31	25,31	0,38	5,15
0,18	5,10	0,25	26,51	0,32	23,79	0,39	5,10
0,19	5,20	0,26	16,68	0,33	18,44	0,40	5,10
0,02	5,87	0,27	10,80	0,34	12,45	0,41	5,09
0,21	8,72	0,28	11,26	0,35	8,22	0,42	5,09
0,22	16,04	0,29	16,05	0,36	6,12	0,43	5,09
0,23	26,35	0,3	21,96	0,37	5,35	0,44	5,09

Engenharia Civil/Ambiental

20.22 As tensões de cisalhamento, em kilopascals (kPa), de nove amostras tomadas a várias profundidades em camadas de barro estão listadas a seguir. Estime a tensão de cisalhamento na profundidade de 4,5 m.

Profundidade, m	1,9	3,1	4,2	5,1	5,8	6,9	8,1	9,3	10,0
Tensão, kPa	14,4	28,7	19,2	43,1	33,5	52,7	71,8	62,2	76,6

20.23 Um estudo de engenharia de transporte foi conduzido para determinar um projeto adequado de faixas para bicicleta. Foram coletados dados sobre a largura das faixas para bicicleta e a distância média entre as bicicletas e os carros trafegando. Os dados de nove ruas são:

Distância, m	2,4	1,5	2,4	1,8	1,8	2,9	1,2	3	1,2
Largura da faixa, m	2,9	2,1	2,3	2,1	1,8	2,7	1,5	2,9	1,5

(a) Trace os dados.
(b) Ajuste uma reta aos dados, por regressão linear. Acrescente essa reta ao gráfico.
(c) Se for considerar a distância mínima média segura entre as bicicletas e os carros trafegando do que é 2 m, determine a largura mínima da faixa correspondente.

20.24 Na engenharia de fontes de água, a estimativa dos tamanhos dos reservatórios depende de estimativas acuradas do escoamento da água no rio que está sendo confinado. Para alguns rios, registros históricos de longa duração de tais dados de escoamento são difíceis de obter. Em contraste, em geral os dados meteorológicos sobre precipitação estão disponíveis para muitos anos do passado. Portanto, é útil determinar a relação entre escoamento e precipitação. Essa relação pode então ser usada para fazer uma estimativa do escoamento nos anos nos quais apenas medidas de precipitação foram feitas. Os seguintes dados estão disponíveis para um rio que deve ser estancado:

Precipitação, cm	88,9	108,5	104,1	139,7	127	94	116,8	99,1
Escoamento, m^3/s	14,6	16,7	15,3	23,2	19,5	16,1	18,1	16,6

(a) Trace os dados.
(b) Ajuste uma reta aos dados, por regressão linear. Superponha essa reta ao seu gráfico.
(c) Use a reta de melhor ajuste para prever o escoamento anual de água se a precipitação for 120 cm.
(d) Se a área de drenagem for 1.100 km^2, estime qual fração da precipitação será perdida por processos como evaporação, infiltração profunda da água no solo e uso destrutivo.

20.25 A concentração total de fósforo (p em mg/m^3) e clorofila a (c em mg/m^3) para cada um dos Grandes Lagos em 1970 era

	p	c
Lago Superior	4,5	0,8
Lago Michigan	8,0	2,0
Lago Huron	5,5	1,2
Lago Erie:		
Bacia Oeste	39,0	11,0
Bacia Central	19,5	4,4
Bacia Leste	17,5	3,8
Lago Ontário	21,0	5,5

A concentração de clorofila a indica quanta vida vegetal está suspensa na água. Como tal, ela indica quão não-clara e feia a água parece. Use os dados citados para determinar a expressão de c como uma função de p. Use essa equação para prever o nível de clorofila que pode ser esperado se for utilizado tratamento de dejetos para baixar a concentração de fósforo na parte oeste do lago Erie para 10 mg/m^3.

20.26 A tensão vertical σ_z sob o canto de uma área retangular sujeita a uma carga uniforme de intensidade q é dada pela solução da equação de Boussinesq:

$$\sigma = \frac{q}{4\pi}\left[\frac{2mn\sqrt{m^2+n^2+1}}{m^2+n^2+1+m^2n^2}\,\frac{m^2+n^2+2}{m^2+n^2+1} + \operatorname{sen}^{-1}\left(\frac{2mn\sqrt{m^2+n^2+1}}{m^2+n^2+1+m^2n^2}\right)\right]$$

Como é difícil resolver manualmente essa equação, ela foi reformulada como

$$\sigma_z = q f_z(m, n)$$

onde $f_z(m, n)$ é chamado valor de influência e m e n são razões adimensionais, com $m = a/z$ e $n = b/z$ e a e b como definidos na Figura P20.26. O valor de influência é então tabulado, uma parte dos quais está dada na Tabela P20.26. Se $a = 4{,}6$ e $b = 14$, use um polinômio interpolador de 3º grau para calcular σ_z a uma profundidade de 10 m abaixo do canto de um piso retangular que está submetido a uma carga total de 100 t (toneladas métricas). Expresse sua resposta em tones por metros quadrados. Observe que q é igual à carga por área.

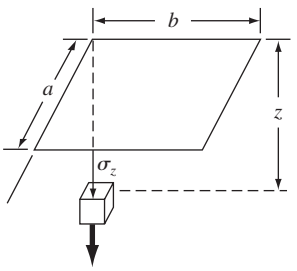


Figura P20.26

TABELA P20.26

m	$n = 1{,}2$	$n = 1{,}4$	$n = 1{,}6$
0,1	0,02926	0,03007	0,03058
0,2	0,05733	0,05894	0,05994
0,3	0,08323	0,08561	0,08709
0,4	0,10631	0,10941	0,11135
0,5	0,12626	0,13003	0,13241
0,6	0,14309	0,14749	0,15027
0,7	0,15703	0,16199	0,16515
0,8	0,16843	0,17389	0,17739

20.27 Três organismos transportadores de doenças decaem exponencialmente em água de lago de acordo com o seguinte modelo:

$$p(t) = Ae^{-1{,}5t} + Be^{-0{,}3t} + Ce^{-0{,}05t}$$

Estime a população inicial de cada organismo (A, B e C) dadas as seguintes medidas:

t, h	0,5	1	2	3	4	5	6	7	9
$p(t)$	6,0	4,4	3,2	2,7	2,2	1,9	1,7	1,4	1,1

20.28 O mastro de um barco a vela tem área da seção transversal 10,65 cm^2 e é construído de uma liga de alumínio experimental. Foram feitos testes para definir a relação entre a tensão e a deformação. Os resultados do teste são

Deformação, cm/cm	0,0032	0,0045	0,0055	0,0016	0,0085	0,0005
Tensão, N/cm^2	4.970	5.170	5.500	3.590	6.900	1.240

A tensão causada pelo vento pode ser calculada como F/A_c, onde F é a força no mastro e A_c é a área da seção transversal do mastro. Esse valor pode então ser substituído na lei de Hooke para determinar a deflexão no mastro, $\Delta L =$ deformação $\times L$, onde L é o comprimento do mastro. Se a força do vento for 25.000 N, use os dados para fazer uma estimativa da deflexão de um mastro de 9 m.

20.29 As reações enzimáticas são usadas extensivamente para caracterizar reações mediadas biologicamente em engenharia ambiental. Expressões propostas para a taxa de uma reação enzimática são dadas a seguir, onde $[S]$ é a concentração de substrato e v_0 é a taxa de reação inicial. Qual fórmula ajusta melhor os dados experimentais?

$$v_0 = k[S] \quad v_0 = \frac{k[S]}{K + [S]} \quad v_0 = \frac{k[S]^2}{K + [S]^2} \quad v_0 = \frac{k[S]^3}{K + [S]^3}$$

[S], M	**Taxa inicial, 10^{-6} M/s**
0,01	$6{,}3636 \times 10^{-5}$
0,05	$7{,}9520 \times 10^{-3}$
0,1	$6{,}3472 \times 10^{-2}$
0,5	6,0049
1	17,690
5	24,425
10	24,491
50	24,500
100	24,500

20.30 Os engenheiros ambientais, lidando com o impacto da chuva ácida, devem determinar o valor do produto iônico da água K_w como função da temperatura. Os cientistas sugeriram a seguinte equação para modelar essa relação:

$$-\log_{10} K_w = \frac{a}{T_a} + b \log_{10} T_a + cT_a + d$$

onde T_a é a temperatura absoluta (K), e a, b, c e d são parâmetros. Use os seguintes dados e regressão para fazer uma estimativa dos parâmetros:

T (K)	273,15	283,15	293,15	303,15	313,15
K_w	$1{,}164 \times 10^{-15}$	$2{,}950 \times 10^{-15}$	$6{,}846 \times 10^{-15}$	$1{,}467 \times 10^{-14}$	$2{,}929 \times 10^{-14}$

Engenharia Elétrica

20.31 Faça os mesmos cálculos que na Seção 20.3, mas analise dados gerados por $f(t) = 4\cos(5t) - 7\,\text{sen}(3t) + 6$.

20.32 Você mediu a queda de voltagem V através de um resistor para diversos valores diferentes da corrente i. Os resultados são

i	0,25	0,75	1,25	1,5	2,0
V	−0,45	−0,6	0,70	1,88	6,0

Use interpolação polinomial de primeiro a quarto graus para fazer uma estimativa da queda de voltagem para $i = 1{,}15$. Interprete seus resultados.

20.33 Repita os cálculos do Problema 20.32, mas use regressão polinomial para determinar as equações de melhor ajuste de primeiro a quarto graus utilizando todos os dados. Faça um gráfico de seus resultados e uma avaliação deles.

20.34 A corrente em um fio é medida com grande precisão como uma função do tempo:

t	0	0,1250	0,2500	0,3750	0,5000
i	0	6,24	7,75	4,85	0,0000

Determine i em $t = 0{,}23$.

20.35 Os seguintes dados foram tirados de uma experiência que mediu a corrente em um fio para várias voltagens impostas:

V, V	2	3	4	5	7	10
i, A	5,2	7,8	10,7	13	19,3	27,5

(a) Com base em uma regressão linear desses dados, determine a corrente para a voltagem de 3,5 V. Trace a reta e os dados e avalie o ajuste.

(b) Refaça a regressão e force a intersecção com o eixo y a ser zero.

20.36 É conhecido que a queda de voltagem através de um indutor segue a lei de Faraday:

$$V_L = L\frac{di}{dt}$$

onde V_L é a queda de voltagem (em volts), L é a indutância (em henrys; 1 H = 1 V· s/A) e i é a corrente (em ampères). Use os seguintes dados para fazer uma estimativa de L:

di/dt, A/s	1	2	4	6	8	10
V_L, V	5,5	12,5	17,5	32	38	49

Qual o significado, se houver algum, da intersecção com o eixo y da equação de regressão deduzida a partir desses dados?

20.37 A lei de Ohm afirma que a queda de voltagem V através de um resistor ideal é linearmente proporcional à corrente i passando pelo resistor como em $V = iR$, onde R é a resistência. Entretanto, resistores reais podem nem sempre obedecer à lei de Ohm. Suponha que você fez algumas experiências muito precisas para medir a queda de voltagem e a corrente correspondente para um resistor. Os resultados, listados na Tabela P20.37, sugerem uma relação curvilínea em vez da reta representada pela lei de Ohm. Para quantificar essa relação, uma curva deve ser ajustada aos dados. Por causa dos erros nas medidas, a regressão, tipicamente, seria o método preferido de ajuste de curva para analisar tais dados experimentais. Entretanto, o fato dessa relação ser lisa e a precisão dos métodos experimentais sugerem que a interpolação poderia ser apropriada. Use um polinômio interpolador de Newton para ajustar os dados e calcule V para $i = 0{,}10$. Qual é o grau do polinômio que foi usado para gerar os dados?

TABELA P20.37 Dados experimentais para a queda de voltagem através de um resistor sujeito a diversos níveis de corrente.

i	−2	−1	−0,5	0,5	1	2
V	−637	−96,5	−20,5	20,5	96,5	637

20.38 Repita o Problema 20.37, mas determine os coeficientes do polinômio (Seção 18.4) que ajusta os dados da Tabela P20.37.

20.39 É feita uma experiência para determinar a elongação porcentual de um material condutor elétrico como uma função da temperatura. Os dados resultantes estão listados a seguir. Faça uma previsão para a elongação porcentual para a temperatura de 400 °C.

Temperatura, °C	200	250	300	375	425	475	600
Elongação, %	7,5	8,6	8,7	10	11,3	12,7	15,3

20.40 As funções de Bessel aparecem com freqüência na análise avançada da engenharia tal como no estudo de campos elétricos. Essas funções usualmente não podem ser calculadas de forma simples e, portanto, em geral são compiladas em tabelas matemáticas padrão. Por exemplo,

x	1,8	2	2,2	2,4	2,6
$J_1(x)$	0,5815	0,5767	0,556	0,5202	0,4708

Estime $J_1(2,1)$, usando (**a**) um polinômio interpolador e (**b**) splines cúbicos. Observe que o valor verdadeiro é 0,568292.

20.41 A população (p) de uma pequena comunidade na periferia de uma cidade cresce rapidamente durante um período de 20 anos:

t	0	5	10	15	20
p	100	200	450	950	2.000

Como um engenheiro trabalhando para uma companhia de serviços públicos, você deve prever qual será a população daqui a cinco anos para antecipar a demanda por energia elétrica. Use um modelo exponencial e regressão linear para fazer essa previsão.

Engenharia Mecânica/Aeroespacial

20.42 Com base na Tabela 20.4, use interpolação linear e quadrática para calcular Q para $D = 1,23$ pé e $S = 0,001$ pé/pé. Compare seus resultados com o mesmo valor calculado com a fórmula deduzida na Seção 20.4.

20.43 Reproduza a Seção 20.4, mas determine uma equação para prever a inclinação como uma função do diâmetro e do escoamento. Compare seus resultados com a fórmula da Seção 20.4 e discuta seus resultados.

20.44 A viscosidade dinâmica da água $\mu(10^{-3}$ N· s/m^2) está relacionada com a temperatura T(ºC) da seguinte maneira:

T	0	5	10	20	30	40
μ	1,787	1,519	1,307	1,002	0,7975	0,6529

(a) Trace esses dados.
(b) Use interpolação para prever μ em $T = 7,5$ ºC.
(c) Use regressão polinomial para ajustar uma parábola aos dados e fazer a mesma previsão.

20.45 A *lei de Hooke*, que é válida quando uma mola não for esticada demais, significa que a extensão da mola e a força aplicada estão linearmente relacionadas. A proporcionalidade é parametrizada pela constante da mola k. Um valor para esse parâmetro pode ser estabelecido experimentalmente colocando-se pesos conhecidos sobre a mola e medindo-se a compressão resultante. Tais dados estão contidos na Tabela P20.45 e traçados na Figura P20.45. Observe que, acima de um peso de 40×10^4 N, a relação linear entre a força e o deslocamento deixa de ser válida. Esse tipo de comportamento é típico do que é chamado "enrijecimento da mola". Use regressão linear para determinar um valor de k para a parte linear desse sistema. Além disso, ajuste uma relação não-linear para a parte não-linear.

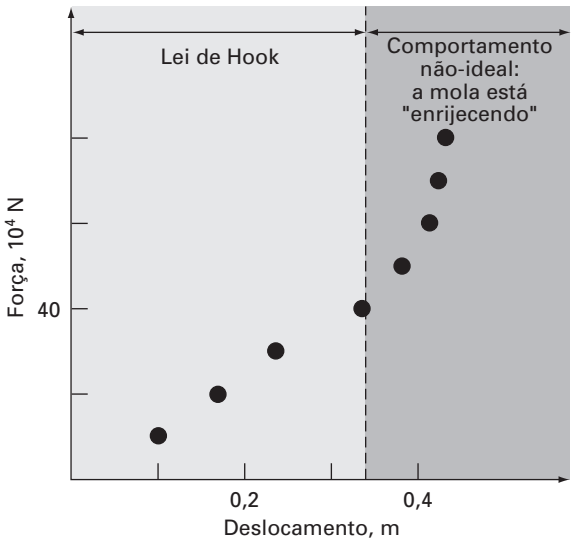


Figura P20.45
Gráfico da força (em 10^4 newtons) em função do deslocamento (em metros) para a mola de um sistema de suspensão de um automóvel.

20.46 Repita o Problema 20.45, mas ajuste uma curva de potência a todos os dados na Tabela P20.45. Comente seus resultados.

20.47 A distância necessária para parar um automóvel consiste nas componentes tanto de pensar como de frear, cada uma das quais é uma função de sua velocidade. Os seguintes dados experimentais foram coletados para quantificar essa relação. Determine uma equação de melhor ajuste para as componentes de pensar e frear. Use essas equações para fazer uma estimativa da distância de parada total para um carro movendo-se a 110 m/h.

Velocidade, m/h	30	45	60	75	90	120
Pensar, m	5,6	8,5	11,1	14,5	16,7	22,4
Frear, m	5,0	12,3	21,0	32,9	47,6	84,7

20.48 Uma experiência é feita para definir uma relação entre a tensão aplicada e o tempo de fratura para um tipo de aço inoxidável. Oito valores diferentes de tensão foram aplicados e os dados resultantes são

Tensão aplicada x, kg/mm^2	5	10	15	20	25	30	35	40
Tempo de fratura y, h	40	30	25	40	18	20	22	15

Trace os dados e então determine uma equação de melhor ajuste para prever o tempo de fratura para uma tensão aplicada de 20 kg/mm^2.

20.49 A aceleração da gravidade a uma altitude y acima da superfície da Terra é dada por

TABELA P20.45 Valores experimentais para a elongação x e para a força F em uma mola no sistema de suspensão de um carro.

Deslocamento, m	0,10	0,17	0,27	0,35	0,39	0,42	0,43	0,44
Força, 10^4 N	10	20	30	40	50	60	70	80

y, m	0	30.000	60.000	90.000	120.000
g, m/s^2	9,8100	9,7487	9,6879	9,6278	9,5682

Calcule g em $y = 55.000$ m.

20.50 A taxa de arrasto $\dot{\varepsilon}$ é a taxa no tempo pela qual a deformação aumenta e os dados de tensão a seguir foram obtidos de um procedimento de teste. Usando um ajuste por uma curva do tipo lei de potências,

$$\dot{\varepsilon} = B\sigma^m$$

determine os valores de B e m. Trace seus resultados usando uma escala log/log.

Taxa de arrasto, min^{-1}	0,0004	0,0011	0,0021	0,0031
Tensão, MPa	5,775	8,577	10,874	12,555

20.51 É uma prática comum, ao examinar o comportamento viscoso de um fluido, traçar a taxa de cisalhamento (gradiente da velocidade)

$$\frac{dv}{dy} = \dot{\gamma}$$

na abscissa *versus* a deformação de cisalhamento (τ) na ordenada. Quando um fluido tem um comportamento linear entre essas duas variáveis, é chamado *fluido newtoniano,* e a relação resultante é:

$$\tau = \mu\dot{\gamma}$$

onde μ é a viscosidade do fluido. Muitos fluidos comuns, como a água, o leite e o óleo, seguem esse comportamento. Fluidos que não se comportam dessa maneira são chamados *não-newtonianos*. Alguns exemplos de fluidos não-newtonianos são mostrados na Figura P20.51.

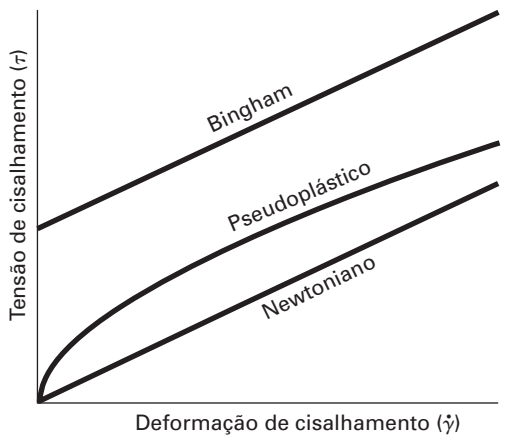


Figura P20.51

Para os *plásticos de Bingham*, existe uma tensão de cisalhamento inicial τ_y que deve ser vencida antes de o escoamento começar,

$$\tau = \tau_y + \mu\dot{\gamma}$$

Um exemplo comum é a pasta de dentes.

Para os *pseudoplásticos*, a taxa de cisalhamento é elevada à potência n,

$$\tau = \mu\dot{\gamma}^n$$

Exemplos comuns são o iogurte e o xampu.

Os seguintes dados mostram a relação entra a tensão de cisalhamento τ e a taxa de deformação de cisalhamento $\dot{\gamma}$ para um plástico de Bingham fluido. A tensão de cisalhamento inicial τ_y é a quantidade de tensão que deve ser vencida antes de o escoamento começar. Encontre a viscosidade μ (inclinação), τ_y, e o valor de r^2 usando o método da regressão.

Tensão τ, N/m^2	3,58	3,91	4,98	5,65	6,15
Taxa de deformação de cisalhamento $\dot{\gamma}$, 1/s	1	2	3	4	5

20.52 A relação entre a tensão τ e a taxa de deformação de cisalhamento $\dot{\gamma}$ para um pseudoplástico fluido (ver Problema 20.51) pode ser expressa pela equação $\tau = \mu\dot{\gamma}^n$. Os seguintes dados são de uma solução de 0,5% de hidroxi-etil-celulose em água. Usando um ajuste por uma lei de potência, encontre os valores de μ e n.

Taxa de deformação de cisalhamento $\dot{\gamma}$, 1/s	50	70	90	110	130
Tensão τ, N/m^2	6,01	7,48	8,59	9,19	10,21

20.53 A velocidade u do ar escoando por uma superfície plana é medida a diferentes distâncias y da superfície. Ajuste uma curva a esses dados, supondo que a velocidade seja zero na superfície ($y = 0$). Use seu resultado para determinar a tensão de cisalhamento ($\mu\ du/dy$) na superfície. ($\mu = 1{,}8 \times 10^{-5}$ N · s/m^2)

y, m	0,002	0,006	0,012	0,018	0,024
u, m/s	0,287	0,899	1,915	3,048	4,299

20.54 A *equação de Andrade* foi proposta como um modelo para o efeito da temperatura sobre a viscosidade,

$$\mu = De^{B/T_a}$$

onde μ é a viscosidade dinâmica da água (10^{-3} N · s/m^2), T_a é a temperatura absoluta (K), e D e B são parâmetros. Ajuste esse modelo aos dados para a água do Problema 20.44.

20.55 Determine equações para ajustar os calores específicos ideais c_p (kJ/kg · K), como uma função da temperatura T (K), para diversos gases, como listado na Tabela P20.55.

20.56 As temperaturas são medidas em diversos pontos de uma placa aquecida (Tabela P20.56). Faça uma estimativa da temperatura em **(a)** $x = 4$, $y = 3{,}2$ e **(b)** $x = 4{,}3$, $y = 2{,}7$.

20.57 Os dados na Tabela P20.57 foram obtidos de um teste de arrasto feito à temperatura ambiente em um fio composto de 40% de estanho, 60% de chumbo e solda sólida. Isso foi feito medindo-se o aumento na deformação ao longo do tempo enquanto uma carga constante era aplicada à amostra de teste. Usando o método de regressão linear, encontre (**a**) a equação da reta que melhor ajusta os dados e (**b**) o valor de r^2. Faça o gráfico de seus resultados. Na reta que passa pela origem — ou seja, no instante 0 — deveria haver alguma deformação? Se a reta não passar pela origem, force-a a fazer isso. Essa nova reta representa a tendência dos dados? Sugira uma nova equação que satisfaça a deformação zero no instante zero e que também represente bem a tendência dos dados.

TABELA P20.55 Calores específicos ideais, c_p (kJ/kg · K) como uma função da temperatura para diversos gases.

Gás	250 K	300 K	350 K	450 K	550 K	650 K	800 K	900 K	1000 K
H_2	14,051	14,307	14,427	14,501	14,53	14,571	14,695	14,822	14,983
CO_2	0,791	0,846	0,895	0,978	1,046	1,102	1,169	1,204	1,234
O_2	0,913	0,918	0,928	0,956	0,988	1,017	1,054	1,074	1,09
N_2	1,039	1,039	1,041	1,049	1,065	1,086	1,121	1,145	1,167

TABELA P20.56 Temperatura (°C) em diversos pontos de uma placa aquecida quadrada.

	$x = 0$	$x = 2$	$x = 4$	$x = 6$	$x = 8$
$y = 0$	100,00	90,00	80,00	70,00	60,00
$y = 2$	85,00	64,49	53,50	48,15	50,00
$y = 4$	70,00	48,90	38,43	35,03	40,00
$y = 6$	55,00	38,78	30,39	27,07	30,00
$y = 8$	40,00	35,00	30,00	25,00	20,00

TABELA P20.57

Tempo, min	Deformação, %	Tempo, min	Deformação, %	Tempo, min	Deformação, %
0,085	0,10	3,589	0,26	7,092	0,43
0,586	0,13	4,089	0,30	7,592	0,45
1,086	0,16	4,590	0,32	8,093	0,47
1,587	0,18	5,090	0,34	8,593	0,50
2,087	0,20	5,591	0,37	9,094	0,52
2,588	0,23	6,091	0,39	9,594	0,54
3,088	0,25	6,592	0,41	10,097	0,56



EPÍLOGO: PARTE CINCO

PT5.4 PRÓS E CONTRAS

A Tabela PT5.4 fornece um resumo dos prós e contras envolvidos no ajuste de curvas. As técnicas foram divididas em duas grandes categorias, dependendo das incertezas dos dados. Para medidas imprecisas, a regressão é usada para determinar uma "melhor" curva que ajuste a tendência geral dos dados sem necessariamente passar por nenhum ponto individual. Para medidas precisas, a interpolação é usada para determinar uma curva que passe diretamente por cada um dos pontos.

Todos os métodos de regressão são desenvolvidos para ajustar funções que minimizem a soma dos quadrados dos resíduos entre os dados e a função. Esses métodos são chamados regressão por mínimos quadrados. A regressão por mínimos quadrados linear é usada nos casos em que uma variável dependente e uma independente estão relacionadas entre si de uma forma linear. Para situações nas quais uma variável dependente e uma independente exibem uma relação curvilínea, diversas opções estão disponíveis. Em alguns casos, podem ser utilizadas transformações que linearizem a relação, e a regressão linear pode ser aplicada às variáveis transformadas para determinar a melhor reta. Alternativamente, a regressão polinomial pode ser usada para ajustar uma curva diretamente aos dados.

A regressão linear múltipla é utilizada quando uma variável dependente for uma função linear de duas ou mais variáveis independentes. Transformações logarítmicas também podem ser aplicadas a esse tipo de regressão para alguns casos nos quais a dependência múltipla é curvilínea.

TABELA PT5.4 Comparação das características dos métodos de ajustes de curvas.

Método	Erro Associado aos Dados	Coincidência com Pontos Individuais	Número de Pontos com Concordância Exata	Esforço de Programação	Comentários
Regressão					
Regressão linear	Grande	Aproximada	0	Fácil	
Regressão polinomial	Grande	Aproximada	0	Moderado	Os erros de arredondamento se tornam pronunciados para versões de grau mais alto
Regressão linear múltipla	Grande	Aproximada	0	Moderado	
Regressão não-linear	Grande	Aproximada	0	Difícil	
Interpolação					
Polinômio por diferenças divididas de Newton	Pequeno	Exata	$n + 1$	Fácil	Usualmente preferido para análise exploratória
Polinômios de Lagrange	Pequeno	Exata	$n + 1$	Fácil	Usualmente preferido quando o grau é conhecido
Splines cúbicos	Pequeno	Exata	Ajuste por partes dos pontos dados	Moderado	Primeira e segunda derivadas iguais nos nós

As regressões polinomial e linear múltipla (observe que uma regressão linear simples é um membro de ambas) pertencem a uma classe mais geral de modelos lineares de mínimos quadrados. Elas são classificadas dessa forma porque são lineares com relação a seus coeficientes. Esses modelos são tipicamente implementados usando sistemas algébricos lineares que são, algumas vezes, mal condicionados. Entretanto, em muitas aplicações de engenharia (isto é, para ajuste de grau mais baixo), isso não desempenha um papel importante. No caso de haver problemas, estão disponíveis abordagens alternativas. Por exemplo, uma técnica chamada polinômios ortogonais está disponível para fazer regressão polinomial (ver Seção PT5.6).

Equações que não são lineares com relação aos seus coeficientes são chamadas não-lineares. Técnicas especiais de regressão estão disponíveis para ajustar tais equações. Esses são métodos aproximados que começam com uma estimativa inicial do parâmetro e então, de modo iterativo, se aproximam dos valores que minimizam a soma dos quadrados.

A interpolação polinomial foi desenvolvida para ajustar um único polinômio de grau n que passe exatamente pelos $n + 1$ dados precisos. Esse polinômio é apresentado de duas formas alternativas. O polinômio interpolador por diferenças divididas de Newton idealmente é adequado para os casos nos quais o grau adequado do polinômio não é conhecido. O polinômio de Newton é apropriado para tais situações porque é facilmente programável em um formato que compara resultados com graus diferentes. Além disso, uma estimativa de erro pode ser incorporada de forma simples nessa técnica. Logo, você pode comparar e escolher a partir dos resultados usando polinômios de diversos graus diferentes.

O polinômio interpolador de Lagrange é uma formulação alternativa apropriada quando o grau é conhecido *a priori*. Para essas situações, a versão de Lagrange é um pouco mais simples de programar e não exige o cálculo e o armazenamento de diferenças divididas finitas.

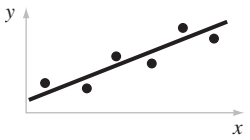
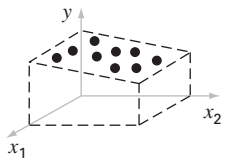
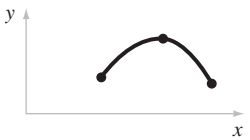
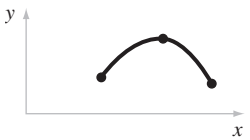
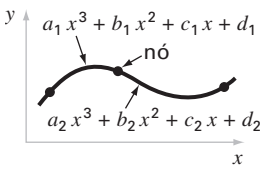
Outra abordagem para ajustar curvas é a interpolação por splines, técnica que ajusta um polinômio de grau baixo a cada intervalo entre os dados. O ajuste é tornado liso igualando-se as derivadas dos polinômios adjacentes no seu ponto de ligação. O spline cúbico é a versão mais comum. Os splines têm grande utilidade no ajuste de dados que no geral são lisos, mas exibem regiões locais de variação abrupta. Tais dados tendem a induzir grandes oscilações em polinômios interpoladores de grau mais alto. Os splines cúbicos são menos sujeitos a essas oscilações porque estão limitados a variações de terceira ordem.

O método final coberto nesta parte do livro é a aproximação de Fourier. Essa área trata do uso de funções trigonométricas para aproximar formas ondulatórias. Em contraste com outras técnicas, a ênfase principal dessa abordagem não é ajustar uma curva aos pontos dados. Em vez disso, a curva ajustada é usada para analisar as freqüências características de um sinal. Em particular, uma transformada rápida de Fourier está disponível para transformar de maneira muito eficiente uma função do domínio do tempo para o domínio da freqüência para elucidar sua estrutura harmônica subjacente.

PT5.5 RELAÇÕES E FÓRMULAS IMPORTANTES

A Tabela PT5.5 resume informações importantes que foram apresentadas na Parte Cinco. Essa tabela pode ser consultada para acessar rapidamente relações e fórmulas importantes.

TABELA PT5.5 Resumo das informações importantes apresentadas na Parte Cinco.

Método	Formulação	Interpretação Gráfica	Erros
Regressão linear	$y = a_0 + a_1 x$ onde $a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$ $a_0 = \bar{y} - a_1 \bar{x}$		$s_{y/x} = \sqrt{\frac{S_r}{n-2}}$ $r^2 = \frac{S_t - S_r}{S_t}$
Regressão polinomial	$y = a_0 + a_1 x + \cdots + a_m x^m$ (Cálculo dos a's equivalente à solução de $m + 1$ equações algébricas lineares)		$s_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}$ $r^2 = \frac{S_t - S_r}{S_t}$
Regressão linear múltipla	$y = a_0 + a_1 x_1 + \cdots + a_m x_m$ (Cálculo dos a's equivalente à solução de $m + 1$ equações algébricas lineares)		$s_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}$ $r^2 = \frac{S_t - S_r}{S_t}$
Polinômio interpolador por diferenças divididas de Newton*	$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1)$ onde $b_0 = f(x_0)$ $b_1 = f[x_1, x_0]$ $b_2 = f[x_2, x_1, x_0]$		$R_2 = (x - x_0)(x - x_1)(x - x_2)\frac{f^{(3)}(\xi)}{6}$ ou $R_2 = (x - x_0)(x - x_1)(x - x_2) f[x_3, x_2, x_1, x_0]$
Polinômio interpolador de Lagrange*	$f_2(x) = f(x_0)\left(\frac{x - x_1}{x_0 - x_1}\right)\left(\frac{x - x_2}{x_0 - x_2}\right) + f(x_1)\left(\frac{x - x_0}{x_1 - x_0}\right)\left(\frac{x - x_2}{x_1 - x_2}\right) + f(x_2)\left(\frac{x - x_0}{x_2 - x_0}\right)\left(\frac{x - x_1}{x_2 - x_1}\right)$		$R_2 = (x - x_0)(x - x_1)(x - x_2)\frac{f^{(3)}(\xi)}{6}$ ou $R_2 = (x - x_0)(x - x_1)(x - x_2) f[x_3, x_2, x_1, x_0]$
Splines cúbicos	Uma função cúbica: $a_i x^3 + b_i x^2 + c_i x + d_i$ é ajustada a cada intervalo entre os nós. Primeira e segunda derivadas iguais em cada nó.		

*Observação: por simplicidade, é mostrada a versão de 2º grau.

PT5.6 MÉTODOS AVANÇADOS E REFERÊNCIAS ADICIONAIS

Embora a regressão polinomial com equações normais seja adequada para muitas aplicações da engenharia, há contextos de problemas nos quais a sensibilidade a erros de arredondamento pode representar uma limitação séria. Uma abordagem alternativa baseada em *polinômios ortogonais* pode diminuir esses efeitos. Deve ser observado que essa abordagem não fornece uma equação de melhor ajuste, mas, em vez disso, fornece previsões individuais para valores dados para a variável independente. Informações sobre polinômios ortogonais podem ser encontradas em Shampine e Allen (1973) e Guest (1961).

A técnica dos polinômios ortogonais é útil no desenvolvimento de uma regressão polinomial, mas não representa uma solução para a instabilidade para o modelo de regressão linear geral [Equação (17.23)]. Uma abordagem alternativa baseada na *decomposição por valor singular,* chamada de método SVD, está disponível para esse propósito. Forsythe et al. (1977), Lawson e Hanson (1974) e Press et al. (1992) contêm informações sobre essa abordagem.

Além do algoritmo de Gauss-Newton, há diversos métodos de otimização que podem ser usados para desenvolver diretamente um ajuste por mínimos quadrados para uma equação não-linear. Essas técnicas de regressão não-linear incluem os métodos de Marquardt e do declive máximo (lembre-se da Parte Quatro). Informações gerais sobre regressão podem ser encontradas em Draper e Smith (1981).

Todos os métodos na Parte Cinco foram expostos em termos de ajustar uma curva a pontos dados. Além disso, você pode também querer ajustar uma curva a outra curva. O motivo principal para tal *aproximação funcional* é representar funções complicadas por uma versão mais simples que é mais fácil de manipular. Uma forma de fazer isso é usar a função complicada para gerar uma tabela de valores. Então, as técnicas discutidas nesta parte do livro podem ser usadas para ajustar polinômios a esses valores discretos.

Uma abordagem alternativa é baseada no *princípio minimax* (lembre-se da Figura 17.2*c*). Esse princípio especifica que os coeficientes do polinômio aproximador sejam escolhidos de modo que a discrepância máxima seja tão pequena quanto possível. Logo, embora a aproximação possa não ser tão boa quanto a dada pela série de Taylor no ponto-base, em geral é melhor ao longo do intervalo todo do ajuste. A economia de *Chebyshev* é um exemplo de uma abordagem para aproximação funcional baseada em tal estratégia (Ralston e Rabinowitz, 1978; Gerald e Wheatley, 1989; Carnahan, Luther e Wilkes, 1969).

Uma área importante no ajuste de curvas é a combinação de splines com regressão por mínimos quadrados. Assim, é gerado um spline cúbico que não passe por todos os pontos, mas, em vez disso, minimize a soma dos quadrados dos resíduos entre os pontos dados e as curvas do spline. A abordagem envolve o uso dos assim chamados *splines B* como funções-base. Eles levam esse nome por causa de seu uso como funções *base*, mas também por causa de sua forma característica de sino (*bell*). Tais curvas são consistentes com uma abordagem por splines, pois os seus valores e os de suas primeira e segunda derivadas seriam contínuos nas suas extremidades. Logo, a continuidade de $f(x)$ e de suas derivadas de ordem mais baixa é garantida nos nós. Wold (1974), Prenter (1975) e Cheney e Kincaid (1994) apresentam discussões sobre essa abordagem.

Em resumo, o objetivo da discussão anterior é fornecer a você avenidas para uma exploração mais profunda do assunto. Além disso, todas essas referências dadas fornecem descrições das técnicas básicas cobertas na Parte Cinco. Nós o encorajamos a consultar essas fontes alternativas para alargar seu entendimento dos métodos numéricos para ajuste de curvas.

PARTE SEIS



INTEGRAÇÃO E DERIVAÇÃO NUMÉRICAS

PT6.1 MOTIVAÇÃO

O cálculo é a matemática da variação. Como os engenheiros devem lidar continuamente com sistemas e processos que variam, o cálculo é uma ferramenta essencial de sua profissão. No cerne do cálculo estão os conceitos matemáticos relacionados de derivação (ou diferenciação) e integração.

De acordo com a definição do dicionário, *diferenciar* significa "marcar por diferenças; distinguir;... perceber a diferença entre". Matematicamente, a *derivada*, que serve como o veículo fundamental da derivação, representa a taxa de variação de uma variável dependente com relação a uma variável independente. Como mostrado na Figura PT6.1, a definição matemática da derivada começa com uma aproximação por diferenças:

$$\frac{\Delta y}{\Delta x} = \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x} \tag{PT6.1}$$

em que y e $f(x)$ são representações alternativas para a variável dependente e x é a variável independente. Se for permitido que Δx se aproxime de zero, como ocorre ao nos movermos da Figura PT6.1*a* para *c*, a diferença se torna a derivada

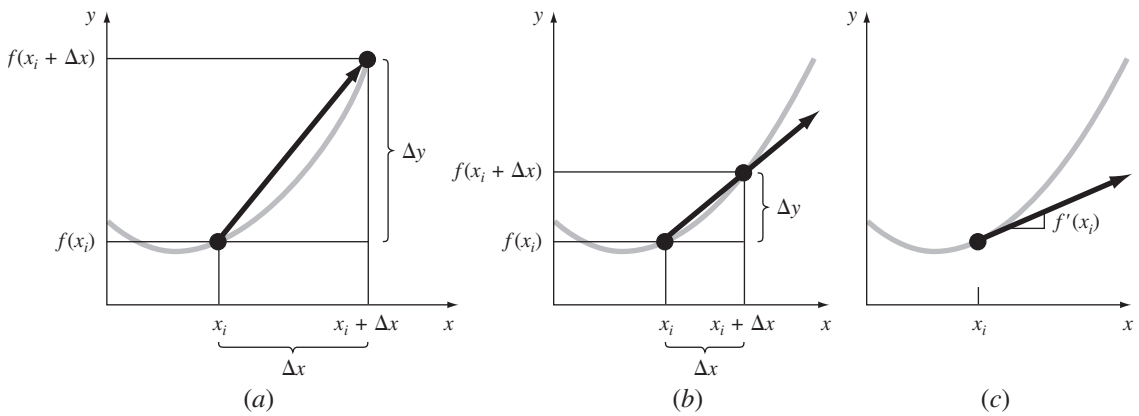
$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x}$$

em que dy/dx [que também pode ser designada por y' ou $f'(x_i)$] é a primeira derivada de y com relação a x, calculada em x_i. Como visto na descrição visual da Figura PT6.1*c*, a derivada é a inclinação da reta tangente à curva em x_i.

O processo inverso da derivação no cálculo é a integração. De acordo com a definição do dicionário, *integrar* significa "juntar as partes em um todo; unir, indicar a quantidade total...". Matematicamente, a integração é representada por

$$I = \int_a^b f(x)\,dx \tag{PT6.2}$$

FIGURA PT6.1
A definição gráfica de uma derivada: conforme Δx se aproxima de zero ao passar de (*a*) para (*c*), a aproximação por diferença se torna a derivada.



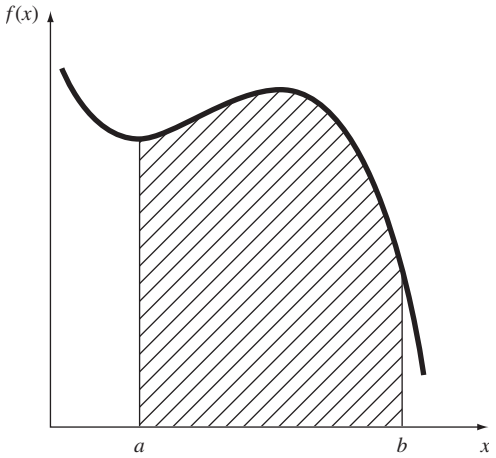


FIGURA PT6.2
Representação gráfica da integral de $f(x)$ entre os extremos $x = a$ e b. A integral é equivalente à área abaixo da curva.

o que indica a integral de uma função $f(x)$ em relação à variável independente x, calculada entre os extremos $x = a$ e $x = b$. A função $f(x)$ na Equação (PT6.2) é conhecida como o *integrando*.

Como sugerido pela definição do dicionário, o "significado" da Equação (PT6.2) é o *valor total* ou a *soma* de $f(x)\,dx$ no intervalo de $x = a$ a b. De fato, o símbolo $\int$ é, na realidade, um S maiúsculo estilizado usado para indicar a conexão íntima entre integração e soma.

A Figura PT6.2 representa uma descrição gráfica do conceito. Para funções que estão acima do eixo x, a integral expressa pela Equação (PT6.2) corresponde à área sob a curva de $f(x)$ entre $x = a$ e b.[1]

Como já delineado, a "separação" ou "discriminação" da derivação e o "juntar" da integração são processos intimamente ligados, que são, de fato, inversamente relacionados (Figura PT6.3). Por exemplo, se for dada uma função $y(t)$ que especifica a posição de um objeto como uma função do tempo, a derivação fornece um meio de determinar sua velocidade, como em (Figura PT6.3*a*):

$$v(t) = \frac{d}{dt}y(t)$$

Reciprocamente, se for dada a velocidade como uma função do tempo, a integração pode ser usada para determinar sua posição (Figura PT6.3*b*),

$$y(t) = \int_0^t v(t)\,dt$$

Logo, pode-se fazer a afirmação geral de que o cálculo da integral

$$I = \int_a^b f(x)\,dx$$

é equivalente à resolução da equação diferencial

$$\frac{dy}{dx} = f(x)$$

para a determinação de $y(b)$ dada a condição inicial $y(a) = 0$.

[1] Deve ser observado que o processo representado pela Equação (PT6.2) e pela Figura PT6.2 é chamado *integração definida*. Existe um outro tipo chamado integração indefinida, no qual os extremos a e b não são especificados. Como será discutido na Parte Sete, a *integração indefinida* trata da determinação de uma função cuja derivada é dada.

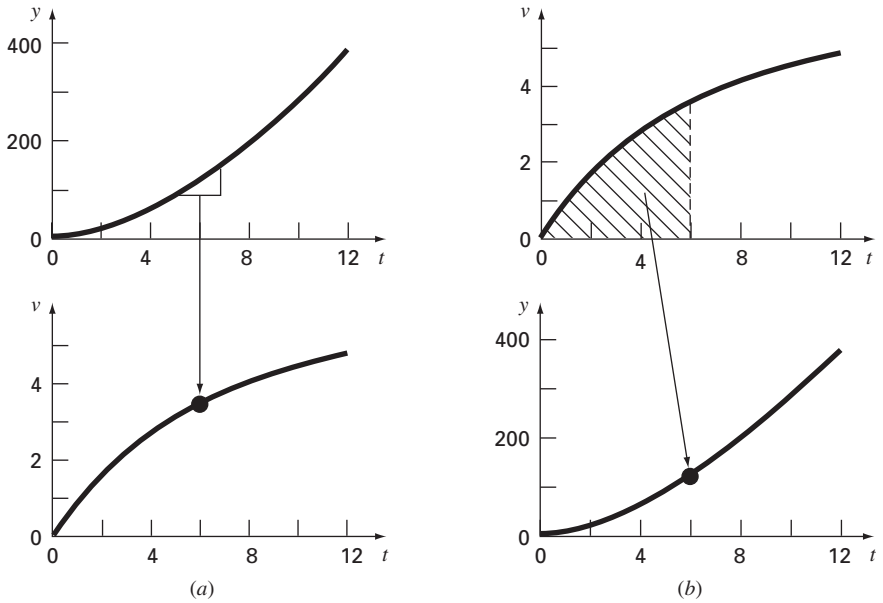


FIGURA PT6.3
O contraste entre (*a*) derivação e (*b*) integração.

Por causa de sua relação íntima, optou-se por dedicar esta parte do livro a ambos os processos. Entre outras coisas, isso fornecerá a oportunidade de destacar suas semelhanças e diferenças de uma perspectiva numérica. Além disso, a discussão será relevante para as próximas partes do livro nas quais cobriremos as equações diferenciais.

PT6.1.1 Métodos Não-Computacionais para Diferenciação e Integração

A função a ser diferenciada ou integrada estará tipicamente em uma das três formas:

1. Uma função no contínuo simples tal como um polinômio, uma exponencial ou uma função trigonométrica.
2. Uma função no contínuo complicada que é difícil ou impossível de derivar ou integrar diretamente.
3. Uma função tabulada em que valores de x e $f(x)$ são dados em vários pontos discretos, como em geral é o caso com dados experimentais ou de campo.

No primeiro caso, a derivada ou integral de uma função simples pode ser calculada analiticamente usando-se o cálculo. Para o segundo caso, as soluções analíticas são freqüentemente não práticas e, em alguns casos, impossíveis de obter. Nesses casos, bem como no terceiro caso, de dados discretos, devem ser empregados métodos aproximados.

Um método não-computacional para determinar derivadas dos dados é chamado de *derivação gráfica por áreas iguais*. Nesse método, os dados (x, y) são tabulados e, para cada intervalo, uma diferença dividida simples $\Delta y/\Delta x$ é usada para fazer uma estimativa da inclinação. Então, esses valores são colocados em um gráfico com degraus, em função de x (Figura PT6.4). A seguir, é desenhada uma curva lisa que tenta aproximar a área sob a curva em degraus. Isto é, ela é desenhada de modo que, visualmente, as áreas positivas e negativas fiquem balanceadas. As taxas nos valores dados de x podem então ser lidas na curva.

No mesmo espírito, abordagens visualmente orientadas foram usadas para integrar dados tabulados e funções complicadas na era pré-computador. Uma abordagem intuitiva simples é traçar a função em um reticulado (Figura PT6.5) e contar o número de retângulos que aproximam a área. Esse número é multiplicado pela área de cada retângulo, fornecendo uma estimativa grosseira da área sob a curva. Essa estimativa pode ser refinada, com o custo de um esforço adicional, usando-se um reticulado mais fino.

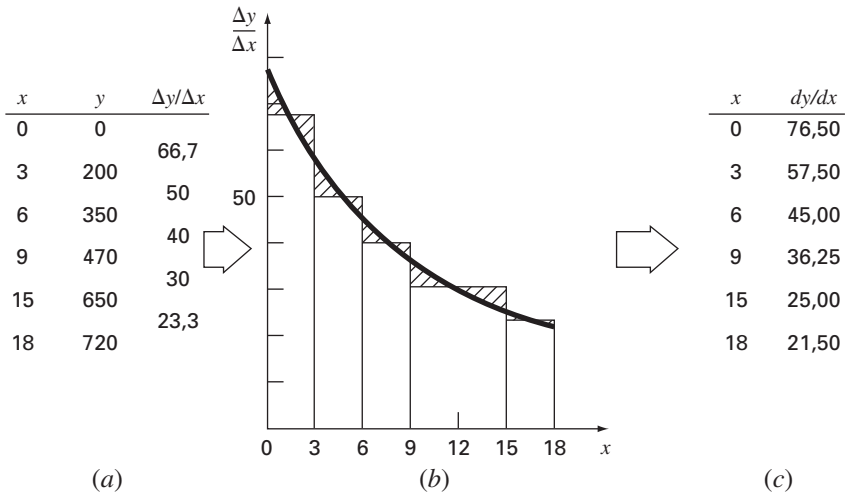


FIGURA PT6.4
Derivação por áreas iguais. (*a*) Diferenças divididas finitas centrais são usadas para obter estimativas da derivada para cada intervalo entre os pontos dados. (*b*) As estimativas das derivadas são traçadas em um gráfico de barras. Uma curva lisa é superposta a esse gráfico para aproximar a área sob o gráfico de barras. Isso é conseguido desenhando-se a curva de modo que áreas iguais positivas e negativas estejam balanceadas. (*c*) Valores de *dy*/*dx* podem então ser lidos na curva lisa.

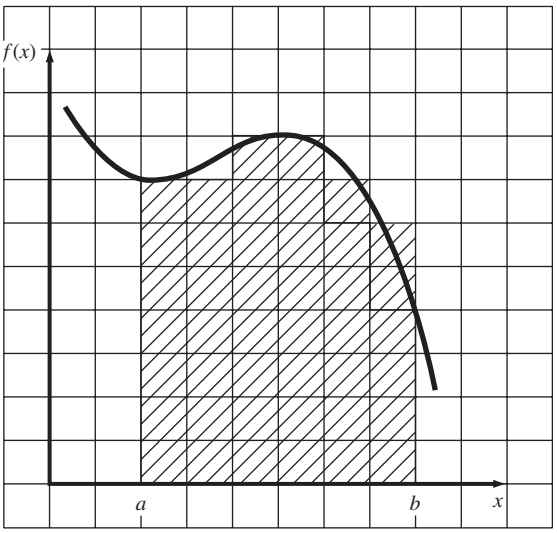


FIGURA PT6.5
O uso de um reticulado para obter uma aproximação para uma integral.

Uma outra abordagem pelo bom senso é dividir a área em segmentos verticais ou faixas, com uma altura igual ao valor da função no ponto médio de cada faixa (Figura PT6.6). A área dos retângulos pode então ser calculada e somada para se obter uma estimativa da área total. Nessa abordagem, é suposto que o valor no ponto médio fornece uma aproximação válida da altura média da função em cada faixa. Como no caso do reticulado, estimativas refinadas são possíveis usando-se mais faixas (e mais finas) para aproximar a integral.

Embora tais abordagens simples tenham utilidade para estimativas rápidas, técnicas numéricas alternativas estão disponíveis para o mesmo propósito. Não surpreendentemente, o mais simples desses métodos é parecido, de forma geral, às técnicas não-computacionais.

Para a derivação, a técnica numérica mais fundamental usa diferenças divididas finitas para obter estimativas das derivadas. Para dados com erros, uma abordagem alternativa é ajustar uma curva lisa aos dados com uma técnica como a regressão por mínimos quadrados e então derivar essa curva para obter estimativas para as derivadas.

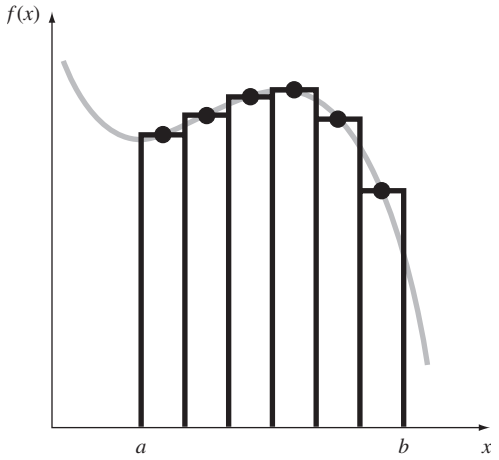


FIGURA PT6.6
O uso de retângulos ou faixas para aproximar uma integral.

De forma parecida, a integração numérica ou métodos de *quadratura* estão disponíveis para se obter integrais. Tais métodos, que são, na realidade, mais fáceis de implementar do que a abordagem por reticulado, são parecidos com o método da faixa. Isto é, as alturas da função são multiplicadas pelas larguras das faixas e somadas para se obter uma estimativa da integral. Entretanto, com escolhas inteligentes de fatores de peso, a estimativa resultante pode-se tornar mais acurada que aquela do modelo de faixas simples.

Como no método das faixas simples, as técnicas de derivação e integração numéricas usam dados em pontos discretos. Como a informação tabulada já está em tal forma, ela é naturalmente compatível com muitas das abordagens numéricas. Embora as funções no contínuo não estejam originalmente na forma discreta, em geral é uma tarefa simples usar a equação dada para gerar uma tabela de valores. Como mostrado na Figura PT6.7, essa tabela pode ser analisada por um método numérico.

PT6.1.2 Derivação e Integração Numéricas em Engenharia

A derivação e a integração de uma função têm tantas aplicações em engenharia que foi necessário que você fizesse cursos de cálculo diferencial e integral em seu primeiro ano na faculdade. Muitos exemplos específicos de tais aplicações podem ser dados em todos os campos da engenharia.

A derivação é rotineira na engenharia porque muito de nosso trabalho envolve a caracterização da mudança das variáveis no tempo e no espaço. De fato, muitas das leis e outras generalizações que desempenham um papel tão importante em nosso trabalho são baseadas na forma previsível pelas quais as variações se manifestam no mundo físico. Um exemplo importante é a segunda lei de Newton, que não é enunciada em termos da posição de um objeto, mas, em vez disso, em termos da variação da posição com relação ao tempo.

Além de tais exemplos temporais, diversas leis que governam o comportamento espacial das variáveis são expressas em termos de derivadas. Entre as mais comuns estão as leis envolvendo potenciais ou gradientes. Por exemplo, a *lei de Fourier de condução de calor* quantifica a observação de que o calor flui de regiões de alta para regiões de baixa temperatura.

$$\text{Fluxo de calor} = -k' \frac{dT}{dx}$$

Logo, a derivada fornece uma medida da intensidade da variação de temperatura, ou do *gradiente*, que governa a transferência de calor. Leis parecidas fornecem modelos práticos em muitas outras áreas da engenharia, incluindo a modelagem da mecânica dos fluidos, a transferência de massa e o fluxo eletromagnético. A habilidade de obter estimativas acuradas para as derivadas é uma faceta importante em nossa capacidade de trabalhar eficientemente nessas áreas.

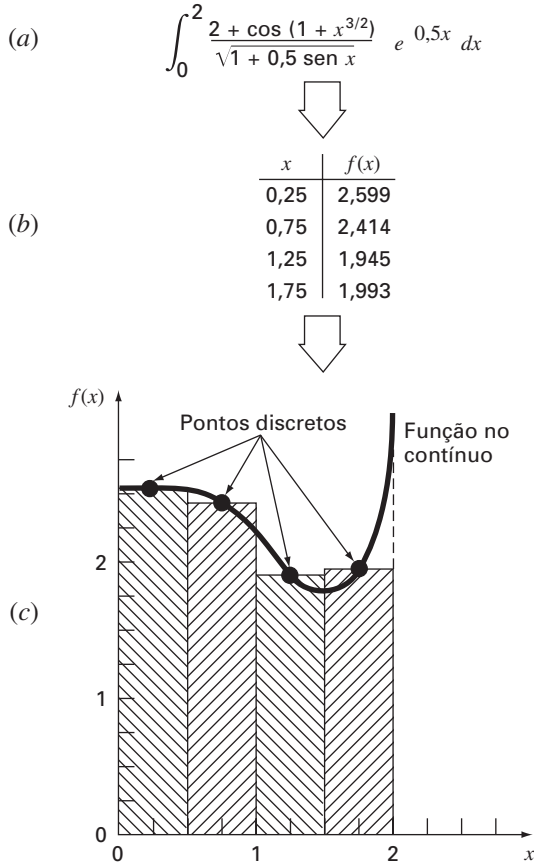
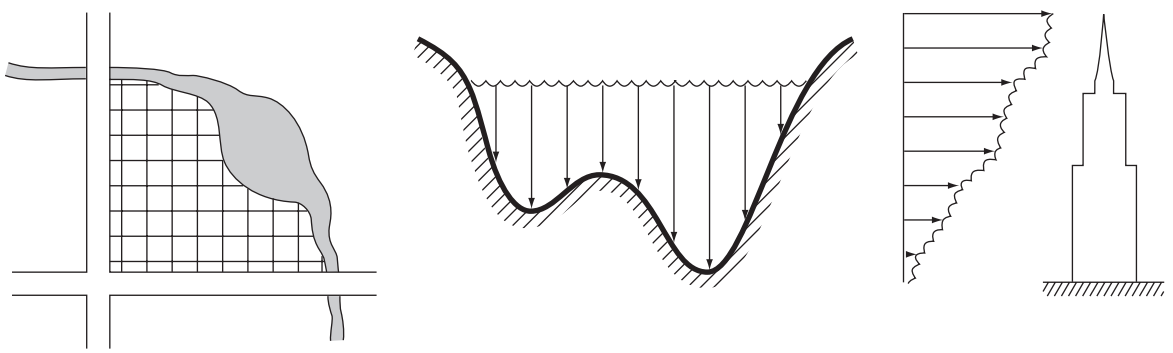


FIGURA PT6.7
Aplicação de um método de integração numérica: (*a*) Uma função no contínuo complicada. (*b*) Tabela de valores discretos de *f*(*x*) gerada a partir da função. (*c*) Uso de um método numérico (aqui, o método das faixas) para obter uma estimativa da integral com base nos pontos discretos. Para uma função tabulada, os dados já estão na forma tabular (*b*); portanto, o passo (*a*) não é necessário.

Do mesmo modo como estimativas acuradas das derivadas são importantes na engenharia, o cálculo de integrais é igualmente valioso. Diversos exemplos relacionam diretamente a idéia de integral com a área sob a curva. A Figura PT6.8 descreve uns poucos casos nos quais a integração é usada para esse propósito.

FIGURA PT6.8
Exemplos de como a integração é usada para calcular áreas em aplicações de engenharia. (*a*) Um inspetor pode precisar saber a área de um campo limitado por um riacho sinuoso e duas estradas. (*b*) Um engenheiro de recursos hidráulicos pode precisar saber a área da seção transversal de um rio. (*c*) Um engenheiro de estruturas poderia precisar determinar a força média decorrente de um vento não-uniforme soprando contra o lado de um arranha-céu.



(*a*) (*b*) (*c*)

Outra aplicação comum da integral se relaciona com a analogia entre integração e soma. Por exemplo, uma aplicação comum é determinar a média de uma função contínua. Na Parte Cinco, foi apresentado o conceito de média de n pontos discretos [lembre-se da Equação (PT5.1)]:

$$\text{Média} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{PT6.3}$$

em que y_i são medidas individuais. A determinação da média de pontos discretos é ilustrada na Figura PT6.9*a*.

Em contraste, suponha que y seja uma função no contínuo de uma variável independente x, como descrito na Figura PT6.9*b*. Nesse caso, existe um número infinito de valores entre a e b. Do mesmo modo que a Equação (PT6.3) pode ser aplicada para determinar a média de leituras discretas, você poderia também estar interessado em calcular a média de uma função no contínuo $y = f(x)$ no intervalo de a até b. A integração é usada para esse propósito, como especificado pela fórmula

$$\text{Média} = \frac{\int_a^b f(x)\,dx}{b - a} \tag{PT6.4}$$

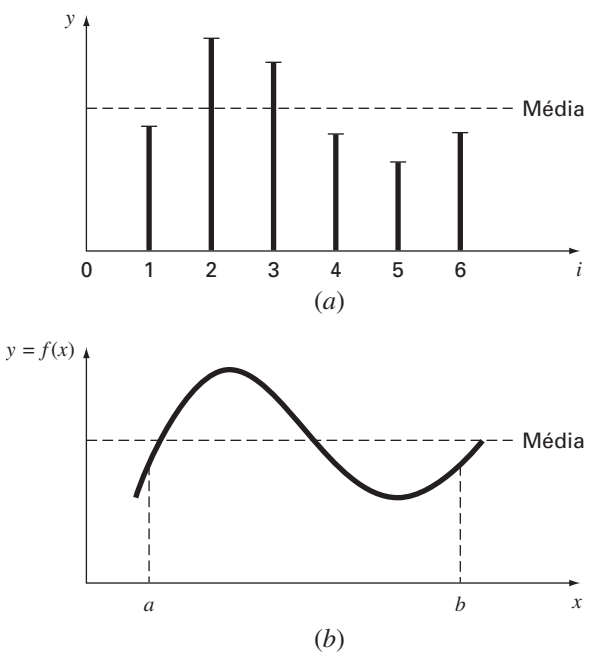
Essa fórmula tem centenas de aplicações em engenharia. Por exemplo, é usada para calcular o centro de gravidade de objetos irregulares na mecânica e na engenharia civil e para determinar a corrente eficaz na engenharia elétrica.

As integrais também são usadas pelos engenheiros para calcular a quantidade total de uma dada variável física. A integral pode ser calculada sobre uma reta, uma área ou um volume. Por exemplo, a massa total de um composto químico em um reator é dada pelo produto da concentração do composto químico pelo volume do reator, ou

$$\text{Massa} = \text{concentração} \times \text{volume}$$

em que a concentração tem unidades de massa por volume. Entretanto, suponha que a concentração varie de uma posição para outra dentro do reator. Nesse caso, é necessário

FIGURA PT6.9
Uma ilustração da média para dados (*a*) discretos e (*b*) contínuos.



somar o produto das concentrações locais c_i pelos elementos de volumes correspondentes ΔV_i:

$$\text{Massa} = \sum_{i=1}^{n} c_i \, \Delta V_i$$

em que n é o número de volumes discretos. Para o caso contínuo, em que $c(x, y, z)$ é uma função conhecida e x, y e z são as variáveis independentes que indicam a posição em coordenadas cartesianas, a integração pode ser usada para o mesmo propósito:

$$\text{Massa} = \iiint c(x, y, z)\, dx\, dy\, dz$$

ou

$$\text{Massa} = \iiint_V c(V)\, dV$$

que é conhecida como uma *integral de volume*. Observe a forte analogia entre a soma e a integração.

Exemplos parecidos poderiam ser dados em outros campos da engenharia. Por exemplo, a taxa total de transferência de energia através de um plano no qual o fluxo (em calorias por centímetro quadrado) é uma função da posição é dada por

$$\text{Transferência de calor} = \iint_A \text{fluxo}\, dA$$

que é conhecida como uma *integral de área*, em que A é a área.

Analogamente, para o caso unidimensional, a massa total de uma haste de densidade variável com área da seção transversal constante é dada por

$$m = A \int_0^L \rho(x)\, dx$$

em que m é a massa total (kg), L é o comprimento da haste (em metros), $\rho(x)$ é a densidade conhecida (kg/m^3) como função do comprimento x (em metros) e A é a área da seção transversal da haste (m^2).

Finalmente, as integrais são usadas para resolver equações diferenciais ou de taxa. Suponha que a velocidade de uma partícula é uma função conhecida contínua do tempo $v(t)$,

$$\frac{dy}{dt} = v(t)$$

A distância total y percorrida por essa partícula até o instante t é dada por (Figura PT6.3*b*)

$$y = \int_0^t v(t)\, dt \tag{PT6.5}$$

Essas são apenas algumas poucas aplicações da derivação e da integração com as quais você pode se deparar regularmente no exercício de sua profissão. Quando as funções que devem ser analisadas forem simples, você normalmente preferirá calculá-las analiticamente. Por exemplo, no problema do pára-quedista em queda livre, determinamos a solução para a velocidade como uma função do tempo [Equação (1.10)]. Essa relação poderia ser substituída na Equação (PT6.5), a qual poderia então ser facilmente integrada para determinar quanto o pára-quedista cai até o instante t. Nesse caso, é simples calcular a integral. Entretanto, isso é difícil ou impossível quando a função é complicada, como é tipicamente o caso em exemplos mais realistas. Além disso, a função subjacente, em geral, é desconhecida e é definida apenas por medidas em pontos discretos. Para ambos os casos, você precisa ter a habilidade de obter valores aproximados para derivadas e integrais usando técnicas numéricas. Diversas dessas técnicas serão discutidas nesta parte do livro.

PT6.2 FUNDAMENTOS MATEMÁTICOS

No ensino médio ou durante seu primeiro ano na faculdade, você foi apresentado ao *cálculo diferencial e integral*. Lá você aprendeu técnicas para obter derivadas e integrais analítica ou exatamente.

Quando diferenciamos uma função analiticamente, geramos uma segunda função que pode ser usada para calcular a derivada para diferentes valores da variável independente. Estão disponíveis regras gerais para esse propósito. Por exemplo, no caso de um monômio

$$y = x^n$$

a seguinte regra simples se aplica ($n \neq 0$):

$$\frac{dy}{dx} = nx^{n-1}$$

que é a expressão da regra mais geral para

$$y = u^n$$

em que u é uma função de x. Para essa equação, a derivada é calculada por

$$\frac{dy}{dx} = nu^{n-1}\frac{du}{dx}$$

Duas outras fórmulas úteis se aplicam ao produto e ao quociente de funções. Por exemplo, se o produto de duas funções de x (u e v) for representado por $y = uv$, então a derivada pode ser calculada por

$$\frac{dy}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}$$

Para a divisão, $y = u/v$, a derivada pode ser calculada por

$$\frac{dy}{dx} = \frac{v\dfrac{du}{dx} - u\dfrac{dv}{dx}}{v^2}$$

Outras fórmulas úteis estão resumidas na Tabela PT6.1.

Fórmulas parecidas estão disponíveis para integração definida, que trata da determinação de uma integral entre extremos especificados, como em

$$I = \int_a^b f(x)\,dx \tag{PT6.6}$$

De acordo com o *teorema fundamental do cálculo integral*, a Equação (PT6.6) é calculada por

$$\int_a^b f(x)\,dx = F(x)\Big|_a^b$$

em que $F(x)$ é uma primitiva de $f(x)$ — isto é, qualquer função tal que $F'(x) = f(x)$. A notação no lado direito significa

$$F(x)\Big|_a^b = F(b) - F(a) \tag{PT6.7}$$

Um exemplo de uma integral definida é

$$I = \int_0^{0,8} (0,2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5)\,dx \tag{PT6.8}$$

Para esse caso, a função é um simples polinômio que pode ser integrado analiticamente pelo cálculo de cada termo de acordo com a regra

$$\int_a^b x^n\,dx = \frac{x^{n+1}}{n+1}\Big|_a^b \tag{PT6.9}$$

TABELA PT6.1 Algumas derivadas comumente usadas.

$\frac{d}{dx}\,\text{sen}\,x = \cos x$	$\frac{d}{dx}\,\text{cotg}\,x = -\text{cosec}^2 x$
$\frac{d}{dx}\cos x = -\text{sen}\,x$	$\frac{d}{dx}\sec x = \sec x\,\text{tg}\,x$
$\frac{d}{dx}\,\text{tg}\,x = \sec^2 x$	$\frac{d}{dx}\,\text{cossec}\,x = -\text{cossec}\,x\,\text{cotg}\,x$
$\frac{d}{dx}\ln x = \frac{1}{x}$	$\frac{d}{dx}\log_a x = \frac{1}{x \ln a}$
$\frac{d}{dx}e^x = e^x$	$\frac{d}{dx}a^x = a^x \ln a$

em que n não pode ser igual a -1. Aplicando essa regra a cada termo na Equação (PT6.8), obtemos

$$I = 0{,}2x + 12{,}5x^2 - \frac{200}{3}x^3 + 168{,}75x^4 - 180x^5 + \frac{400}{6}x^6\Big|_0^{0{,}8} \qquad \text{(PT6.10)}$$

a qual pode ser calculada de acordo com a Equação (PT6.7) como sendo $I = 1{,}6405333$. Esse valor é igual à área sob o polinômio original [Equação (PT6.8)] entre $x = 0$ e 0,8.

A integração anterior depende do conhecimento da regra expressa pela Equação (PT6.9). Outras funções seguem regras diferentes. Essas "regras" são todas meros exemplos de *primitivação*, isto é, de encontrar $F(x)$ tal que $F'(x) = f(x)$. Conseqüentemente, a integração analítica depende do conhecimento anterior da resposta. Esse conhecimento é adquirido por treinamento e experiência. Muitas das regras estão resumidas em manuais e em tabelas de integrais. Listamos na Tabela PT6.2 algumas integrais comumente encontradas. Entretanto, muitas funções de importância prática são muito complicadas para estarem contidas em tais tabelas. Uma razão pela qual as técnicas nesta parte do livro são tão valiosas é que elas fornecem uma maneira de calcular relações tais como a Equação (PT6.8) sem conhecimento das regras (*knowledge of the rules*).

TABELA PT6.2 Algumas integrais simples que são usadas na Parte Seis. Os a e b nesta tabela são constantes e não devem ser confundidos com os extremos de integração discutidos no texto.

$$\int u\,dv = uv - \int v\,du$$

$$\int u^n\,du = \frac{u^{n+1}}{n+1} + C \qquad n \neq -1$$

$$\int a^{bx}\,dx = \frac{a^{bx}}{b \ln a} + C \qquad a > 0,\ a \neq 1$$

$$\int \frac{dx}{x} = \ln|x| + C \qquad x \neq 0$$

$$\int \text{sen}\,(ax+b)\,dx = -\frac{1}{a}\cos(ax+b) + C$$

$$\int \cos(ax+b)\,dx = \frac{1}{a}\,\text{sen}\,(ax+b) + C$$

$$\int \ln|x|\,dx = x\ln|x| - x + C$$

$$\int e^{ax}\,dx = \frac{e^{ax}}{a} + C$$

$$\int xe^{ax}\,dx = \frac{e^{ax}}{a^2}(ax-1) + C$$

$$\int \frac{dx}{a+bx^2} = \frac{1}{\sqrt{ab}}\,\text{tg}^{-1}\,\frac{\sqrt{ab}}{a}\,x + C$$

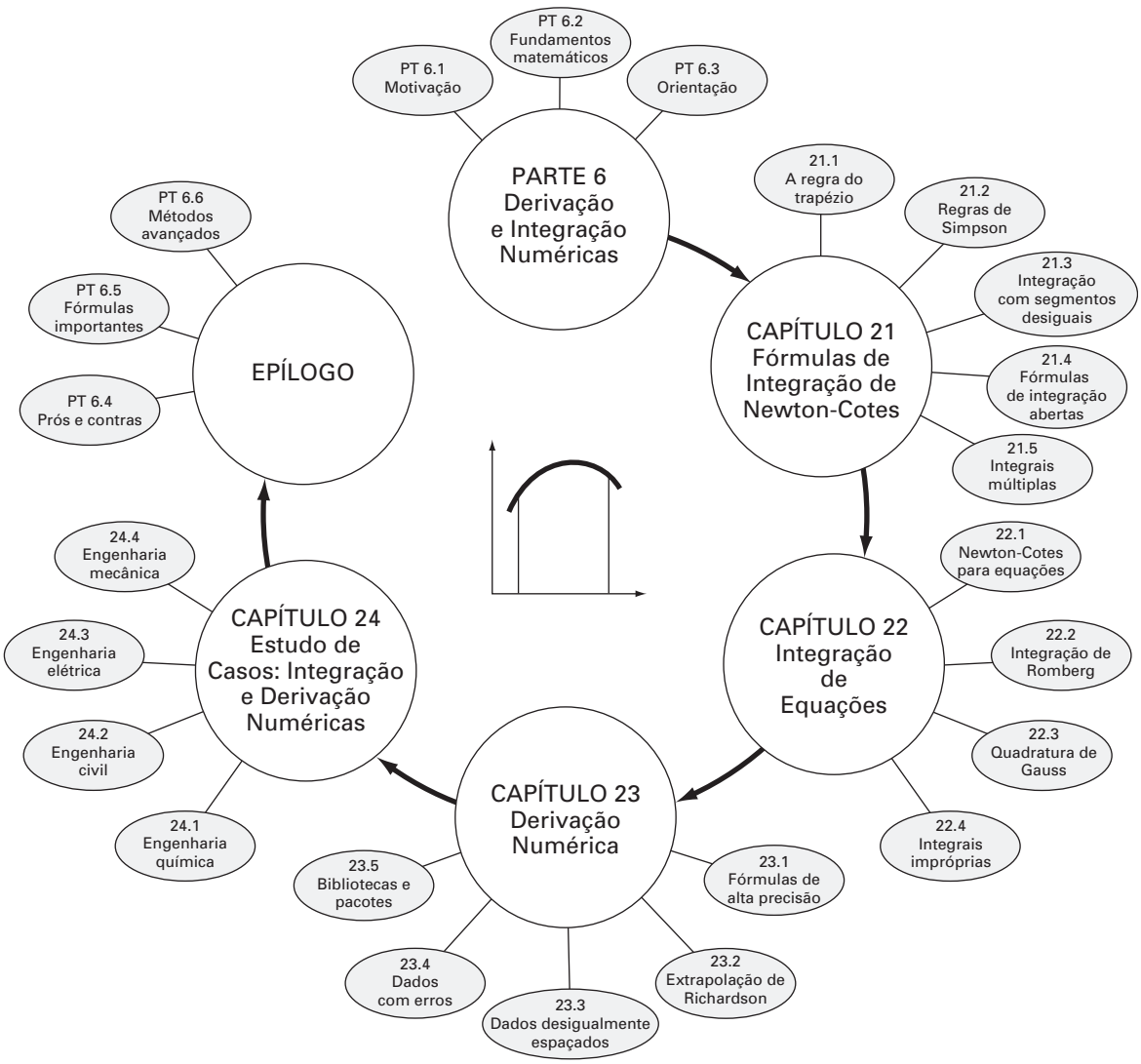
PT6.3 ORIENTAÇÃO

Antes de prosseguir para os métodos numéricos para integração, um pouco mais de orientação pode ser útil. O material a seguir foi planejado como uma visão geral do que é discutido na Parte Seis. Além disso, forram formulados alguns objetivos para ajudá-lo a focalizar seus esforços quando estudar este material.

PT6.3.1 Escopo e Visão Geral

A Figura PT6.10 fornece uma visão geral da Parte Seis. O Capítulo 21 é dedicado às abordagens mais comuns da integração numérica — *as fórmulas de Newton-Cotes*. Essas relações são baseadas na substituição de uma função complicada ou de dados tabulados por um polinômio simples que é fácil de integrar. Três das mais amplamente usadas fórmulas de Newton-Cotes serão discutidas em detalhes: a *regra do trapézio*, a *regra 1/3 de Simpson* e a *regra 3/8 de Simpson*. Todas essas fórmulas se destinam aos casos nos quais os dados a serem integrados estão igualmente espaçados. Além disso, também incluímos a discussão da integração numérica de dados desigualmente espaçados. Esse é um tópico muito importante, porque muitas aplicações do mundo real tratam de dados que têm essa forma.

FIGURA PT6.10
Esquema da organização do material na Parte Seis: Integração e Derivação Numéricas.



Todo o material citado diz respeito à integração fechada, na qual os valores da função nos extremos de integração são conhecidos. No final do Capítulo 21, serão apresentadas *fórmulas de integração abertas*, nas quais os extremos de integração se estendem para além do intervalo de dados conhecidos. Embora não sejam comumente usadas para integração definida, as fórmulas de integração abertas estão apresentadas aqui porque são extensamente usadas na solução de equações diferenciais ordinárias na Parte Sete.

As formulações cobertas no Capítulo 21 podem ser usadas para analisar tanto dados tabulados quanto equações. O Capítulo 22 trata de duas técnicas que foram projetadas especialmente pra integrar equações e funções: a *integração de Romberg* e a *quadratura de Gauss*. São fornecidos algoritmos computacionais para ambos os métodos. Além disso, são discutidos métodos para o cálculo de *integrais impróprias*.

No Capítulo 23, são apresentadas informações adicionais sobre *derivação numérica* para suplementar o material introdutório do Capítulo 4. Os tópicos incluem fórmulas de diferenças finitas de alta precisão, extrapolação de Richardson e derivação de dados desigualmente espaçados. Os efeitos dos erros na derivação e na integração numéricas são discutidos. Finalmente, o capítulo é concluído com uma descrição da aplicação de diversos pacotes e bibliotecas de software para integração e derivação.

O Capítulo 24 mostra como os métodos podem ser aplicados na resolução de problemas. Como nas outras partes do livro, as aplicações são tiradas de todos os setores da engenharia.

Uma seção de revisão ou *epílogo* é incluída no final da Parte Seis. Essa revisão inclui uma discussão dos prós e contras relevantes para a implementação na prática da engenharia. Além disso, são resumidas as fórmulas importantes. Finalmente, é apresentada uma pequena revisão de métodos avançados e referências alternativas que facilitarão seus estudos adicionais de integração e derivação numéricas.

PT6.3.2 Metas e Objetivos

Objetivos de Estudo. Depois de completar a Parte Seis, você deverá ser capaz de resolver muitos problemas de integração e derivação numéricas e apreciar suas aplicações na resolução de problemas. Você deverá esforçar-se para dominar diversas técnicas e avaliar sua confiabilidade. Você deveria entender os prós e contras envolvidos na escolha do "melhor" método (ou métodos) para qualquer problema particular. Além desses objetivos gerais, os conceitos específicos listados na Tabela PT6.3 devem ser assimilados e dominados.

TABELA PT6.3 Objetivos de estudo específicos para a Parte Seis.

1. Entender a dedução das fórmulas de Newton-Cotes; saber como deduzir a regra do trapézio e como montar a dedução de ambas as regras de Simpson; reconhecer que as regras dos trapézios e 1/3 e 3/8 de Simpson representam as áreas sob polinômios de 1º, 2º e 3º graus, respectivamente.
2. Saber as fórmulas e equações de erro para (*a*) a regra do trapézio, (*b*) a aplicação múltipla da regra do trapézio, (*c*) a regra 1/3 de Simpson, (*d*) a regra 3/8 de Simpson e (*e*) a aplicação múltipla da regra de Simpson. Ser capaz de escolher a "melhor" entre essas fórmulas para qualquer contexto particular de problema.
3. Reconhecer que a regra 1/3 de Simpson é de quarta ordem mesmo sendo baseada em apenas três pontos; perceber que todas as fórmulas de Newton-Cotes com número par de segmentos e ímpar de pontos têm precisão melhorada similar.
4. Saber como calcular a integral e a derivada de dados desigualmente espaçados.
5. Reconhecer a diferença entre fórmulas de integração abertas e fechadas.
6. Entender a base teórica da extrapolação de Richardson e como ela é aplicada ao algoritmo da integração de Romberg e na derivação numérica.
7. Entender a diferença fundamental entre as fórmulas de Newton-Cotes e as de quadratura de Gauss.
8. Reconhecer por que tanto a integração de Romberg quanto a quadratura de Gauss são úteis para integrar equações (em oposição a dados tabulados ou discretos).
9. Saber como as fórmulas de integração são usadas para calcular integrais impróprias.
10. Entender a aplicação de fórmulas de derivação numérica de alta precisão.
11. Saber como derivar dados desigualmente espaçados.
12. Reconhecer os diferentes efeitos dos erros nos dados nos processos de derivação e integração numéricas.

Objetivos Computacionais. Serão fornecidos software e algoritmos computacionais simples para implementar as técnicas discutidas na Parte Seis. Todos são úteis como ferramentas educacionais.

São fornecidos algoritmos para a maioria dos outros métodos na Parte Seis. Essa informação vai permitir-lhe expandir sua biblioteca de softwares além da regra do trapézio. Por exemplo, você pode achar útil, do ponto de vista profissional, ter softwares capazes de implementar integração e derivação numéricas de dados desigualmente espaçados. E pode querer também desenvolver seu próprio software para as regras de Simpson, integração de Romberg e quadratura de Gauss, as quais são usualmente mais eficientes e acuradas do que a regra do trapézio.

Finalmente, um de seus objetivos mais importantes deveria ser dominar diversos dos pacotes de software de propósitos gerais que estão amplamente disponíveis. Em particular, você deve tornar-se adepto do uso dessas ferramentas para implementar métodos numéricos na resolução de problemas de engenharia.

CAPÍTULO 21

Fórmulas de Integração de Newton-Cotes

As *fórmulas de Newton-Cotes* são os esquemas mais comuns de integração numérica. Elas são baseadas na estratégia de substituir uma função complicada ou dados tabulados por uma função aproximadora simples que seja fácil de integrar:

$$I = \int_a^b f(x)\,dx \cong \int_a^b f_n(x)\,dx \tag{21.1}$$

em que $f_n(x)$ é um polinômio da forma

$$f_n(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + a_n x^n$$

em que n é o grau do polinômio. Por exemplo, na Figura 21.1*a*, um polinômio de primeiro grau (uma reta) é usado como uma aproximação. Na Figura 21.1*b*, é usada uma parábola para o mesmo propósito.

A integral também pode ser aproximada usando uma série de polinômios aplicados por partes à função ou dados em segmentos de comprimento constante. Por exemplo, na Figura 21.2, três segmentos de reta são usados para aproximar a integral. Polinômios de grau mais alto podem ser usados para o mesmo propósito. Com esse cenário, reconhecemos que o "método das faixas" na Figura PT6.6 empregava uma série de polinômios de grau zero (isto é, constantes) para aproximar a integral.

Estão disponíveis formas abertas e fechadas das fórmulas de Newton-Cotes. As *formas fechadas* são aquelas nas quais os dados nos extremos inicial e final de integração são conhecidos (Figura 21.3*a*). As *fórmulas abertas* têm extremos de integração que se estendem além do intervalo dos dados (Figura 21.3*b*). Nesse sentido, elas são parecidas com a extrapolação discutida na Seção 18.5. As fórmulas abertas de Newton-Cotes, em geral, não são usadas para integração definida. Entretanto, são utilizadas para calcular integrais impróprias e na solução de equações diferenciais ordinárias. Este capítulo enfatiza as formas fechadas. No entanto, material sobre as fórmulas de Newton-Cotes abertas é introduzido brevemente no seu final.

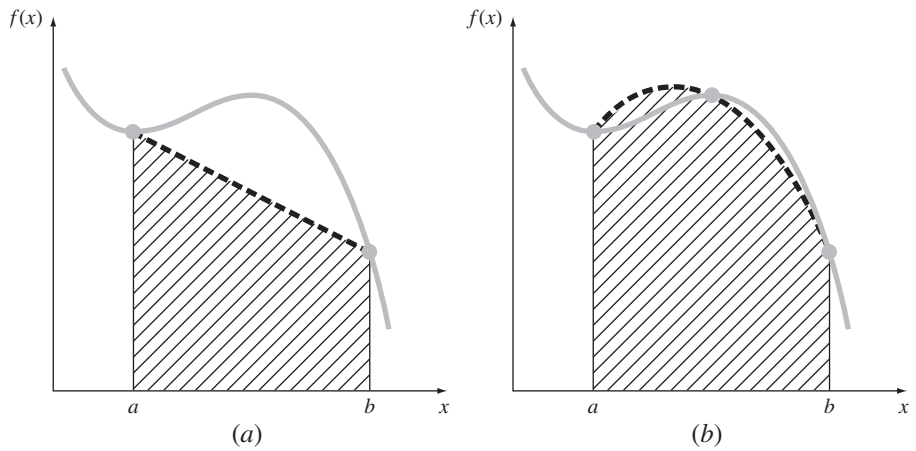


FIGURA 21.1
A aproximação de uma integral pela área sob (*a*) uma única reta e (*b*) uma única parábola.

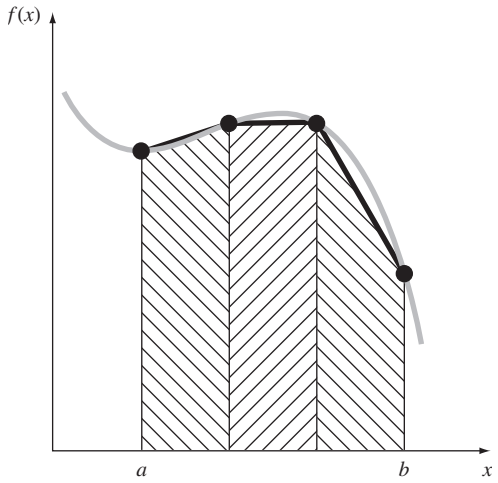


FIGURA 21.2
A aproximação de uma integral pela área sob três segmentos de reta.

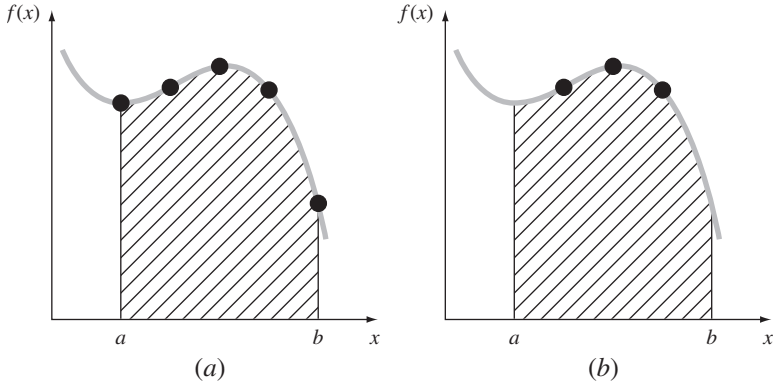


FIGURA 21.3
A diferença entre fórmulas de integração (*a*) fechadas e (*b*) abertas.

21.1 A REGRA DO TRAPÉZIO

A *regra do trapézio* é a primeira fórmula de integração fechada de Newton-Cotes. Ela corresponde ao caso no qual o polinômio na Equação (21.1) é de primeiro grau:

$$I = \int_a^b f(x)\,dx \cong \int_a^b f_1(x)\,dx$$

Lembre-se, do Capítulo 18, de que uma reta pode ser representada por [Equação (18.2)]

$$f_1(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \tag{21.2}$$

A área sob essa reta é uma estimativa da integral de $f(x)$ entre os extremos a e b:

$$I = \int_a^b \left[f(a) + \frac{f(b) - f(a)}{b - a}(x - a) \right] dx$$

O resultado da integração (veja o Quadro 21.1 para detalhes) é

$$I = (b - a)\frac{f(a) + f(b)}{2} \tag{21.3}$$

que é chamada *regra do trapézio*.

Quadro 21.1 Dedução da Regra do Trapézio

Antes da integração, a Equação (21.2) pode ser expressa como

$$f_1(x) = \frac{f(b) - f(a)}{b - a}x + f(a) - \frac{af(b) - af(a)}{b - a}$$

Agrupando os dois últimos termos, obtemos

$$f_1(x) = \frac{f(b) - f(a)}{b - a}x + \frac{bf(a) - af(a) - af(b) + af(a)}{b - a}$$

ou

$$f_1(x) = \frac{f(b) - f(a)}{b - a}x + \frac{bf(a) - af(b)}{b - a}$$

que pode ser integrada entre $x = a$ e $x = b$ para fornecer

$$I = \frac{f(b) - f(a)}{b - a}\frac{x^2}{2} + \frac{bf(a) - af(b)}{b - a}x\Big|_a^b$$

Esse resultado pode ser calculado por

$$I = \frac{f(b) - f(a)}{b - a}\frac{(b^2 - a^2)}{2} + \frac{bf(a) - af(b)}{b - a}(b - a)$$

Agora, como $b^2 - a^2 = (b - a)(b + a)$,

$$I = [f(b) - f(a)]\frac{b + a}{2} + bf(a) - af(b)$$

Fazendo a multiplicação e agrupando os termos, obtemos

$$I = (b - a)\frac{f(a) + f(b)}{2}$$

que é a fórmula da regra dos trapézios.

Geometricamente, a regra dos trapézios é equivalente a aproximar a integral pela área do trapézio sob a reta ligando $f(a)$ e $f(b)$ na Figura 21.4. Lembre-se, da geometria, que a fórmula para calcular a área de um trapézio é a altura vezes a média das bases (Figura 21.5*a*). Em nosso caso, o conceito é o mesmo, mas o trapézio está sobre o seu lado (Figura 21.5*b*). Portanto, a estimativa da integral pode ser representada por

$$I \cong \text{largura} \times \text{altura média} \tag{21.4}$$

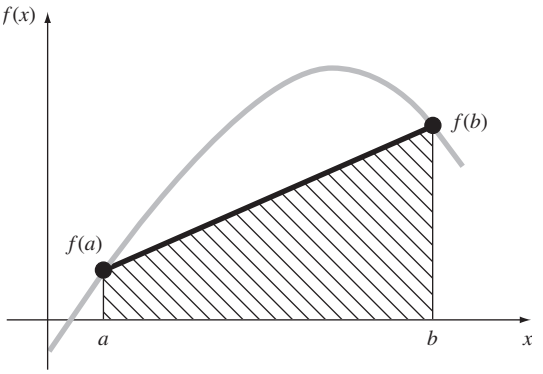
ou

$$I \cong (b - a) \times \text{altura média} \tag{21.5}$$

em que, para a regra dos trapézios, a altura média é a média dos valores da função nas extremidades, ou $[f(a) + f(b)]/2$.

Todas as fórmulas de Newton-Cotes fechadas podem ser expressas na forma geral da Equação (21.5). De fato, elas diferem somente com relação à formulação da altura média.

FIGURA 21.4
Descrição gráfica da regra do trapézio.



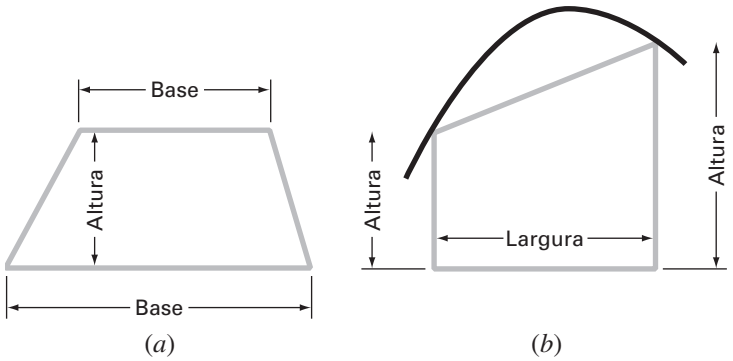


FIGURA 21.5
(*a*) A fórmula para calcular a área de um trapézio: altura vezes a média das bases. (*b*) Para a regra dos trapézios, o conceito é o mesmo, mas o trapézio está sobre seu lado.

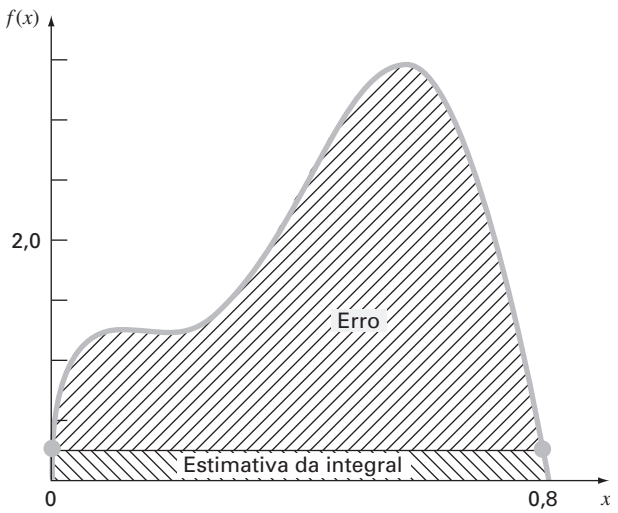
21.1.1 Erro na Regra do Trapézio

Quando empregamos a integral sob um segmento de reta para aproximar a integral sob a curva, obviamente incorremos em um erro que pode ser substancial (Figura 21.6). Uma estimativa para o erro de truncamento local de uma única aplicação da regra do trapézio é (Quadro 21.2)

$$E_t = -\frac{1}{12} f''(\xi)(b-a)^3 \tag{21.6}$$

em que ξ está em algum ponto do intervalo entre a e b. A Equação (21.6) indica que, se a função que está sendo integrada for linear, a regra dos trapézios será exata. Caso contrário, para funções com derivadas de segunda ordem e de ordem superior não-nulas (isto é, com curvatura), pode ocorrer algum erro.

FIGURA 21.6
Descrição gráfica do uso de uma única aplicação da regra do trapézio para aproximar a integral de $f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$ de $x = 0$ a $0{,}8$.



Quadro 21.2 Dedução e Estimativa de Erro da Regra do Trapézio

É possível uma dedução alternativa da regra do trapézio integrando o polinômio interpolador progressivo de Newton-Gregory. Lembre-se de que, para a versão de primeiro grau com termo de erro, a integral seria (Quadro 18.2)

$$I = \int_a^b \left[f(a) + \Delta f(a)\alpha + \frac{f''(\xi)}{2}\alpha(\alpha - 1)h^2 \right] dx \quad \text{(Q21.2.1)}$$

Para simplificar a análise, perceba que, como $\alpha = (x - a)/h$,

$$dx = h\, d\alpha$$

Considerando que $h = b - a$ (para a regra do trapézio com um segmento), os extremos de integração a e b correspondem a 0 e 1, respectivamente. Portanto, a Equação (Q21.2.1) pode ser expressa como

$$I = h \int_0^1 \left[f(a) + \Delta f(a)\alpha + \frac{f''(\xi)}{2}\alpha(\alpha - 1)h^2 \right] d\alpha$$

Se for suposto que, para h pequeno, o termo $f''(\xi)$ é aproximadamente constante, essa equação pode ser integrada:

$$I = h \left[\alpha f(a) + \frac{\alpha^2}{2}\Delta f(a) + \left(\frac{\alpha^3}{6} - \frac{\alpha^2}{4} \right) f''(\xi)h^2 \right]_0^1$$

e calculada como

$$I = h \left[f(a) + \frac{\Delta f(a)}{2} \right] - \frac{1}{12} f''(\xi)h^3$$

Como $\Delta f(a) = f(b) - f(a)$, o resultado pode ser escrito como

$$I = \underbrace{h\frac{f(a) + f(b)}{2}}_{\text{Regra do trapézio}} \underbrace{- \frac{1}{12} f''(\xi)h^3}_{\text{Erro de truncamento}}$$

Logo, o primeiro termo é a regra do trapézio e o segundo é uma aproximação para o erro.

EXEMPLO 21.1 Uma Única Aplicação da Regra dos Trapézios

Enunciado do Problema. Use a Equação (21.3) para integrar numericamente

$$f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

de $a = 0$ a $b = 0{,}8$. Lembre-se, da Seção PT6.2, que o valor exato da integral pode ser determinado analiticamente como sendo 1,640533.

Solução. Os valores da função

$$f(0) = 0{,}2$$
$$f(0{,}8) = 0{,}232$$

podem ser substituídos na Equação (21.3) para fornecer

$$I \cong 0{,}8\frac{0{,}2 + 0{,}232}{2} = 0{,}1728$$

o que representa um erro de

$$E_t = 1{,}640533 - 0{,}1728 = 1{,}467733$$

que corresponde a um erro relativo porcentual de $\varepsilon_t = 89{,}5\%$. A razão para esse erro grande é evidente na descrição gráfica na Figura 21.6. Observe que a área sob a reta despreza uma parte significativa da integral, que fica acima da reta.

Em situações reais, não teríamos conhecimento prévio do valor verdadeiro. Portanto, é necessária uma estimativa de erro aproximada. Para obter essa estimativa, a segunda derivada da função pode ser obtida derivando-se a função original duas vezes para obter

$$f''(x) = -400 + 4.050x - 10.800x^2 + 8.000x^3$$

O valor médio da segunda derivada pode ser calculado usando-se a Equação (PT6.4):

$$\bar{f}''(x) = \frac{\displaystyle\int_0^{0{,}8} (-400 + 4.050x - 10.800x^2 + 8.000x^3)\, dx}{0{,}8 - 0} = -60$$

o qual pode ser substituído na Equação (21.6) para fornecer

$$E_a = -\frac{1}{12}(-60)(0{,}8)^3 = 2{,}56$$

que é da mesma ordem de grandeza e do mesmo sinal que o erro verdadeiro. Entretanto, existe uma discrepância por causa do fato que, para um intervalo desse tamanho, a média da segunda derivada não é uma aproximação acurada de $f''(\xi)$. Logo, indicamos que o erro é aproximado usando a notação E_a em vez de E_t, que seria o erro exato.

21.1.2 A Aplicação Múltipla da Regra do Trapézio

Uma maneira de melhorar a acurácia da regra do trapézio é dividir o intervalo de integração de a a b em diversos segmentos e aplicar o método a cada segmento (Figura 21.7).

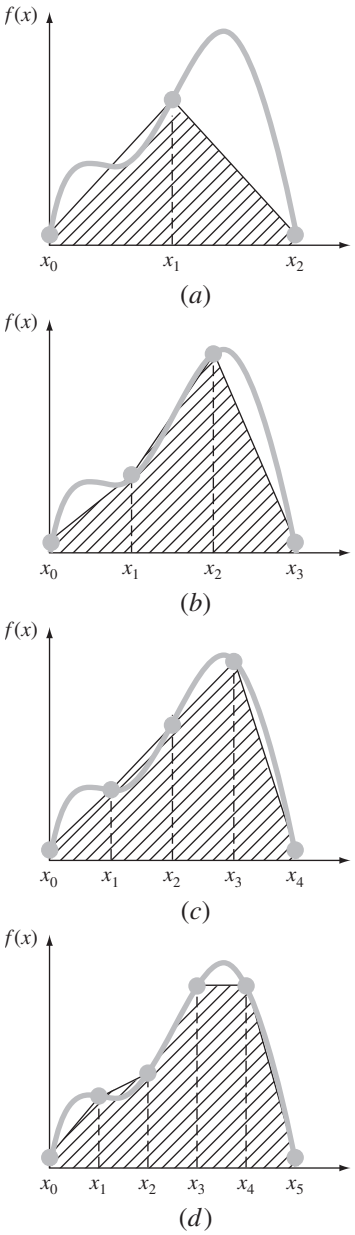


FIGURA 21.7
Ilustração da aplicação múltipla da regra do trapézio. (*a*) Dois segmentos, (*b*) três segmentos, (*c*) quatro segmentos e (*d*) cinco segmentos.

As áreas correspondentes aos segmentos individuais podem então ser somadas para fornecer a integral para o intervalo inteiro. As equações resultantes são chamadas *fórmulas de integração por aplicações múltiplas* ou *compostas*.

A Figura 21.8 mostra a forma geral e a nomenclatura que será usada para caracterizar as integrais por aplicações múltiplas. Existem $n + 1$ pontos base igualmente espaçados $(x_0, x_1, x_2, \ldots, x_n)$. Conseqüentemente, existem n segmentos de largura igual:

$$h = \frac{b - a}{n} \tag{21.7}$$

Se a e b forem designados por x_0 e x_n, respectivamente, a integral total pode ser representada como

$$I = \int_{x_0}^{x_1} f(x)\,dx + \int_{x_1}^{x_2} f(x)\,dx + \cdots + \int_{x_{n-1}}^{x_n} f(x)\,dx$$

Substituindo cada integral pela regra do trapézio, obtém-se

$$I = h\frac{f(x_0) + f(x_1)}{2} + h\frac{f(x_1) + f(x_2)}{2} + \cdots + h\frac{f(x_{n-1}) + f(x_n)}{2} \tag{21.8}$$

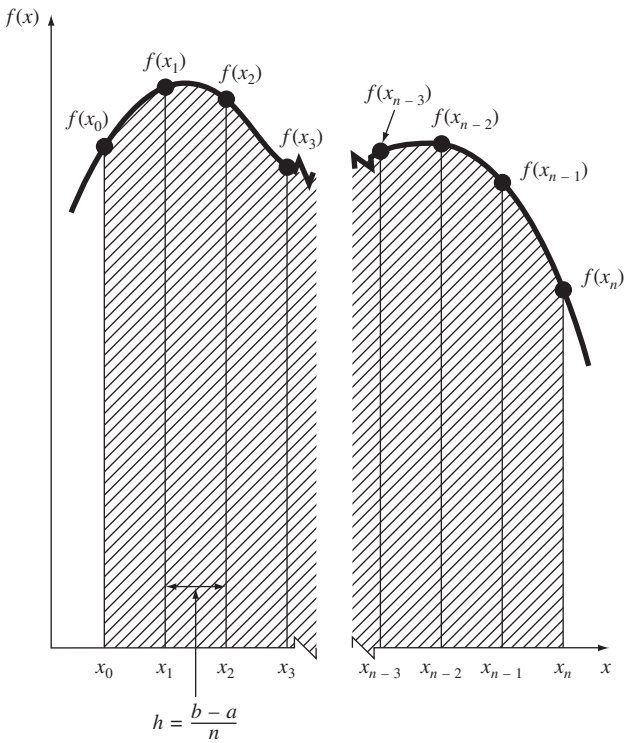
ou, agrupando termos,

$$I = \frac{h}{2}\left[f(x_0) + 2\sum_{i=1}^{n-1} f(x_i) + f(x_n)\right] \tag{21.9}$$

ou, usando a Equação (21.7) para expressar a Equação (21.9) na forma geral da Equação (21.5),

$$I = \underbrace{(b - a)}_{\text{Largura}} \underbrace{\frac{f(x_0) + 2\sum_{i=1}^{n-1} f(x_i) + f(x_n)}{2n}}_{\text{Altura média}} \tag{21.10}$$

FIGURA 21.8
A forma geral e a nomenclatura para integrais por aplicação múltipla.



Como a soma dos coeficientes de $f(x)$ no numerador dividido por $2n$ é igual a 1, a altura média representa uma média ponderada dos valores da função. De acordo com a Equação (21.10), os pontos interiores têm um peso duas vezes maior do que as extremidades $f(x_0)$ e $f(x_n)$.

Um erro para a aplicação múltipla da regra do trapézio pode ser obtido pela soma dos erros individuais em cada segmento, o que dá

$$E_t = -\frac{(b-a)^3}{12n^3}\sum_{i=1}^{n} f''(\xi_i) \tag{21.11}$$

em que $f''(\xi_i)$ é a segunda derivada em um ponto ξ_i localizado no segmento i. Esse resultado pode ser simplificado por uma estimativa do valor médio da segunda derivada no intervalo todo como [Equação (PT6.3)]

$$\bar{f}'' \cong \frac{\sum_{i=1}^{n} f''(\xi_i)}{n} \tag{21.12}$$

Portanto, $\Sigma f''(\xi_i) \cong n\bar{f}''$ e a Equação (21.11) podem ser reescritas como

$$E_a = -\frac{(b-a)^3}{12n^2}\bar{f}'' \tag{21.13}$$

Logo, se o número de segmentos for dobrado, o erro de truncamento será dividido por quatro. Observe que a Equação (21.13) é um erro aproximado por causa da natureza aproximada da Equação (21.12).

EXEMPLO 21.2 Aplicação Múltipla da Regra dos Trapézios

Enunciado do Problema. Use a regra do trapézio com dois segmentos para obter uma estimativa da integral de

$$f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

de $a = 0$ a $b = 0{,}8$. Empregue a Equação (21.13) para fazer uma estimativa do erro. Lembre-se que o valor correto da integral é 1,640533.

Solução. $n = 2$ ($h = 0{,}4$):

$$f(0) = 0{,}2 \qquad f(0{,}4) = 2{,}456 \qquad f(0{,}8) = 0{,}232$$

$$I = 0{,}8\frac{0{,}2 + 2(2{,}456) + 0{,}232}{4} = 1{,}0688$$

$$E_t = 1{,}640533 - 1{,}0688 = 0.57173 \qquad \varepsilon_t = 34{,}9\%$$

$$E_a = -\frac{0{,}8^3}{12(2)^2}(-60) = 0{,}64$$

em que -60 é a média da segunda derivada determinada anteriormente no Exemplo 21.1.

Os resultados do exemplo anterior, junto com os da aplicação da regra do trapézio para 3 a 10 segmentos, estão resumidos na Tabela 21.1. Observe como o erro diminui conforme o número de segmentos aumenta. Entretanto, observe também que a taxa de diminuição é gradual. Isso acontece porque o erro está inversamente relacionado com o quadrado de n [Equação (21.13)]. Portanto, dobrar o número de segmentos faz que o erro fique quatro vezes menor. Nas seções subseqüentes, serão desenvolvidas fórmulas de ordem superior que são mais acuradas e que convergem mais rapidamente para a integral verdadeira conforme o número de segmentos aumenta. Mas, antes de se investigar essas fórmulas, serão primeiro discutidos algoritmos computacionais para implementar a regra do trapézio.

21.1.3 Algoritmos Computacionais para a Regra do Trapézio

Dois algoritmos simples para a regra do trapézio estão listados na Figura 21.9. O primeiro (Figura 21.9*a*) é para a versão com um único segmento. O segundo (Figura 21.9*b*) é para a versão com segmentos múltiplos, com um comprimento constante dos segmentos. Observe que ambos foram desenvolvidos para dados que estejam na forma tabular. Um programa geral deveria ser capaz de calcular funções conhecidas e também equações. Vamos ilustrar como as funções são tratadas no Capítulo 22.

TABELA 21.1 Resultados da aplicação múltipla da regra do trapézio para obter uma estimativa da integral de $f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$ de $x = 0$ a 0,8. O valor exato é 1,640533.

n	h	I	ε_t (%)
2	0,4	1,0688	34,9
3	0,2667	1,3695	16,5
4	0,2	1,4848	9,5
5	0,16	1,5399	6,1
6	0,1333	1,5703	4,3
7	0,1143	1,5887	3,2
8	0,1	1,6008	2,4
9	0,0889	1,6091	1,9
10	0,08	1,6150	1,6

(*a*) Um único segmento

```
FUNCTION Trap (h, f0, f1)
  Trap = h * (f0 + f1)/2
END Trap
```

(*b*) Múltiplos segmentos

```
FUNCTION Trapm (h, n, f)
  soma = f0
  DOFOR i = 1, n - 1
    soma = soma + 2 * fi
  END DO
  soma = soma + fn
  Trapm = h * soma / 2
END Trapm
```

FIGURA 21.9
Algoritmos para a regra dos trapézios com (*a*) um único segmento e (*b*) segmentos múltiplos.

EXEMPLO 21.3 Calculando Integrais com o Computador

Enunciado do Problema. Use um software baseado na Figura 21.9*b* para resolver um problema relacionado com nosso amigo, o pára-quedista em queda livre. Como você se lembra, do Exemplo 1.1, a velocidade do pára-quedista é dada pela seguinte função do tempo:

$$v(t) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{E21.3.1}$$

em que v é a velocidade (m/s), $g = 9{,}8$ m/s^2 é a constante gravitacional, $m = 68{,}1$ kg é a massa do pára-quedista e $c = 12{,}5$ kg/s é o coeficiente de arrasto. O modelo prevê a velocidade do pára-quedista como função do tempo como descrito no Exemplo 1.1.

Suponha que quiséssemos saber quanto o pára-quedista caiu depois de um certo intervalo de tempo t. A distância é dada por [Equação (PT6.5)]

$$d = \int_0^t v(t)\,dt$$

em que d é a distância em metros. Substituindo a Equação (E21.3.1),

$$d = \frac{gm}{c}\int_0^t \left(1 - e^{-(c/m)t}\right) dt$$

Use seu software para determinar essa integral com a regra do trapézio com segmentos múltiplos, utilizando diferentes números de segmentos. Observe que fazer a integral analiticamente e substituir os valores dados dos parâmetros resulta em um valor exato de $d = 289{,}43515$ m.

Solução. No caso em que $n = 10$ segmentos, é obtida uma integral aproximada de 288,7491. Logo, obtivemos a integral com acurácia de três algarismos significativos. Resultados para outros números de segmentos podem facilmente ser gerados.

Segmentos	Tamanho do Segmento	Estimativa de d, m	ε_t (%)
10	1,0	288,7491	0,237
20	0,5	289,2636	0,0593
50	0,2	289,4076	$9{,}5 \times 10^{-3}$
100	0,1	289,4282	$2{,}4 \times 10^{-3}$
200	0,05	289,4336	$5{,}4 \times 10^{-4}$
500	0,02	289,4348	$1{,}2 \times 10^{-4}$
1.000	0,01	289,4360	$-3{,}0 \times 10^{-4}$
2.000	0,005	289,4369	$-5{,}9 \times 10^{-4}$
5.000	0,002	289,4337	$5{,}2 \times 10^{-4}$
10.000	0,001	289,4317	$1{,}2 \times 10^{-3}$

Até cerca de 500 segmentos, a aplicação múltipla da regra do trapézio alcança uma acurácia excelente. Entretanto, observe como o erro muda de sinal e começa a aumentar em valor absoluto além do caso de 500 segmentos. O caso de 10.000 segmentos, na realidade, parece divergir do valor verdadeiro. Isso decorre da intrusão dos erros de arredondamento, por causa do grande número de cálculos para essa quantidade de segmentos. Logo, o nível de precisão é limitado e nunca atingiríamos o resultado exato de 289,4351 obtido analiticamente. Essa limitação e formas de superá-la serão discutidas em mais detalhes no Capítulo 22.

Três conclusões principais podem ser tiradas do Exemplo 21.3:

- Para aplicações individuais em funções bem comportadas, a aplicação múltipla da regra do trapézio é decididamente boa para obter o tipo de acurácia necessária em muitas aplicações de engenharia.
- Se for necessária uma alta acurácia, a regra do trapézio com segmentos múltiplos exige um grande esforço computacional. Embora esse esforço possa ser desprezível para uma única aplicação, pode ser muito importante quando (*a*) muitas integrais estão sendo calculadas ou (*b*) a própria função leva muito tempo para ser calculada. Para tais casos, abordagens mais eficientes (como aquelas no restante deste capítulo e no próximo) podem ser necessárias.
- Finalmente, os erros de arredondamento podem limitar nossa habilidade de determinar integrais. Isso decorre tanto da precisão da máquina quanto da grande quantidade de cálculos envolvida em técnicas simples como a regra do trapézio com segmentos múltiplos.

Voltamo-nos agora para uma maneira pela qual a eficiência é melhorada. Ou seja, ao uso de polinômio de grau mais alto para obter uma aproximação da integral.

21.2 REGRAS DE SIMPSON

Além de aplicar a regra do trapézio com segmentos menores, outra forma de obter uma estimativa mais acurada de uma integral é usar polinômios de grau mais alto para ligar os pon-

tos. Por exemplo, se existir um ponto extra no ponto médio entre $f(a)$ e $f(b)$, os três pontos podem ser ligados por uma parábola (Figura 21.10*a*). Se existirem dois pontos igualmente espaçados entre $f(a)$ e $f(b)$, os quatro pontos podem ser ligados por um polinômio de 3º grau (Figura 21.10*b*). As fórmulas que resultam de tomar as integrais desses polinômios são chamadas de *regras de Simpson*.

21.2.1 A Regra 1/3 de Simpson

A regra 1/3 de Simpson é obtida quando um polinômio interpolador de segundo grau é substituído na Equação (21.1):

$$I = \int_a^b f(x)\,dx \cong \int_a^b f_2(x)\,dx$$

Se a e b forem designados por x_0 e x_2 e se $f_2(x)$ for representado por um polinômio de Lagrange de segundo grau [Equação (18.23)], a integral se torna

$$I = \int_{x_0}^{x_2} \left[\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} f(x_2) \right] dx$$

Depois da integração e de manipulações algébricas, obtém-se a seguinte fórmula:

$$I \cong \frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)] \tag{21.14}$$

em que, para esse caso, $h = (b-a)/2$. Essa equação é conhecida como a *regra* 1/3 *de Simpson*. Ela é a segunda fórmula de integração fechada de Newton-Cotes. A designação "1/3" vem do fato que h está dividido por 3 na Equação (21.14). Uma dedução alternativa é mostrada no Quadro 21.3, no qual o polinômio de Newton-Gregory é integrado para se obter a mesma fórmula.

A regra 1/3 de Simpson também pode ser expressa usando o formato da Equação (21.5):

$$I \cong \underbrace{(b-a)}_{\text{Largura}} \underbrace{\frac{f(x_0) + 4f(x_1) + f(x_2)}{6}}_{\text{Altura média}} \tag{21.15}$$

em que $a = x_0$, $b = x_2$ e x_1 é o ponto médio entre a e b, o qual é dado por $(b + a)/2$. Observe que, de acordo com a Equação (21.15), o ponto médio tem peso de dois terços, e os dois pontos extremos, de um sexto.

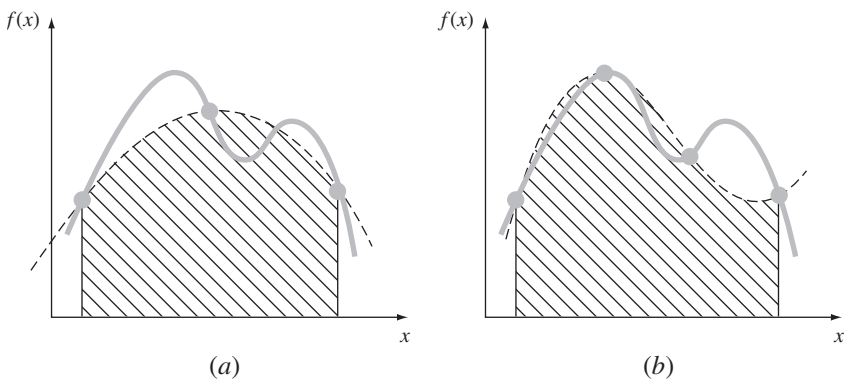


FIGURA 21.10
(*a*) Descrição gráfica da regra 1/3 de Simpson: ela consiste em tomar a área sob uma parábola ligando três pontos. (*b*) Descrição gráfica da regra 3/8 de Simpson: ela consiste em tomar a área sob uma equação cúbica ligando quatro pontos.

Quadro 21.3 Dedução e Estimativa de Erro da Regra 1/3 de Simpson

Como foi feito no Quadro 21.2 para a regra do trapézio, a regra 1/3 de Simpson pode ser deduzida pela integração do polinômio interpolador progressivo de Newton-Gregory (Quadro 18.2):

$$I = \int_{x_0}^{x_2} \left[f(x_0) + \Delta f(x_0)\alpha + \frac{\Delta^2 f(x_0)}{2}\alpha(\alpha-1) + \frac{\Delta^3 f(x_0)}{6}\alpha(\alpha-1)(\alpha-2) + \frac{f^{(4)}(\xi)}{24}\alpha(\alpha-1)(\alpha-2)(\alpha-3)h^4 \right] dx$$

Observe que escrevemos o polinômio até o termo de quarto grau em vez do de terceiro como seria esperado. A razão para isso ficará logo aparente. Observe também que os extremos de integração são de x_0 a x_2. Portanto, quando as substituições simplificadoras são feitas (lembre-se do Quadro 21.2), a integral vai de $\alpha = 0$ a 2:

$$I = h \int_0^2 \left[f(x_0) + \Delta f(x_0)\alpha + \frac{\Delta^2 f(x_0)}{2}\alpha(\alpha-1) + \frac{\Delta^3 f(x_0)}{6}\alpha(\alpha-1)(\alpha-2) + \frac{f^{(4)}(\xi)}{24}\alpha(\alpha-1)(\alpha-2)(\alpha-3)h^4 \right] d\alpha$$

que pode ser integrada para fornecer

$$I = h\left[\alpha f(x_0) + \frac{\alpha^2}{2}\Delta f(x_0) + \left(\frac{\alpha^3}{6} - \frac{\alpha^2}{4}\right)\Delta^2 f(x_0) + \left(\frac{\alpha^4}{24} - \frac{\alpha^3}{6} + \frac{\alpha^2}{6}\right)\Delta^3 f(x_0) + \left(\frac{\alpha^5}{120} - \frac{\alpha^4}{16} + \frac{11\alpha^3}{72} - \frac{\alpha^2}{8}\right) f^{(4)}(\xi)h^4 \right]_0^2$$

e, calculada nos extremos, dá

$$I = h\left[2f(x_0) + 2\Delta f(x_0) + \frac{\Delta^2 f(x_0)}{3} + (0)\Delta^3 f(x_0) - \frac{1}{90} f^{(4)}(\xi)h^4 \right] \quad \text{(Q21.3.1)}$$

Observe o resultado significativo de que o coeficiente da terceira diferença dividida é nulo. Como $\Delta f(x_0) = f(x_1) - f(x_0)$ e $\Delta^2 f(x_0) = f(x_2) - 2f(x_1) + f(x_0)$, a Equação (Q21.3.1) pode ser reescrita como

$$I = \underbrace{\frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)]}_{\text{Regra 1/3 de Simpson}} - \underbrace{\frac{1}{90} f^{(4)}(\xi)h^5}_{\text{Erro de truncamento}}$$

Logo, o primeiro termo é a regra 1/3 de Simpson e o segundo é o erro de truncamento. Como a terceira diferença dividida desapareceu, obtemos o resultado significativo de que a fórmula é acurada até terceira ordem.

É possível mostrar que a aplicação da regra 1/3 de Simpson para um único segmento tem um erro de truncamento de (Quadro 21.3)

$$E_t = -\frac{1}{90}h^5 f^{(4)}(\xi)$$

ou, como $h = (b-a)/2$,

$$E_t = -\frac{(b-a)^5}{2.880} f^{(4)}(\xi) \quad (21.16)$$

em que ξ é algum ponto no intervalo entre a e b. Logo, a regra 1/3 de Simpson é mais acurada do que a regra do trapézio. A comparação com a Equação (21.6) indica que ela é mais acurada do que o esperado. Em vez de ser proporcional à terceira derivada, o erro é proporcional à quarta derivada. Isso ocorre porque, como mostrado no Quadro 21.3, o coeficiente do termo de terceiro grau se anula durante a integração do polinômio interpolador. Conseqüentemente, a regra 1/3 de Simpson é acurada até terceira ordem, mesmo sendo baseada em apenas três pontos. Em outras palavras, ela fornece resultados exatos para polinômios cúbicos, mesmo tendo sido deduzida a partir de uma parábola!

EXEMPLO 21.4 Uma Única Aplicação da Regra 1/3 de Simpson

Enunciado do Problema. Use a Equação (21.15) para integrar

$$f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

de $a = 0$ a $b = 0{,}8$. Lembre-se que a integral exata é 1,640533.

Solução.

$$f(0) = 0{,}2 \qquad f(0{,}4) = 2{,}456 \qquad f(0{,}8) = 0{,}232$$

Portanto, a Equação (21.15) pode ser usada para calcular

$$I \cong 0{,}8\frac{0{,}2 + 4(2{,}456) + 0{,}232}{6} = 1{,}367467$$

o que representa um erro exato de

$$E_t = 1{,}640533 - 1{,}367467 = 0{,}2730667 \qquad \varepsilon_t = 16{,}6\%$$

o que é aproximadamente 5 vezes mais acurado do que uma única aplicação da regra do trapézio (Exemplo 21.1).

O erro estimado é [Equação 21.16)]

$$E_a = -\frac{(0{,}8)^5}{2.880}(-2.400) = 0{,}2730667$$

em que -2.400 é o valor médio da quarta derivada para o intervalo, obtido usando a Equação (PT6.4). Como foi o caso no Exemplo 21.1, o erro é aproximado (E_a) porque o valor médio da quarta derivada não é uma estimativa exata de $f^{(4)}(\xi)$. Entretanto, como esse caso trata de um polinômio de quinto grau, o resultado coincide.

21.2.2 Aplicações Múltiplas da Regra 1/3 de Simpson

Do mesmo modo como no caso da regra do trapézio, a regra de Simpson pode ser melhorada dividindo-se o intervalo de integração em diversos segmentos de mesmo comprimento (Figura 21.11):

$$h = \frac{b - a}{n} \tag{21.17}$$

A integral total pode ser representada como

$$I = \int_{x_0}^{x_2} f(x)\,dx + \int_{x_2}^{x_4} f(x)\,dx + \cdots + \int_{x_{n-2}}^{x_n} f(x)\,dx$$

Substituindo cada integral individual pela regra 1/3 de Simpson, obtemos

$$I \cong 2h\frac{f(x_0) + 4f(x_1) + f(x_2)}{6} + 2h\frac{f(x_2) + 4f(x_3) + f(x_4)}{6}$$
$$+ \cdots + 2h\frac{f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)}{6}$$

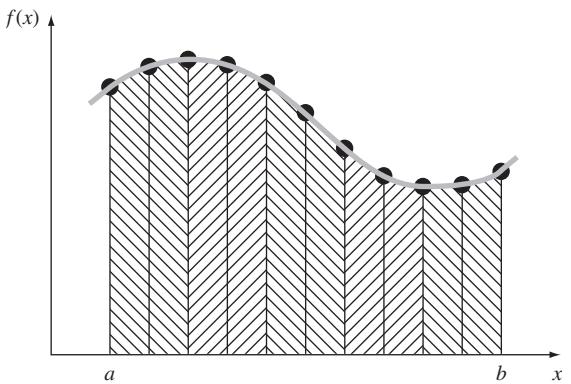


FIGURA 21.11
Representação gráfica da aplicação múltipla da regra 1/3 de Simpson. Observe que o método pode ser usado apenas se o número de intervalos for par.

ou, combinando os termos e usando a Equação (21.17),

$$I \cong \underbrace{(b-a)}_{\text{Largura}} \underbrace{\frac{f(x_0) + 4\sum_{i=1,3,5}^{n-1} f(x_i) + 2\sum_{j=2,4,6}^{n-2} f(x_j) + f(x_n)}{3n}}_{\text{Altura média}} \tag{21.18}$$

Observe que, como ilustrado na Figura 21.11, deve ser utilizado um número par de segmentos para implementar o método. Além disso, os coeficientes "4" e "2" na Equação (21.18) poderiam, a princípio, parecer peculiares. Entretanto, eles seguem naturalmente da regra 1/3 de Simpson. Os pontos ímpares representam o termo médio para cada aplicação e, portanto, levam o peso 4 da Equação (21.15). Os pontos pares são comuns a aplicações adjacentes e, então, são contados duas vezes.

Uma estimativa de erro para a aplicação múltipla da regra de Simpson é obtida da mesma maneira que para a regra trapezoidal, somando-se os erros individuais para os segmentos e fazendo a média da derivada, resultando em

$$E_a = -\frac{(b-a)^5}{180n^4}\bar{f}^{(4)} \tag{21.19}$$

em que $\bar{f}^{(4)}$ é o valor médio da quarta derivada no intervalo.

EXEMPLO 21.5 Versão com Aplicações Múltiplas da Regra 1/3 de Simpson

Enunciado do Problema. Use a Equação (21.18) com $n = 4$ para obter uma estimativa da integral de

$$f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

de $a = 0$ a $b = 0{,}8$. Lembre que a integral exata é 1,640533.

Solução. $n = 4$ ($h = 0{,}2$):

$$f(0) = 0{,}2 \qquad f(0{,}2) = 1{,}288$$
$$f(0{,}4) = 2{,}456 \qquad f(0{,}6) = 3{,}464$$
$$f(0{,}8) = 0{,}232$$

Da Equação (21.18),

$$I = 0{,}8\,\frac{0{,}2 + 4(1{,}288 + 3{,}464) + 2(2{,}456) + 0{,}232}{12} = 1{,}623467$$

$$E_t = 1{,}640533 - 1{,}623467 = 0{,}017067 \qquad \varepsilon_t = 1{,}04\%$$

O erro estimado [Equação (21.19)] é

$$E_a = -\frac{(0{,}8)^5}{180(4)^4}(-2.400) = 0{,}017067$$

O exemplo anterior ilustra que a versão com aplicações múltiplas da regra 1/3 de Simpson fornece resultados muito acurados. Por essa razão, ela é considerada superior à regra do trapézio na maioria das aplicações. Entretanto, como mencionado anteriormente, ela é limitada aos casos nos quais os valores estão igualmente espaçados. Além disso, é limitada a situações nas quais haja um número par de segmentos e um número ímpar de pontos. Conseqüentemente, como discutido na próxima seção, uma fórmula para um

número ímpar de segmentos e par de pontos, conhecida como regra 3/8 de Simpson, é usada em conjunto com a regra 1/3 para permitir o cálculo tanto para um número par quanto para um número ímpar de segmentos.

21.2.3 Regra 3/8 de Simpson

De uma maneira parecida com a dedução da regra do trapézio e da regra 1/3 de Simpson, um polinômio de Lagrange de ordem quatro pode ser ajustado a quatro pontos e integrado:

$$I = \int_a^b f(x)\,dx \cong \int_a^b f_3(x)\,dx$$

para fornecer

$$I \cong \frac{3h}{8}[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)]$$

em que $h = (b - a)/3$. Essa equação é chamada *regra 3/8 de Simpson* porque h é multiplicada por 3/8. Ela é a terceira fórmula de integração fechada de Newton-Cotes. A regra 3/8 pode também ser expressa na forma da Equação (21.5):

$$I \cong \underbrace{(b-a)}_{\text{Largura}} \underbrace{\frac{f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)}{8}}_{\text{Altura média}} \tag{21.20}$$

Assim, os dois pontos interiores têm pesos de três oitavos, enquanto as extremidades têm peso de um oitavo. A regra 3/8 de Simpson tem um erro de

$$E_t = -\frac{3}{80}h^5 f^{(4)}(\xi)$$

ou, como $h = (b - a)/3$,

$$E_t = -\frac{(b-a)^5}{6.480} f^{(4)}(\xi) \tag{21.21}$$

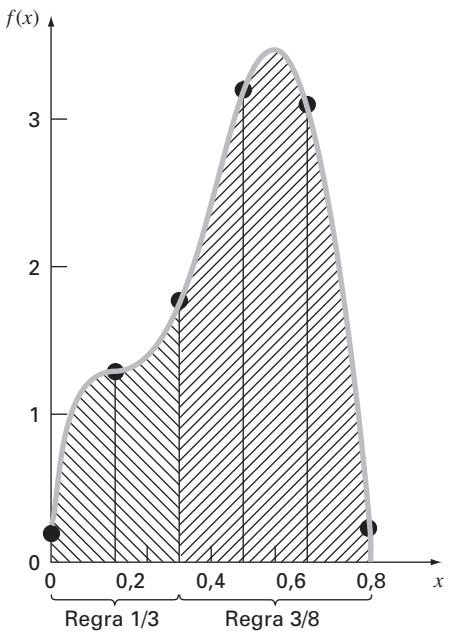


FIGURA 21.12
Ilustração de como as regras 1/3 e 3/8 de Simpson podem ser usadas em conjunto para tratar aplicações múltiplas com um número ímpar de intervalos.

Como o denominador da Equação (21.21) é maior do que o da Equação (21.16), a regra 3/8 é um pouco mais acurada do que a regra 1/3.

A regra 1/3 de Simpson é usualmente o método preferido, pois alcança uma acurácia de terceira ordem com três pontos em vez dos quatro pontos necessários para a versão 3/8. Entretanto, a regra 3/8 tem utilidade quando o número de segmentos é ímpar. No caso do Exemplo 21.5, usamos a regra de Simpson para integrar a função com quatro segmentos. Suponha que você quisesse uma estimativa usando cinco segmentos. Uma opção seria usar a versão com aplicação múltipla da regra do trapézio, como foi feito nos Exemplos 21.2 e 21.3. Entretanto, isso pode não ser aconselhável por causa do grande erro de truncamento associado com esse método. Uma alternativa seria aplicar a regra 1/3 de Simpson aos primeiros dois segmentos e a regra 3/8 de Simpson aos últimos três (Figura 21.12). Dessa maneira, você poderia obter uma estimativa acurada até ordem três em todo o intervalo.

EXEMPLO 21.6 A Regra 3/8 de Simpson

Enunciado do Problema.

(**a**) Use a regra 3/8 de Simpson para integrar

$$f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

de $a = 0$ a $b = 0{,}8$.

(**b**) Use-a em conjunto com a regra 1/3 de Simpson para integrar a mesma função usando cinco segmentos.

Solução.

(**a**) Uma única aplicação da regra 3/8 de Simpson exige quatro pontos igualmente espaçados:

$$f(0) = 0{,}2 \qquad f(0{,}2667) = 1{,}432724$$
$$f(0{,}5333) = 3{,}487177 \qquad f(0{,}8) = 0{,}232$$

Usando a Equação (21.20),

$$I \cong 0{,}8\,\frac{0{,}2 + 3(1{,}432724 + 3{,}487177) + 0{,}232}{8} = 1{,}519170$$

$$E_t = 1{,}640533 - 1{,}519170 = 0{,}1213630 \qquad \varepsilon_t = 7{,}4\%$$

$$E_a = -\frac{(0{,}8)^5}{6.480}(-2.400) = 0{,}1213630$$

(**b**) Os dados necessários para uma aplicação com cinco segmentos ($h = 0.16$) são

$$f(0) = 0{,}2 \qquad f(0{,}16) = 1{,}296919$$
$$f(0{,}32) = 1{,}743393 \qquad f(0{,}48) = 3{,}186015$$
$$f(0{,}64) = 3{,}181929 \qquad f(0{,}80) = 0{,}232$$

A integral para os primeiros dois segmentos é obtida usando-se a regra 1/3 de Simpson:

$$I \cong 0{,}32\,\frac{0{,}2 + 4(1{,}296919) + 1{,}743393}{6} = 0{,}3803237$$

Para os três últimos segmentos, a regra 3/8 pode ser usada para se obter

$$I \cong 0{,}48\,\frac{1{,}743393 + 3(3{,}186015 + 3{,}181929) + 0{,}232}{8} = 1{,}264754$$

A integral total é calculada pela soma destes dois resultados:

$$I = 0{,}3803237 + 1{,}264753 = 1{,}645077$$
$$E_t = 1{,}640533 - 1{,}645077 = -0{,}00454383 \qquad \varepsilon_t = -0{,}28\%$$

21.2.4 Algoritmos Computacionais para as Regras de Simpson

Pseudocódigos para diversas formas da regra de Simpson estão delineados na Figura 21.13. Observe que todos são desenvolvidos para dados que estejam em uma forma tabulada. Um programa geral deveria ser capaz de calcular funções conhecidas e equações. Ilustraremos como as funções são tratadas no Capítulo 22.

Observe que o programa na Figura 21.13*d* é apresentado de modo que tanto um número par quanto um ímpar de segmentos possam ser usados. Para o caso par, a regra 1/3 de Simpson é aplicada para cada par de segmentos e os resultados são somados para se calcular a integral total. Para o caso ímpar, a regra 3/8 de Simpson é aplicada aos três últimos segmentos e a regra 1/3 é aplicada a todos os segmentos anteriores.

21.2.5 Fórmulas Fechadas de Newton-Cotes de Ordem Superior

Como observado previamente, a regra do trapézio e ambas as regras de Simpson são membros de uma família de equações de integração conhecidas como fórmulas de integração fechadas de Newton-Cotes. Algumas das fórmulas estão resumidas na Tabela 21.2 junto com suas estimativas de erro.

(*a*)

```
FUNCTION Simp13 (h, f0, f1, f2)
  Simp13 = 2*h* (f0+4*f1+f2) / 6
END Simp13
```

(*b*)

```
FUNCTION Simp38 (h, f0, f1, f2, f3)
  Simp38 = 3*h* (f0+3*(f1+f2)+f3) / 8
END Simp38
```

(*c*)

```
FUNCTION Simp13m (h, n, f)
  soma = f(0)
  DOFOR i = 1, n − 2, 2
    soma = soma + 4 * f_i + 2 * f_i+1
  END DO
  soma = soma + 4 * f_n−1 + f_n
  Simp13m = h * soma / 3
END Simp13m
```

(*d*)

```
FUNCTION SimpInt(a,b,n,f)
  h = (b − a) / n
  IF n = 1 THEN
    soma = Trap(h, f_n−1, f_n)
  ELSE
    m = n
    ímpar = n / 2 − INT(n / 2)
    IF ímpar > 0 AND n > 1 THEN
      soma = soma+Simp38(h, f_n−3, f_n−2, f_n−1, f_n)
      m = n − 3
    END IF
    IF m > 1 THEN
      soma = soma + Simp13m(h,m,f)
    END IF
  END IF
  SimpInt = soma
END SimpInt
```

FIGURA 21.13
Pseudocódigos para as regras de Simpson. (*a*) Uma única aplicação da regra 1/3 de Simpson, (*b*) uma única aplicação da regra 3/8 de Simpson, (*c*) aplicação múltipla da regra 1/3 de Simpson e (*d*) aplicação múltipla da regra de Simpson tanto para número par quanto ímpar de segmentos. Observe que, em todos os casos, *n* deve ser maior ou igual a 1.

TABELA 21.2 Fórmulas de integração fechadas de Newton-Cotes. As fórmulas são apresentadas no formato da Equação (21.5) de modo que os pesos dos pontos dados na estimativa da altura média são aparentes. O tamanho do passo é dado por $h = (b - a)/n$.

Segmentos (n)	Pontos	Nome	Fórmula	Erro de Truncamento
1	2	Regra do trapézio	$(b-a)\dfrac{f(x_0)+f(x_1)}{2}$	$-(1/12)h^3f''(\xi)$
2	3	Regra 1/3 de Simpson	$(b-a)\dfrac{f(x_0)+4f(x_1)+f(x_2)}{6}$	$-(1/90)h^5f^{(4)}(\xi)$
3	4	Regra 3/8 de Simpson	$(b-a)\dfrac{f(x_0)+3f(x_1)+3f(x_2)+f(x_3)}{8}$	$-(3/80)h^5f^{(4)}(\xi)$
4	5	Regra de Boole	$(b-a)\dfrac{7f(x_0)+32f(x_1)+12f(x_2)+32f(x_3)+7f(x_4)}{90}$	$-(8/945)h^7f^{(6)}(\xi)$
5	6		$(b-a)\dfrac{19f(x_0)+75f(x_1)+50f(x_2)+50f(x_3)+75f(x_4)+19f(x_5)}{288}$	$-(275/12.096)h^7f^{(6)}(\xi)$

Observe que, como aconteceu no caso das regras 1/3 e 3/8 de Simpson, as fórmulas de cinco e seis pontos têm erros da mesma ordem. Essa característica geral vale para fórmulas de mais pontos e leva ao resultado que fórmulas para um número par de segmentos e ímpar de pontos (por exemplo, a regra 1/3 e a regra de Boole) em geral são os métodos preferidos.

Entretanto, deve também ser enfatizado que, na prática da engenharia, as fórmulas de ordem superior (isto é, maior do que quatro pontos) raramente são usadas. As regras de Simpson são suficientes para a maioria das aplicações. A acurácia pode ser melhorada usando-se a versão com aplicações múltiplas. Além disso, quando a função for conhecida e for necessária alta acurácia, métodos como a integração de Romberg ou a quadratura de Gauss, descritas no Capítulo 22, oferecem alternativas viáveis e atrativas.

21.3 INTEGRAÇÃO COM SEGMENTOS DESIGUAIS

Até esse ponto, todas as fórmulas para integração numérica foram baseadas em dados igualmente espaçados. Na prática, existem muitas situações nas quais essa hipótese não é válida e precisamos lidar com segmentos de tamanhos distintos. Por exemplo, dados obtidos experimentalmente, muitas vezes, são desse tipo. Para tais casos, um método é aplicar a regra do trapézio para cada segmento e somar os resultados:

$$I = h_1 \frac{f(x_0) + f(x_1)}{2} + h_2 \frac{f(x_1) + f(x_2)}{2} + \cdots + h_n \frac{f(x_{n-1}) + f(x_n)}{2} \tag{21.22}$$

em que h_i é a largura do segmento i. Observe que essa foi a mesma abordagem usada na aplicação múltipla da regra do trapézio. A única diferença entre as Equações (21.8) e (21.22) é que os h's na primeira são constantes. Conseqüentemente, a Equação (21.8) podia ser simplificada agrupando-se os termos para obter a Equação (21.9). Embora essa simplificação não possa ser aplicada à Equação (21.22), é possível desenvolver facilmente um programa de computador para acomodar segmentos de tamanhos diferentes. Antes de descrever tal algoritmo, vamos ilustrar no seguinte exemplo como a Equação (21.22) é aplicada para calcular uma integral.

EXEMPLO 21.7 Regra do Trapézio com Segmentos Desiguais

Enunciado do Problema. A informação na Tabela 21.3 foi gerada usando o mesmo polinômio utilizado no Exemplo 21.1. Use a Equação (21.22) para determinar a integral a partir desses dados. Lembre que a resposta correta é 1,640533.

Solução. Aplicando a Equação (21.22) aos dados na Tabela 21.3, obtemos

$$I = 0{,}12\frac{1{,}309729 + 0{,}2}{2} + 0{,}10\frac{1{,}305241 + 1{,}309729}{2} + \cdots + 0{,}10\frac{0{,}232 + 2{,}363}{2}$$
$$= 0{,}090584 + 0{,}130749 + \cdots + 0{,}12975 = 1{,}594801$$

o que representa um erro relativo porcentual absoluto de $\varepsilon_t = 2.8\%$.

TABELA 21.3 Dados para $f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$, com valores de x desigualmente espaçados.

x	$f(x)$	x	$f(x)$
0,0	0,200000	0,44	2,842985
0,12	1,309729	0,54	3,507297
0,22	1,305241	0,64	3,181929
0,32	1,743393	0,70	2,363000
0,36	2,074903	0,80	0,232000
0,40	2,456000		

Os dados do Exemplo 21.7 estão mostrados na Figura 21.14. Observe que alguns segmentos adjacentes têm largura igual e, conseqüentemente, poderiam ter sido calculados usando as regras de Simpson. Isso usualmente leva a resultados mais precisos, como ilustrado no exemplo a seguir.

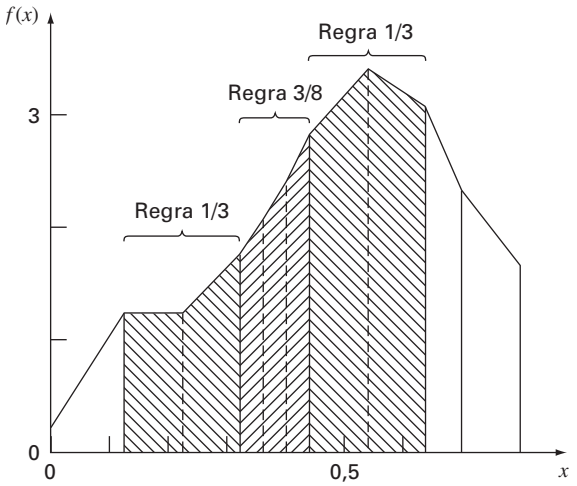


FIGURA 21.14
Uso da regra do trapézio para determinar a integral de dados desigualmente espaçados. Observe como os segmentos hachurados podem ser tratados com a regra de Simpson para se alcançar maior acurácia.

EXEMPLO 21.8 Inclusão das Regras de Simpson no Cálculo com Dados Desiguais

Enunciado do Problema. Recalcule a integral para os dados na Tabela 21.3, mas use as regras de Simpson nos segmentos em que forem apropriadas.

Solução. A área correspondente ao primeiro segmento é calculada com a regra do trapézio:

$$I = 0{,}12\frac{1{,}309729 + 0{,}2}{2} = 0{,}09058376$$

Como os próximos dois segmentos de $x = 0{,}12$ a $0{,}32$ têm comprimentos iguais, sua integral pode ser calculada com a regra $1/3$ de Simpson:

$$I = 0{,}2\frac{1{,}743393 + 4(1{,}305241) + 1{,}309729}{6} = 0{,}2758029$$

Os próximos três segmentos também são iguais e, desse modo, podem ser tratados pela regra $3/8$, o que fornece $I = 0{,}2726863$. Analogamente, a regra $1/3$ pode ser aplicada aos dois segmentos de $x = 0{,}44$ a $0{,}64$ para fornecer $I = 0{,}6684701$. Finalmente, os dois últimos segmentos, cada um com um tamanho diferente, podem ser tratados com a regra do trapézio para fornecer os valores 0,1663479 e 0,1297500, respectivamente. As áreas correspondentes a esses segmentos individuais podem ser somadas para fornecer uma integral total de 1,603641. Isso representa um erro de $\varepsilon_t = 2{,}2\%$, o que é melhor do que o resultado usando-se a regra do trapézio, no Exemplo 21.7.

Programas Computacionais para Dados Desigualmente Espaçados. É uma tarefa muito simples programar a Equação (21.22). Tal algoritmo está listado na Figura 21.15*a*.

Entretanto, como ilustrado pelo Exemplo 21.8, a abordagem é melhorada se as regras de Simpson forem implementadas sempre que possível. Por essa razão, desenvolvemos um segundo algoritmo que incorpora esse recurso. Como descrito na Figura 21.15*b*,

o algoritmo verifica o comprimento de segmentos adjacentes. Se dois segmentos consecutivos tiverem comprimentos iguais, a regra 1/3 de Simpson é aplicada. Se três forem iguais, a regra 3/8 é usada. Quando os segmentos adjacentes forem de comprimentos diferentes, a regra do trapézio é implementada.

Logo, ele não apenas permite o tratamento de dados com segmentos desiguais, mas, se informação igualmente espaçada for usada, ele se reduz ao uso das regras de Simpson. Como tal, ele representa um algoritmo básico, de propósito geral, para a determinação da integral de dados tabulados.

(*a*)

```
FUNCTION Trapun (x, y, n)
  LOCAL i, soma
  soma = 0
  DOFOR i = 1, n
    soma = soma + (x_i − x_{i−1})*(y_{i−1} + y_i)/2
  END DO
  Trapun = soma
END Trapun
```

(*b*)

```
FUNCTION Desigualmente (n,x,f)
  h = x_1 − x_0
  k = 1
  soma = 0.
  DOFOR j = 1, n
    hf = x_{j+1} − x_j
    IF ABS (h − hf) < .000001 THEN
      IF k = 3 THEN
        soma = soma + Simp13 (h, f_{j−3}, f_{j−2}, f_{j−1})
        k = k − 1
      ELSE
        k = k + 1
      END IF
    ELSE
      IF k = 1 THEN
        soma = soma + Trap (h, f_{j−1}, f_j)
      ELSE
        IF k = 2 THEN
          soma = soma + Simp13 (h, f_{j−2}, f_{j−1}, f_j)
        ELSE
          soma = soma + Simp38 (h, f_{j−3}, f_{j−2}, f_{j−1}, f_j)
        END IF
        k = 1
      END IF
    END IF
    h = hf
  END DO
  Desigualmente = soma
END Desigualmente
```

FIGURA 21.15
Pseudocódigo para integrar dados desigualmente espaçados. (*a*) Regra do trapézio e (*b*) combinação das regras de Simpson e do trapézio.

TABELA 21.4 Fórmulas de integração abertas de Newton-Cotes. As fórmulas são apresentadas no formato da Equação (21.5) de modo que os pesos dos dados na estimativa da altura média sejam aparentes. O tamanho do passo é $h = (b - a)/n$.

Segmentos (n)	Pontos	Nome	Fórmula	Erro de Truncamento
2	1	Método do ponto médio	$(b-a)\, f(x_1)$	$(1/3)h^3 f''(\xi)$
3	2		$(b-a)\dfrac{f(x_1)+f(x_2)}{2}$	$(3/4)h^3 f''(\xi)$
4	3		$(b-a)\dfrac{2f(x_1)-f(x_2)+2f(x_3)}{3}$	$(14/45)h^5 f^{[4]}(\xi)$
5	4		$(b-a)\dfrac{11f(x_1)+f(x_2)+f(x_3)+11f(x_4)}{24}$	$(95/144)h^5 f^{[4]}(\xi)$
6	5		$(b-a)\dfrac{11f(x_1)-14f(x_2)+26f(x_3)-14f(x_4)+11f(x_5)}{20}$	$(41/140)h^7 f^{[6]}(\xi)$

21.4 FÓRMULAS DE INTEGRAÇÃO ABERTAS

Lembre-se, da Figura 21.3*b*, que as fórmulas de integração abertas têm extremos que se estendem além do intervalo dos dados. A Tabela 21.4 resume as fórmulas de integração abertas de Newton-Cotes. As fórmulas estão expressas no formato da Equação (21.5) de modo que os fatores de peso são evidentes. Como com as versões fechadas, pares sucessivos de fórmulas têm erros de mesma ordem. As fórmulas para números pares de segmentos e ímpares de pontos são, em geral, os métodos preferidos, pois exigem menos pontos para se obter a mesma acurácia do que as fórmulas para números ímpares de segmentos e pares de pontos.

As fórmulas abertas, em geral, não são usadas para o cálculo de integrais definidas. Entretanto, como discutido no Capítulo 22, terão utilidade na análise de integrais impróprias. Além disso, serão relevantes na discussão de métodos de passos múltiplos para a resolução de equações diferenciais ordinárias no Capítulo 26.

21.5 INTEGRAIS MÚLTIPLAS

As integrais múltiplas são amplamente usadas na engenharia. Por exemplo, uma equação geral para calcular a média de uma função bidimensional pode ser escrita como (lembre-se da Equação (PT6.4))

$$\bar{f} = \frac{\int_c^d \left(\int_a^b f(x, y)\, dx \right) dy}{(d - c)(b - a)} \tag{21.23}$$

O numerador é chamado de integral múltipla.

As técnicas discutidas neste capítulo (e no capítulo seguinte) podem ser facilmente usadas para calcular integrais múltiplas. Um exemplo simples seria tomar a integral dupla de uma função sobre uma área retangular (Figura 21.16).

Lembre-se, do cálculo, que tais integrais podem ser calculadas como integrais iteradas

$$\int_c^d \left(\int_a^b f(x, y)\, dx \right) dy = \int_a^b \left(\int_c^d f(x, y)\, dy \right) dx \tag{21.24}$$

Logo, a integral em uma das dimensões é calculada primeiro. O resultado dessa primeira integração é integrada na segunda dimensão. A Equação (21.24) afirma que a ordem da integração não é importante.

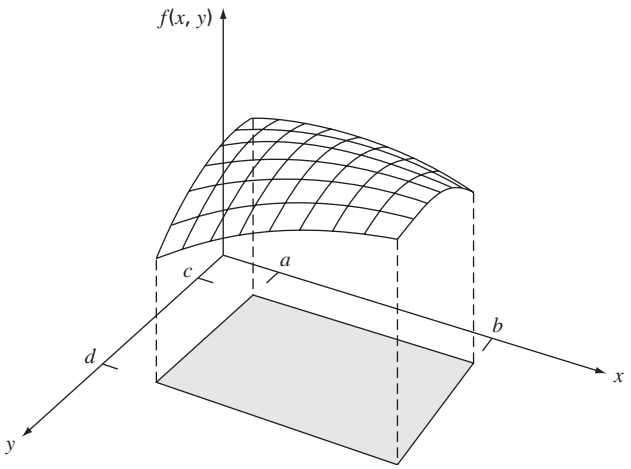


FIGURA 21.16
Integral dupla de uma função sobre uma área retangular.

Uma integral numérica dupla seria baseada na mesma idéia. Primeiro, métodos como a regra do trapézio ou de Simpson para múltiplos segmentos seriam aplicados na primeira dimensão para cada valor da segunda dimensão mantido constante. Então, o método seria aplicado para integrar na segunda dimensão. A abordagem é ilustrada no exemplo seguinte.

EXEMPLO 21.9 Usando uma Integral Dupla para Determinar a Temperatura Média

Enunciado do Problema. Suponha que a temperatura de uma placa retangular aquecida seja descrita pela seguinte função:

$$T(x, y) = 2xy + 2x - x^2 - 2y^2 + 72$$

Se a placa tiver 8 m de comprimento (dimensão x) e 6 m de largura (dimensão y), calcule a temperatura média.

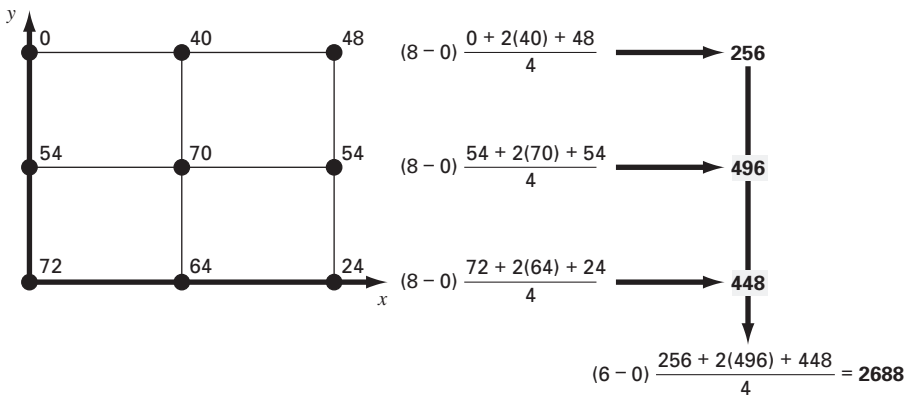


FIGURA 21.17
Cálculo numérico de uma integral dupla usando a regra do trapézio para dois segmentos.

Solução. Primeiro, vamos simplesmente usar aplicações da regra do trapézio com dois segmentos em cada dimensão. As temperaturas nos valores necessário de x e y estão descritas na Figura 21.17. Observe que uma média simples desses valores é 47,33. A média da função pode ser calculada analiticamente, fornecendo um resultado de 58,66667.

Para fazer o mesmo cálculo numericamente, a regra do trapézio é primeiro implementada ao longo da dimensão x para cada valor de y. Esses valores, então, são integrados ao longo da direção y para dar o resultado final de 2.688. Dividir isso pela área fornece a temperatura média como sendo $2.688/(6 \times 8) = 56$.

Agora, podemos aplicar uma regra 1/3 de Simpson para um único segmento da mesma maneira, obtendo uma integral de 2.816 e uma média de 58,66667, que é exata. Por que isso ocorre? Lembre que a regra 1/3 de Simpson fornece resultados perfeitos para polinômios cúbicos. Como o termo de grau mais alto na função é de segundo grau, o mesmo resultado ocorre no caso presente.

Para funções algébricas de grau mais alto, bem como para funções transcendentais, seria necessário usar aplicações de segmentos múltiplos para alcançar uma estimativa acurada da integral. Além disso, o Capítulo 22 introduz técnicas que são mais eficientes do que a fórmula de Newton-Cotes para calcular integrais de funções dadas. Elas, em geral, fornecem uma forma superior para implementar integrações numéricas de integrais múltiplas.

PROBLEMAS

21.1 Calcule a seguinte integral:

$$\int_0^4 \left(1 - e^{-2x}\right) dx$$

(**a**) analiticamente; (**b**) por uma única aplicação da regra do trapézio; (**c**) por aplicações múltiplas da regra do trapézio, com $n = 2$ e 4; (**d**) uma única aplicação da regra 1/3 de Simpson; (**e**) aplicação múltipla da regra 1/3 de Simpson, com $n = 4$; (**f**) uma única aplicação da regra 3/8 de Simpson; e (**g**) aplicação múltipla da regra de Simpson, com $n = 5$. Para cada estimativa numérica de (**b**) a (**g**), determine o erro relativo porcentual com base em (**a**).

21.2 Calcule a seguinte integral:

$$\int_0^{\pi/2} (6 + 3\cos x)\, dx$$

(**a**) analiticamente; (**b**) por uma única aplicação da regra do trapézio; (**c**) por aplicações múltiplas da regra do trapézio, com $n = 2$ e 4; (**d**) uma única aplicação da regra 1/3 de Simpson; (**e**) aplicação múltipla da regra 1/3 de Simpson, com $n = 4$; (**f**) uma única aplicação da regra 3/8 de Simpson; e (**g**) aplicação múltipla da regra de Simpson, com $n = 5$. Para cada estimativa numérica de (**b**) a (**g**), determine o erro relativo porcentual com base em (**a**).

21.3 Calcule a seguinte integral:

$$\int_{-2}^4 (1 - x - 4x^3 + 2x^5)\, dx$$

(**a**) analiticamente; (**b**) por uma única aplicação da regra do trapézio; (**c**) por aplicações múltiplas da regra do trapézio, com $n = 2$ e 4; (**d**) uma única aplicação da regra 1/3 de Simpson; (**e**) regra 3/8 de Simpson; (**f**) regra de Boole. Para cada estimativa numérica de (**b**) a (**f**), determine o erro relativo porcentual com base em (**a**).

21.4 Integre a seguinte função analiticamente e usando a regra do trapézio, com $n = 1$, 2, 3 e 4:

$$\int_1^2 (x + 2/x)^2\, dx$$

Use a solução analítica para calcular os erros relativos porcentuais verdadeiros para avaliar a precisão das aproximações por trapézios.

21.5 Integre a seguinte função tanto analiticamente quanto usando as regras de Simpson com $n = 4$ e 5. Discuta os resultados.

$$\int_{-3}^5 (4x - 3)^3\, dx$$

21.6 Integre a seguinte função tanto analiticamente quanto numericamente. Use as regras do trapézio e 1/3 de Simpson para integrar numericamente a função. Em ambos os casos, use a versão com aplicações múltiplas, com $n = 4$. Calcule os erros relativos porcentuais para os resultados numéricos.

$$\int_0^3 x^2 e^x\, dx$$

21.7 Integre a seguinte função tanto analiticamente quanto numericamente. Para os cálculos numéricos, use (**a**) uma única aplicação da regra do trapézio, (**b**) a regra 1/3 de Simpson, (**c**) a regra 3/8 de Simpson, (**d**) a regra de Boole, (**e**) o método do ponto médio, (**f**) a fórmula de integração aberta com três segmentos e dois pontos e (**g**) a fórmula de integração aberta com quatro segmentos e três pontos. Calcule os erros relativos porcentuais para os resultados numéricos.

$$\int_{0,5}^{1,5} 14^{2x}\, dx$$

21.8 Integre a seguinte função tanto analiticamente quanto numericamente. Para os cálculos numéricos, use (**a**) uma única aplicação da regra do trapézio; (**b**) a regra 1/3 de Simpson; (**c**) a regra 3/8 de Simpson; (**d**) aplicação múltipla das regras de Simpson, com $n = 5$; (**e**) a regra de Boole; (**f**) o método do ponto médio; (**g**) a fórmula de integração aberta com três segmentos e dois pontos e (**h**) a fórmula de integração aberta com quatro segmentos e três pontos.

$$\int_0^3 (5 + 3\cos x)\, dx$$

Calcule os erros relativos porcentuais para os resultados numéricos.

21.9 Suponha que uma força para cima de resistência do ar em um objeto em queda livre seja proporcional ao quadrado da velocidade. Nesse caso, a velocidade pode ser calculada por

$$v(t) = \sqrt{\frac{gm}{c_d}}\, \text{tgh}\left(\sqrt{\frac{gc_d}{m}}t\right)$$

em que c_d é o coeficiente de arrasto de segunda ordem. (**a**) Se $g = 9{,}8$ m/s^2, $m = 68{,}1$ kg e $c_d = 0{,}25$ kg/m, use integração analítica para determinar quanto o objeto cai em 10 s. (**b**) Faça o mesmo cálculo, mas calcule a integral com a regra dos trapézios com segmentos múltiplos. Use um n suficientemente grande para que você consiga uma acurácia de três algarismos significativos.

21.10 Calcule a integral dos seguintes dados tabulares com (**a**) a regra do trapézio e (**b**) regras de Simpson:

x	0	0,1	0,2	0,3	0,4	0,5
$f(x)$	1	8	4	3,5	5	1

21.11 Calcule a integral dos seguintes dados tabulares com (**a**) a regra do trapézio e (**b**) regras de Simpson:

x	−2	0	2	4	6	8	10
$f(x)$	35	5	−10	2	5	3	20

21.12 Determine o valor médio da função

$$f(x) = -46 + 45x - 14x^2 + 2x^3 - 0{,}075x^4$$

entre $x = 2$ e 10: (**a**) traçando o gráfico da função e estimando visualmente o valor médio, (**b**) usando a Equação (PT6.4) e o cálculo analítico da integral e (**c**) usando a Equação (PT6.4) e uma versão da regra de Simpson para cinco segmentos para fazer uma estimativa da integral. Calcule o erro porcentual relativo.

21.13 A função $f(x) = 2e^{-1{,}5x}$ pode ser usada para gerar a seguinte tabela de dados desigualmente espaçados:

x	0	0,05	0,15	0,25	0,35	0,475	0,6
$f(x)$	2	1,8555	1,5970	1,3746	1,1831	0,9808	0,8131

Calcule a integral de $a = 0$ a $b = 0{,}6$ usando (**a**) métodos analíticos, (**b**) a regra do trapézio e (**c**) uma combinação das regras do trapézio e de Simpson; empregue as regras de Simpson sempre que possível para obter a máxima acurácia. Para (**b**) e (**c**), calcule o erro relativo porcentual (ε_t).

21.14 Calcule a seguinte integral dupla:

$$\int_{-1}^1 \int_0^2 (x^2 - 2y^2 + xy^3)\, dx\, dy$$

(**a**) analiticamente; (**b**) usando a aplicação múltipla da regra do trapézio com $n = 2$, e (**c**) usando uma aplicação única da regra 1/3 de Simpson. Para (**b**) e (**c**), calcule o erro relativo porcentual (ε_t).

21.15 Calcule a seguinte integral tripla (**a**) analiticamente e (**b**) usando uma única aplicação da regra 1/3 de Simpson. Para (**b**), calcule o erro relativo porcentual (ε_t).

$$\int_{-2}^{2}\int_{0}^{2}\int_{-3}^{1}(x^3 - 3yz)\,dx\,dy\,dz$$

21.16 Desenvolva um programa de computador amigável ao usuário para a aplicação múltipla da regra do trapézio com base na Figura 21.9. Teste seu programa reproduzindo os cálculos do Exemplo 21.2.

21.17 Desenvolva um programa de computador amigável ao usuário para aplicação múltipla da regra de Simpson com base na Figura 21.13*c*. Teste seu programa reproduzindo os cálculos do Exemplo 21.5.

21.18 Desenvolva um programa de computador amigável ao usuário para integrar dados desigualmente espaçados com base na Figura 21.15*b*. Teste-o reproduzindo os cálculos do Exemplo 21.8.

21.19 Uma viga de 11 m está sujeita a uma carga e a força de cisalhamento segue a equação

$$V(x) = 5 + 0{,}25x^2$$

em que V é a força de cisalhamento e x é a distância ao longo da viga. Sabemos que $V = dM/dx$ e M é o momento de deformação. A integração fornece a relação

$$M = M_o + \int_0^x V\,dx$$

Se M_o for 0 e $x = 11$, calcule M usando (**a**) integração analítica, (**b**) aplicação múltipla da regra do trapézio e (**c**) aplicação múltipla das regras de Simpson. Para (**b**) e (**c**), use incrementos de 1 m.

21.20 O trabalho produzido por um processo termodinâmico a temperatura, constante pode ser calculado por

$$W = \int p\,dV$$

em que W é o trabalho, p é a pressão e V é o volume. Com uma combinação da regra do trapézio, regra 1/3 de Simpson e regra 3/8 de Simpson, use os seguintes dados para calcular o trabalho em kJ (kJ = kN · m):

Pressão (kPa)	336	294,4	266,4	260,8	260,5	249,6	193,6	165,6
Volume (m^3)	0,5	2	3	4	6	8	10	11

21.21 Determine a distância percorrida para os seguintes dados:

t, min	1	2	3,25	4,5	6	7	8	9	9,5	10
v, m/s	5	6	5,5	7	8,5	8	6	7	7	5

(a) Use a regra dos trapézios, (**b**) a melhor combinação das regras do trapézio e de Simpson e (**c**) a integração analítica de polinômios de segundo e terceiro graus determinados por regressão.

21.22 A massa total de uma haste de densidade variável é dada por

$$m = \int_0^L \rho(x)A_c(x)\,dx$$

em que m é a massa, $\rho\,(x)$ é a densidade, $A_c(x)$ é a área da seção transversal, x é a distância ao longo da haste e L é o comprimento total da haste. Os seguintes dados foram medidos para uma haste de 10 m. Determine a massa em quilogramas com a melhor precisão possível.

x, m	0	2	3	4	6	8	10
ρ, g/cm^3	4,00	3,95	3,89	3,80	3,60	3,41	3,30
A_c, cm^2	100	103	106	110	120	133	150

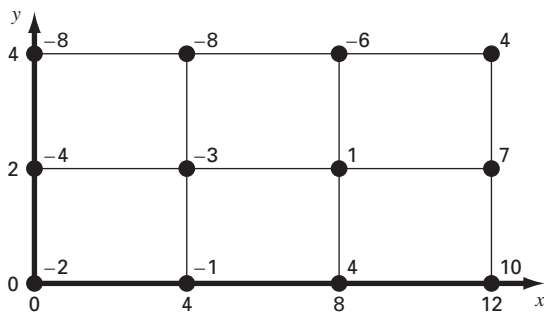
21.23 Um estudo em engenharia de transporte exige que você determine o número de carros que passa por uma intersecção durante o horário de *rush* da manhã. Você fica em um lado da estrada e conta o número de carros que passa a cada 4 minutos em diferentes instantes, como tabulado a seguir. Use o melhor método numérico para determinar (**a**) o número total de carros que passa entre 7:30 e 9:15 e (**b**) a taxa de carros passando pela intersecção por minuto. (*Sugestão*: tenha cuidado com as unidades.)

Horário (h)	7:30	7:45	8:00	8:15	8:45	9:15
Taxa (carros em 4 min.)	18	24	14	24	21	9

21.24 Determine o valor médio para os dados na Figura P21.24. Faça a integral necessária para a média na ordem mostrada na seguinte equação:

$$I = \int_{x_0}^{x_n}\left[\int_{y_0}^{y_m} f(x, y)dy\right]dx$$

Figura P21.24



CAPÍTULO 22

Integração de Equações

Na introdução da Parte Seis, observamos que as funções a serem integradas numericamente terão tipicamente duas formas: uma tabela de valores ou uma equação. A forma dos dados tem uma influência importante nas abordagens usadas para calcular a integral. Para informações tabuladas, fica-se limitado pelo número de pontos dados. Em contraste, se uma equação estiver disponível, é possível gerar tantos valores de $f(x)$ quantos forem necessários para alcançar a precisão aceitável (lembre-se da Figura PT6.7).

Este capítulo é dedicado a duas técnicas desenvolvidas especialmente para analisar casos nos quais a função é dada. Ambas tiram vantagem da possibilidade de gerar valores da função para desenvolver esquemas eficientes para integração numérica. A primeira é baseada na *extrapolação de Richardson*, que é um método para combinar duas estimativas numéricas para a integral para obter um terceiro valor, mais acurado. O algoritmo computacional para implementar a extrapolação de Richardson de uma maneira altamente eficiente é chamado *integração de Romberg*. Essa técnica é recursiva e pode ser usada para gerar uma estimativa da integral dentro de uma tolerância de erro pré-especificada.

O segundo método é chamado *quadratura de Gauss*. Lembre que, no último capítulo, os valores de $f(x)$ para as fórmulas de Newton-Cotes foram determinados em valores especificados de x. Por exemplo, se usássemos a regra do trapézio para determinar uma integral, estaríamos restritos a tomar a média ponderada de $f(x)$ nas extremidades do intervalo. As fórmulas de quadratura de Gauss empregam valores de x que estão posicionados entre a e b de uma maneira que é produzida uma estimativa da integral muito mais acurada.

Além dessas duas técnicas padrão, dedicamos uma seção final ao cálculo de *integrais impróprias*. Nessa discussão, nos concentraremos em integrais com extremos infinitos e mostraremos como uma mudança de variáveis e fórmulas de integração abertas são úteis em tais casos.

22.1 ALGORITMO DE NEWTON-COTES PARA EQUAÇÕES

No Capítulo 21, apresentamos algoritmos para versões com aplicações múltiplas da regra do trapézio e das regras de Simpson. Embora esses pseudocódigos possam certamente ser usados para analisar equações, em nosso esforço para torná-los compatíveis tanto com dados quanto com funções, eles não exploram a conveniência das últimas.

A Figura 22.1 mostra pseudocódigos que são desenvolvidos especialmente para os casos nos quais a função é dada analiticamente. Em particular, observe que nem os valores da variável independente nem os da dependente são passados para a função no código por meio de seus argumentos, como era o caso nos códigos do Capítulo 21. Para a variável independente x, são passados o intervalo de integração (a, b) e o número de segmentos. Essa informação, então, é usada para gerar valores igualmente espaçados de x para essa função. Para a variável dependente, os valores da função na Figura 22.1 são calculados usando chamadas da função $f(x)$ que está sendo analisada.

Desenvolvemos programas em precisão simples baseados nesses pseudocódigos para analisar o esforço envolvido e os erros cometidos conforme usamos progressivamente mais segmentos para fazer uma estimativa da integral de uma função simples. Para uma função definida analiticamente, as equações de erro [Equações (21.13) e (21.19)] indicam que, aumentando o número n de segmentos, obteremos estimativas mais acuradas da integral. Essa observação é evidenciada pela Figura 22.2, que é um gráfico do

(*a*)Regra

```
FUNCTION TrapEq (n, a, b)
  h = (b − a) / n
  x = a
  soma = f(x)
  DOFOR i = 1, n − 1
    x = x + h
    soma = soma + 2 * f(x)
  END DO
  soma = soma + f(b)
  TrapEq = (b − a) * soma / (2 * n)
END TrapEq
```

(*b*)

```
FUNCTION SimpEq (n, a, b)
  h = (b − a) / n
  x = a
  soma = f(x)
  DOFOR i = 1, n − 2, 2
    x = x + h
    soma = soma + 4 * f(x)
    x = x + h
    soma = soma + 2 * f(x)
  END DO
  x = x + h
  soma = soma + 4 * f(x)
  soma = soma + f(b)
  SimpEq = (b − a) * soma/(3 * n)
END SimpEq
```

FIGURA 22.1
Algoritmos para a aplicação múltipla das regras (*a*) do trapézio e (*b*) 1/3 de Simpson, quando a função está disponível.

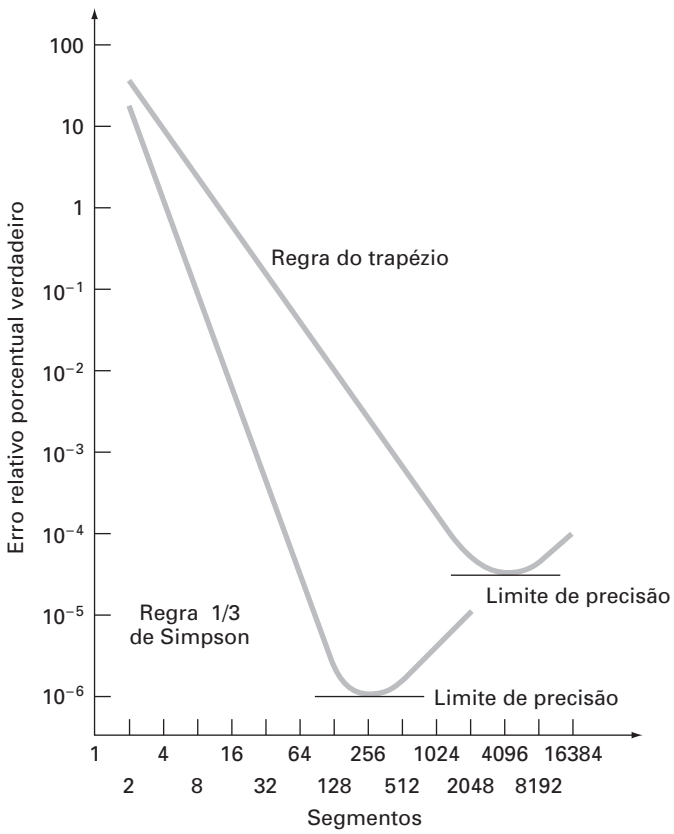


FIGURA 22.2
Valor absoluto do erro relativo porcentual verdadeiro em função do número de segmentos para a determinação da integral de $f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$, calculada de $a = 0$ a $b = 0{,}8$ usando aplicações múltiplas da regra do trapézio e aplicações múltiplas da regra 1/3 de Simpson. Observe que ambos os resultados indicam que, para um grande número de segmentos, os erros de arredondamento limitam a precisão.

erro verdadeiro em função de n para a integral de $f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$. Observe como o erro cai conforme n aumenta. Entretanto, observe também que, para valores grandes de n, o erro começa a aumentar à medida que os erros de arredondamento começam a dominar. Observe ainda que valores muito grandes de cálculos de função (e, portanto, de esforço computacional) são necessários para se alcançar altos níveis de acurácia. Como uma conseqüência desses defeitos, a aplicação múltipla da regra do trapézio e das regras de Simpson algumas vezes é inadequada para contextos de problemas nos quais alta eficiência e pequenos erros são necessários.

22.2 INTEGRAÇÃO DE ROMBERG

A *integração de Romberg* é uma técnica desenvolvida para produzir integrais numéricas eficientes de funções. É bastante parecida com as técnicas discutidas no Capítulo 21, no sentido de que é baseada em aplicações sucessivas da regra dos trapézios. Entretanto, mediante manipulações matemáticas, são obtidos resultados superiores com menos esforço.

22.2.1 Extrapolação de Richardson

Lembre que, na Seção 10.3.3, usamos refinamento iterativo para melhorar a solução de um conjunto de equações lineares simultâneas. Também estão disponíveis técnicas de correções de erros para melhorar os resultados da integração numérica com base nas próprias estimativas da integral. Chamados em geral de *extrapolação de Richardson*, esses métodos usam duas estimativas de uma integral para calcular uma terceira aproximação mais acurada.

A estimativa e o erro associados com a aplicação múltipla da regra do trapézio podem ser representados de modo geral por

$$I = I(h) + E(h)$$

em que I é o valor exato da integral, $I(h)$ é a aproximação a partir da aplicação da regra do trapézio com n segmentos com tamanho do passo $h = (b - a)/n$ e $E(h)$ é o erro de truncamento. Se fizermos duas estimativas separadas usando tamanhos de passo h_1 e h_2 e tivermos os valores exatos para o erro,

$$I(h_1) + E(h_1) = I(h_2) + E(h_2) \tag{22.1}$$

Lembre-se agora de que o erro da aplicação múltipla da regra do trapézio pode ser representado aproximadamente pela Equação (21.13)[com $n = (b - a)/h$]:

$$E \cong -\frac{b-a}{12}h^2\bar{f}'' \tag{22.2}$$

Supondo que $\bar{f}''$ é constante independentemente do tamanho do passo, a Equação (22.2) pode ser usada para determinar que a razão entre os dois erros será

$$\frac{E(h_1)}{E(h_2)} \cong \frac{h_1^2}{h_2^2} \tag{22.3}$$

Esse cálculo tem o efeito importante de remover o termo $\bar{f}''$ do cálculo. Fazendo isso, tornamos possível a utilização da informação embutida na Equação (22.2) sem o conhecimento prévio da segunda derivada da função. Para fazer isso, reorganizamos a Equação (22.3) para obter

$$E(h_1) \cong E(h_2)\left(\frac{h_1}{h_2}\right)^2$$

a qual pode ser substituída na Equação (22.1):

$$I(h_1) + E(h_2)\left(\frac{h_1}{h_2}\right)^2 \cong I(h_2) + E(h_2)$$

a qual pode ser resolvida por

$$E(h_2) \cong \frac{I(h_1) - I(h_2)}{1 - (h_1/h_2)^2}$$

Logo, obteremos uma estimativa para o erro de truncamento em termos das estimativas da integral e seus tamanhos de passo. Essa estimativa pode então ser substituída em

$$I = I(h_2) + E(h_2)$$

para fornecer uma estimativa melhorada da integral:

$$I \cong I(h_2) + \frac{1}{(h_1/h_2)^2 - 1}[I(h_2) - I(h_1)] \tag{22.4}$$

É possível mostrar (Ralston e Rabinowitz, 1978) que o erro nessa estimativa é $O(h^4)$. Logo, combinamos as estimativas de duas regras de trapézio de $O(h^2)$ para obter uma nova estimativa de $O(h^4)$. Para o caso especial em que o intervalo é dividido na metade ($h_2 = h_1/2$), essa equação se torna

$$I \cong I(h_2) + \frac{1}{2^2 - 1}[I(h_2) - I(h_1)]$$

ou, agrupando os termos,

$$I \cong \frac{4}{3}I(h_2) - \frac{1}{3}I(h_1) \tag{22.5}$$

EXEMPLO 22.1 Correções do Erro da Regra do Trapézio

Enunciado do Problema. No capítulo anterior (Exemplo 21.1 e Tabela 21.1), usamos diversos métodos de integração numérica para calcular a integral de $f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$ de $a = 0$ a $b = 0{,}8$. Por exemplo, aplicações única e múltipla da regra do trapézio forneceram os seguintes resultados:

Segmentos	h	Integral	ε_t, %
1	0,8	0,1728	89,5
2	0,4	1,0688	34,9
4	0,2	1,4848	9,5

Use essa informação junto com a Equação (22.5) para calcular estimativas melhoradas da integral.

Solução. As estimativas para um e dois segmentos podem ser combinadas para fornecer

$$I \cong \frac{4}{3}(1{,}0688) - \frac{1}{3}(0{,}1728) = 1{,}367467$$

O erro da integral melhorada é $E_t = 1{,}640533 - 1{,}367467 = 0{,}273067$ ($\varepsilon_t = 16{,}6\%$), que é superior às estimativas sobre as quais é baseado.

Da mesma maneira, as estimativas para dois e quatro segmentos podem ser combinadas para dar

$$I \cong \frac{4}{3}(1{,}4848) - \frac{1}{3}(1{,}0688) = 1{,}623467$$

o que representa um erro de $E_t = 1{,}640533 - 1{,}623467 = 0{,}017067$ ($\varepsilon_t = 1{,}0\%$).

A Equação (22.4) fornece uma forma de combinar duas aplicações da regra do trapézio com erro $O(h^2)$ para calcular uma terceira estimativa com erro $O(h^4)$. Essa abordagem é um subconjunto de um método mais geral para combinar integrais e obter estimativas melhores. No caso do Exemplo 22.1, calculamos duas integrais melhoradas de $O(h^4)$ com base em três estimativas pela regra do trapézio. Essas duas estimativas melhoradas podem, por sua vez, ser combinadas para fornecer um valor ainda melhor com $O(h^6)$. Para o caso especial em que as estimativas originais pela regra do trapézio são

baseadas em dividir o tamanho do passo sucessivamente por 2, a equação usada para uma acurácia $O(h^6)$ é

$$I \cong \frac{16}{15} I_m - \frac{1}{15} I_l \tag{22.6}$$

em que I_m e I_l são as estimativas mais e menos acuradas, respectivamente. Analogamente, dois resultados $O(h^6)$ podem ser combinados para calcular uma integral que é $O(h^8)$ usando

$$I \cong \frac{64}{63} I_m - \frac{1}{63} I_l \tag{22.7}$$

EXEMPLO 22.2 Correção do Erro de Ordem Superior das Estimativas da Integral

Enunciado do Problema. No Exemplo 22.1, usamos a extrapolação de Richardson para calcular duas estimativas da integral de $O(h^4)$. Utilize a Equação (22.6) para combinar essas estimativas para calcular uma integral com $O(h^6)$.

Solução. As duas estimativas da integral $O(h^4)$ obtidas no Exemplo 22.1 foram 1,367467 e 1,623467. Esses valores podem ser substituídos na Equação (22.6) para fornecer

$$I = \frac{16}{15}(1{,}623467) - \frac{1}{15}(1{,}367467) = 1{,}640533$$

que é a resposta correta até os sete algarismos significativos que são usados neste exemplo.

22.2.2 O Algoritmo da Integração de Romberg

Observe que os coeficientes em cada uma das equações de extrapolação [Equações (22.5), (22.6) e (22.7)] somam 1. Logo, eles representam fatores de peso que, conforme a precisão aumenta, colocam peso relativamente maior na estimativa superior da integral. Essas formulações podem ser expressas em uma forma geral que é adequada para a implementação computacional:

$$I_{j,k} \cong \frac{4^{k-1} I_{j+1,k-1} - I_{j,k-1}}{4^{k-1} - 1} \tag{22.8}$$

onde $I_{j+1,k-1}$ e $I_{j,k-1}$ são as integrais mais e menos acuradas, respectivamente, e $I_{j,k}$ é a integral melhorada. O índice k significa o nível da integração, em que $k = 1$ corresponde à estimativa original pela regra do trapézio, $k = 2$ corresponde a $O(h^4)$, $k = 3$ a $O(h^6)$, e assim por diante. O índice j é usado para distinguir entre a estimativa mais acurada $(j + 1)$ e a menos acurada (j). Por exemplo, para $k = 2$ e $j = 1$, a Equação (22.8) se torna

$$I_{1,2} \cong \frac{4I_{2,1} - I_{1,1}}{3}$$

que é equivalente à Equação (22.5).

A forma geral representada pela Equação (22.8) é atribuída a Romberg e sua aplicação sistemática para o cálculo de integrais é conhecida como *integração de Romberg*. A Figura 22.3 é uma descrição gráfica de uma seqüência de estimativas da integral gerada usando essa abordagem. Cada matriz corresponde a uma única iteração. A primeira coluna contém os cálculos da regra do trapézio designados por $I_{j,1}$, em que $j = 1$ é para uma aplicação com um único segmento (o tamanho do passo é $b - a$), $j = 2$ é para

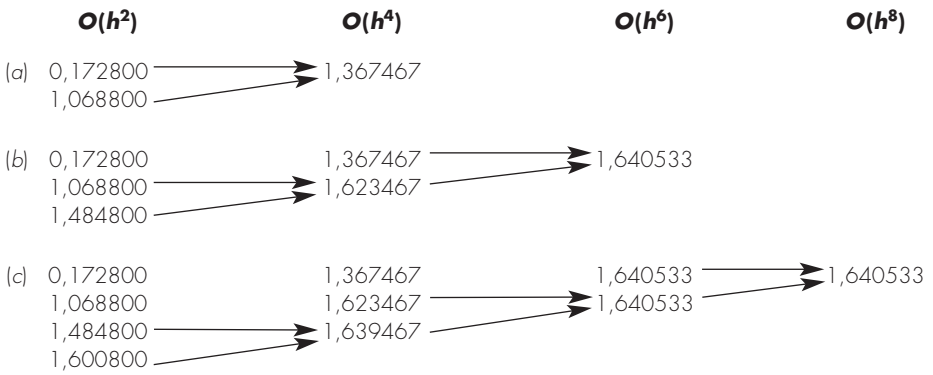


FIGURA 22.3
Descrição gráfica de uma seqüência de estimativas da integral gerada usando a integração de Romberg. (*a*) Primeira iteração. (*b*) Segunda iteração. (*c*) Terceira iteração.

uma aplicação com dois segmentos [o tamanho do passo é $b - a)/2]$, $j = 3$ é para uma aplicação com quatro segmentos [o tamanho do passo é $(b - a)/4]$ e assim por diante. As outras colunas da matriz são geradas pela aplicação sistemática da Equação (22.8) para se obter estimativas sucessivamente melhores da integral.

Por exemplo, a primeira iteração (Figura 22.3*a*) envolve o cálculo de estimativas pela regra do trapézio com um e dois segmentos ($I_{1,1}$ e $I_{2,1}$). A Equação (22.8) é então usada para calcular o elemento $I_{1,2} = 1{,}367467$, o qual tem um erro de $O(h^4)$.

Agora, devemos determinar se esse resultado é adequado para as nossas necessidades. Como em outros métodos aproximadores deste livro, um critério de parada é necessário para avaliar a acurácia dos resultados. Um método que pode ser empregado para o propósito presente é [Equação (3.5)]

$$|\varepsilon_a| = \left| \frac{I_{1,k} - I_{2,k-1}}{I_{1,k}} \right| 100\% \tag{22.9}$$

em que ε_a é uma estimativa do erro porcentual. Assim, como foi feito previamente em outros processos iterativos, comparamos a nova estimativa com um valor anterior. Quando a variação entre os valores antigo e novo, representados por ε_a, ficar abaixo de um critério de erro pré-especificado ε_s, os cálculos são parados. Na Figura 22.3*a*, esse cálculo indica uma variação porcentual de 21,8 durante a primeira iteração.

O objetivo da segunda iteração (Figura 22.3*b*) é obter a estimativa $O(h^6)$ — $I_{1,3}$. Para fazer isso, uma estimativa adicional da regra do trapézio, $I_{3,1} = 1{,}4848$, é determinada. Então, ela é combinada com $I_{2,1}$ usando a Equação (22.8) para gerar $I_{2,2} = 1{,}623467$. O resultado é, por sua vez, combinado com $I_{1,2}$ para fornecer $I_{1,3} = 1{,}640533$. A Equação (22.9) pode ser aplicada para determinar que esse resultado representa uma variação de 1,0% quando comparado com o resultado anterior, $I_{1,2}$.

A terceira iteração (Figura 22.3*c*) continua o processo da mesma forma. Nesse caso, é adicionada uma estimativa pela regra do trapézio à primeira coluna e, então, a Equação (22.8) é aplicada para calcular integrais sucessivamente mais acuradas ao longo da diagonal inferior. Depois de apenas três iterações, como estamos tratando um polinômio de quinto grau, o resultado ($I_{1,4} = 1{,}640533$) é exato.

A integração de Romberg é mais eficiente do que a regra do trapézio e as regras de Simpson discutidas no Capítulo 21. Por exemplo, para determinar a integral como mostrado na Figura 22.1, a regra 1/3 de Simpson exigiria uma aplicação com 256 segmentos para fornecer uma estimativa de 1,640533. Aproximações mais finas não seriam possíveis por causa dos erros de arredondamento. Em contraste, a integração de Romberg fornece um resultado exato (até sete algarismos significativos) baseado na combinação das regras do trapézio para um, dois, quatro e oito segmentos; isto é, com apenas 15 cálculos de função!

A Figura 22.4 apresenta um pseudocódigo para a integração de Romberg. Usando laços, esse algoritmo implementa o método eficientemente. A integração de Romberg é desenvolvida para os casos nos quais a função a ser integrada é conhecida, porque o conhecimento da função permite os cálculos necessários para as implementações iniciais da

```
FUNCTION Rhomberg (a, b, maxit, es)
  LOCAL I(10, 10)
  n = 1
  I(1,1) = TrapEq(n, a, b)
  iter = 0
  DO
    iter = iter + 1
    n = 2^iter
    I(iter+1,1) = TrapEq(n, a, b)
    DOFOR k = 2, iter + 1
      j = 2 + iter - k
      I(j,k) = (4^(k-1) * I(j+1,k-1) - I(j,k-1)) / (4^(k-1) - 1)
    END DO
    ea = ABS((I(1,iter+1) - I(1,iter)) / I(1,iter+1)) * 100
    IF (iter ≥ maxit OR ea ≤ es) EXIT
  END DO
  Rhomberg = I(1,iter+1)
END Rhomberg
```

FIGURA 22.4
Pseudocódigo para a integração de Romberg que usa a versão com segmentos de tamanhos iguais da regra do trapézio da Figura 22.1.

regra do trapézio. Dados tabulados raramente estão na forma em que deveriam para se fazer as sucessivas divisões por dois que são necessárias.

22.3 QUADRATURA DE GAUSS

No Capítulo 21, estudamos o grupo de fórmulas de integração numérica ou quadratura conhecido como equações de Newton-Cotes. Uma característica dessas fórmulas (com a exceção do caso especial da Seção 21.3) era que a estimativa da integral era baseada em valores da função igualmente espaçados. Conseqüentemente, a posição dos pontos base usados nessas equações era predeterminada ou fixa.

Por exemplo, como descrito na Figura 22.5*a*, a regra do trapézio é baseada em tomar a área sob uma reta ligando os valores da função nas extremidades do intervalo de integração. A fórmula que é usada para calcular essa área é

$$I \cong (b - a)\frac{f(a) + f(b)}{2} \tag{22.10}$$

em que a e b são os extremos de integração e $b - a$ é a largura do intervalo de integração. Como a regra do trapézio deve passar pelas extremidades, há casos como a Figura 22.5*a* nos quais a fórmula resulta em um erro grande.

Agora, suponha que a restrição de pontos base fixos seja removida e que estivéssemos livres para calcular a área sob uma reta ligando quaisquer dois pontos da curva. Posicionando sabiamente esses pontos, poderíamos definir uma reta que balanceasse os erros positivos e negativos. Assim, como na Figura 22.5*b*, chegaríamos a uma estimativa melhorada da integral.

Quadratura de Gauss é o nome de uma classe de técnicas para implementar tal estratégia. As fórmulas de quadratura de Gauss descritas nessa seção são chamadas de *fórmulas de Gauss-Legendre*. Antes de descrever essa abordagem, vamos mostrar como as fórmulas de integração numérica tal como a regra do trapézio podem ser deduzidas usando-se o método dos coeficientes indeterminados. Esse método será então empregado para deduzir as fórmulas de Gauss-Legendre.

22.3.1 Método dos Coeficientes Indeterminados

No Capítulo 21, deduzimos a regra do trapézio integrando um polinômio interpolador linear e por argumentação geométrica. O método dos coeficientes indeterminados oferece

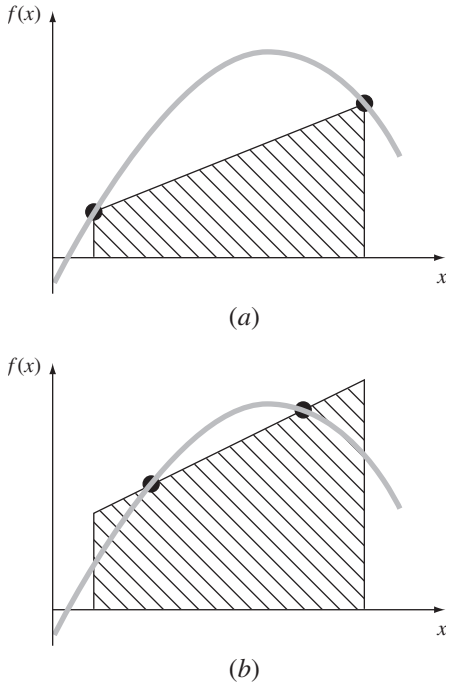


FIGURA 22.5
(*a*) Descrição gráfica da regra do trapézio como a área sob a reta ligando extremidades fixas. (*b*) Uma estimativa melhorada da integral obtida tomando-se a área sob a reta passando por dois pontos intermediários. Posicionando sabiamente esses pontos, os erros positivos e negativos são balanceados e uma estimativa melhorada da integral é obtida.

uma terceira abordagem que também tem utilidade na dedução de outras técnicas de integração, como a quadratura de Gauss.

Para ilustrar essa abordagem, a Equação (22.10) é expressa como

$$I \cong c_0 f(a) + c_1 f(b) \tag{22.11}$$

em que os c's são constantes. Perceba, agora, que a regra do trapézio deveria fornecer resultados exatos quando a função sendo integrada é uma constante ou uma reta. Duas equações simples que representam esses casos são $y = 1$ e $y = x$. Ambas estão ilustradas na Figura 22.6. Logo, as seguintes igualdades devem ser válidas:

$$c_0 + c_1 = \int_{-(b-a)/2}^{(b-a)/2} 1\, dx$$

e

$$-c_0 \frac{b-a}{2} + c_1 \frac{b-a}{2} = \int_{-(b-a)/2}^{(b-a)/2} x\, dx$$

ou, calculando as integrais,

$$c_0 + c_1 = b - a$$

e

$$-c_0 \frac{b-a}{2} + c_1 \frac{b-a}{2} = 0$$

Essas são duas equações com duas incógnitas que podem ser resolvidas por

$$c_0 = c_1 = \frac{b-a}{2}$$

o que, quando se substitui de volta na Equação (22.11), dá

$$I = \frac{b-a}{2} f(a) + \frac{b-a}{2} f(b)$$

o que é equivalente à regra do trapézio.

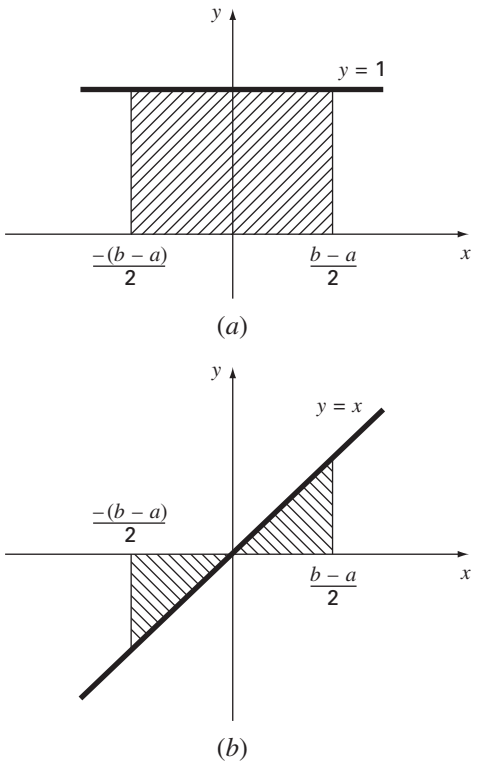


FIGURA 22.6
Duas integrais que devem ser calculadas exatamente pela regra do trapézio: (*a*) uma constante e (*b*) uma reta.

22.3.2 Dedução da Fórmula de Gauss-Legendre de Dois Pontos

Do mesmo modo como no caso da dedução da regra dos trapézios, o objetivo da quadratura de Gauss é determinar os coeficientes de uma equação da forma

$$I \cong c_0 f(x_0) + c_1 f(x_1) \tag{22.12}$$

em que os c's são coeficientes desconhecidos. Entretanto, em contraste com a regra do trapézio que usa extremidades fixas a e b, os argumentos da função x_0 e x_1 não são fixados nas extremidades, mas são incógnitas (Figura 22.7). Assim, agora temos um total de quatro incógnitas que devem ser calculadas e necessitamos de quatro condições para determiná-las exatamente.

Do mesmo modo que no caso da regra dos trapézios, podemos obter duas dessas condições, supondo que a Equação (22.12) calcula a integral de uma função constante e de uma função linear exatamente. Então, para chegar a duas outras condições, simplesmente estendemos esse raciocínio supondo que ele também calculará a integral de uma função parabólica ($y = x^2$) e de uma cúbica ($y = x^3$) exatamente. Fazendo isso, determinamos todas as quatro incógnitas e, de quebra, deduzimos uma fórmula de integração linear de dois pontos que é exata para funções cúbicas. As quatro equações a serem resolvidas são

$$c_0 f(x_0) + c_1 f(x_1) = \int_{-1}^{1} 1\, dx = 2 \tag{22.13}$$

$$c_0 f(x_0) + c_1 f(x_1) = \int_{-1}^{1} x\, dx = 0 \tag{22.14}$$

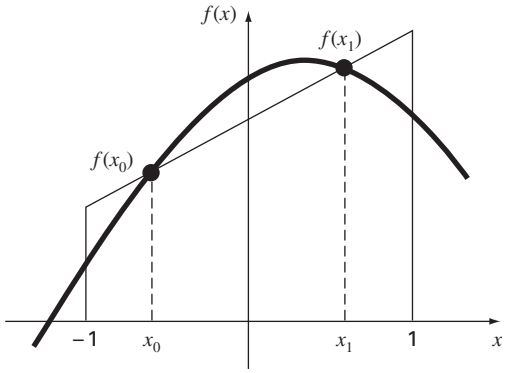


FIGURA 22.7
Descrição gráfica das variáveis desconhecidas x_0 e x_1 da integração por quadratura de Gauss.

$$c_0 f(x_0) + c_1 f(x_1) = \int_{-1}^{1} x^2\,dx = \frac{2}{3} \tag{22.15}$$

$$c_0 f(x_0) + c_1 f(x_1) = \int_{-1}^{1} x^3\,dx = 0 \tag{22.16}$$

As Equações (22.13) até (22.16) podem ser resolvidas simultaneamente por

$$c_0 = c_1 = 1$$

$$x_0 = -\frac{1}{\sqrt{3}} = -0{,}5773503\ldots$$

$$x_1 = \frac{1}{\sqrt{3}} = 0{,}5773503\ldots$$

o que pode ser substituído na Equação (22.12) para fornecer a fórmula de Gauss-Legendre de dois pontos

$$I \cong f\left(\frac{-1}{\sqrt{3}}\right) + f\left(\frac{1}{\sqrt{3}}\right) \tag{22.17}$$

Logo, chegamos ao interessante resultado que uma simples adição de valores da função em $x = 1/\sqrt{3}$ e $-1/\sqrt{3}$ fornece uma estimativa da integral que tem acurácia de terceira ordem.

Observe que os extremos de integração nas Equações (22.13) até (22.16) são de -1 a 1. Isso foi feito para simplificar a matemática e tornar a formulação tão geral quanto possível. Uma simples mudança de variável pode ser usada para transformar outros extremos de integração para essa fórmula, o que se faz supondo que uma nova variável x_d está relacionada à variável original x de uma forma linear como em

$$x = a_0 + a_1 x_d \tag{22.18}$$

Se o extremo inferior, $x = a$, corresponder a $x_d = -1$, esses valores podem ser substituídos na Equação (22.18) para fornecer

$$a = a_0 + a_1(-1) \tag{22.19}$$

Analogamente, o extremo superior, $x = b$, corresponde a $x_d = 1$, e fornece

$$b = a_0 + a_1(1) \tag{22.20}$$

As Equações (22.19) e (22.20) podem ser resolvidas simultaneamente por

$$a_0 = \frac{b+a}{2} \tag{22.21}$$

e

$$a_1 = \frac{b-a}{2} \tag{22.22}$$

o que pode ser substituído na Equação (22.18) para fornecer

$$x = \frac{(b+a)+(b-a)x_d}{2} \tag{22.23}$$

Essa equação pode ser derivada para dar

$$dx = \frac{b-a}{2}\,dx_d \tag{22.24}$$

Os valores de x e dx, nas Equações (22.23) e (22.24), respectivamente, podem ser substituídos na equação a ser integrada. Essas substituições transformam efetivamente o intervalo de integração sem mudar o valor da integral. O exemplo a seguir ilustra como isso é feito na prática.

EXEMPLO 22.3 Fórmula de Gauss-Legendre de Dois Pontos

Enunciado do Problema. Use a Equação (22.17) para calcular a integral de

$$f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

entre os extremos $x = 0$ a 0,8. Lembre-se de que esse é o mesmo problema que resolvemos no Capítulo 21 usando diversas formulações de Newton-Cotes. O valor exato da integral é 1,640533.

Solução. Antes de integrar a função, devemos fazer uma mudança de variáveis de forma que os extremos sejam de -1 a $+1$. Para isso, substituímos $a = 0$ e $b = 0{,}8$ na Equação (22.23) para obter

$$x = 0{,}4 + 0{,}4x_d$$

A derivada dessa relação é [Equação (22.24)]

$$dx = 0{,}4\,dx_d$$

Ambas as equações podem ser substituídas na equação original para fornecer

$$\begin{aligned}\int_0^{0{,}8} &(0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5)\,dx \\ &= \int_{-1}^{1} \big[0{,}2 + 25(0{,}4 + 0{,}4x_d) - 200(0{,}4 + 0{,}4x_d)^2 + 675(0{,}4 + 0{,}4x_d)^3 \\ &\quad - 900(0{,}4 + 0{,}4x_d)^4 + 400(0{,}4 + 0{,}4x_d)^5\big]0{,}4\,dx_d\end{aligned}$$

Portanto, o lado direito está na forma adequada para o cálculo usando a quadratura de Gauss. A função transformada pode ser calculada em $-1/\sqrt{3}$ como sendo igual a 0,516741 e em $1/\sqrt{3}$ como sendo igual a 1,305837. Portanto, de acordo com a Equação (22.17), a integral é

$$I \cong 0{,}516741 + 1{,}305837 = 1{,}822578$$

o que representa um erro relativo porcentual de $-11{,}1\%$. Esse resultado é comparável em módulo à aplicação da regra do trapézio com quatro segmentos (Tabela 21.1) ou a uma única aplicação das regras 1/3 e 3/8 de Simpson (Exemplos 21.4 e 21.6). O último resultado deve ser esperado, já que as regras de Simpson tem acurácia de terceira ordem. Contudo, por causa da escolha inteligente dos pontos base, a quadratura de Gauss alcança essa acurácia com base em apenas dois cálculos da função.

22.3.3 Fórmulas com Mais Pontos

Além da fórmula de dois pontos descrita na seção anterior, versões com mais pontos podem ser desenvolvidas na forma geral

$$I \cong c_0 f(x_0) + c_1 f(x_1) + \cdots + c_{n-1} f(x_{n-1}) \tag{22.25}$$

em que n é o número de pontos. Os valores para os c's e os x's para até (e incluindo) a fórmula de seis pontos estão resumidos na Tabela 22.1.

TABELA 22.1 Fatores de peso, c; e argumentos da função, x; nas fórmulas de Gauss-Legendre.

Pontos	Fatores de Peso	Argumentos da Função	Erro de Truncamento
2	$c_0 = 1{,}0000000$	$x_0 = -0{,}577350269$	$\cong f^{(4)}(\xi)$
	$c_1 = 1{,}0000000$	$x_1 = \ \ 0{,}577350269$	
3	$c_0 = 0{,}5555556$	$x_0 = -0{,}774596669$	$\cong f^{(6)}(\xi)$
	$c_1 = 0{,}8888889$	$x_1 = \ \ 0{,}0$	
	$c_2 = 0{,}5555556$	$x_2 = \ \ 0{,}774596669$	
4	$c_0 = 0{,}3478548$	$x_0 = -0{,}861136312$	$\cong f^{(8)}(\xi)$
	$c_1 = 0{,}6521452$	$x_1 = -0{,}339981044$	
	$c_2 = 0{,}6521452$	$x_2 = \ \ 0{,}339981044$	
	$c_3 = 0{,}3478548$	$x_3 = \ \ 0{,}861136312$	
5	$c_0 = 0{,}2369269$	$x_0 = -0{,}906179846$	$\cong f^{(10)}(\xi)$
	$c_1 = 0{,}4786287$	$x_1 = -0{,}538469310$	
	$c_2 = 0{,}5688889$	$x_2 = \ \ 0{,}0$	
	$c_3 = 0{,}4786287$	$x_3 = \ \ 0{,}538469310$	
	$c_4 = 0{,}2369269$	$x_4 = \ \ 0{,}906179846$	
6	$c_0 = 0{,}1713245$	$x_0 = -0{,}932469514$	$\cong f^{(12)}(\xi)$
	$c_1 = 0{,}3607616$	$x_1 = -0{,}661209386$	
	$c_2 = 0{,}4679139$	$x_2 = -0{,}238619186$	
	$c_3 = 0{,}4679139$	$x_3 = \ \ 0{,}238619186$	
	$c_4 = 0{,}3607616$	$x_4 = \ \ 0{,}661209386$	
	$c_5 = 0{,}1713245$	$x_5 = \ \ 0{,}932469514$	

EXEMPLO 22.4 Fórmula de Gauss-Legendre de Três Pontos

Enunciado do Problema. Use a fórmula de três pontos da Tabela 22.1 para fazer uma estimativa da integral da mesma função do Exemplo 22.3.

Solução. De acordo com a Tabela 22.1, a fórmula de três pontos é

$$I = 0{,}5555556\, f(-0{,}7745967) + 0{,}8888889\, f(0) + 0{,}5555556\, f(0{,}7745967)$$

o que é igual a

$$I = 0{,}2813013 + 0{,}8732444 + 0{,}4859876 = 1{,}640533$$

o que é exato.

Como a quadratura de Gauss exige cálculos da função em pontos que não são uniformemente espaçados dentro do intervalo de integração, não é apropriada para os casos nos quais a função não é conhecida analiticamente. Logo, não é adequada para problemas de engenharia que lidem com dados tabulados. Entretanto, quando a função for conhecida, sua eficiência pode ser uma vantagem decisiva. Isso é particularmente verdade quando muitos cálculos de integrais devam ser feitos.

EXEMPLO 22.5 Aplicando a Quadratura de Gauss ao Problema do Pára-quedista em Queda Livre

Enunciado do Problema. No Exemplo 21.3, usamos a aplicação múltipla da regra do trapézio para calcular

$$d = \frac{gm}{c}\int_0^{10}\left[1 - e^{-(c/m)t}\right]\,dt$$

em que $g = 9{,}8$, $c = 12{,}5$ e $m = 68{,}1$. O valor exato da integral foi determinado pelo cálculo como sendo 289,4351. Lembre-se de que a melhor estimativa obtida usando a regra do trapézio com 500 segmentos foi 289,4348 com um $|\varepsilon_t| \cong 1{,}15 \times 10^{-4}$ por cento. Repita esse cálculo usando a quadratura de Gauss.

Solução. Depois de modificar a função, são obtidos os seguintes resultados:

Estimativa de dois pontos = 290,0145
Estimativa de três pontos = 289,4393
Estimativa de quatro pontos = 289,4352
Estimativa de cinco pontos = 289,4351
Estimativa de seis pontos = 289,4351

Logo, as estimativas de cinco e de seis pontos fornecem resultados que são exatos até sete algarismos significativos.

22.3.4 Análise de Erro para a Quadratura de Gauss

O erro para as fórmulas de Gauss-Legendre é especificado em geral por (Carnahan et al., 1969)

$$E_t = \frac{2^{2n+3}[(n+1)!]^4}{(2n+3)[(2n+2)!]^3} f^{(2n+2)}(\xi) \tag{22.26}$$

em que n é o número de pontos menos um e $f^{(2n+2)}(\xi)$ é a $(2n+2)$-ésima derivada da função depois da mudança de variáveis com ξ localizado em algum ponto no intervalo de -1 a 1. A comparação da Equação (22.26) com a Tabela 21.2 indica a superioridade da quadratura de Gauss sobre as fórmulas de Newton-Cotes, desde que as derivadas de ordem mais alta não cresçam substancialmente com o aumento de n. O Problema 22.8 no final deste capítulo ilustra um caso no qual as fórmulas de Gauss-Legendre têm um desempenho ruim. Nessas situações, a aplicação múltipla da regra de Simpson ou a integração de Romberg seriam preferíveis. Entretanto, para muitas funções confrontadas na prática da engenharia, a quadratura de Gauss fornece uma maneira eficiente de calcular integrais.

22.4 INTEGRAIS IMPRÓPRIAS

Até este ponto, tratamos exclusivamente de integrais que tenham extremos finitos e integrandos limitados. Embora esses tipos sejam comuns na engenharia, algumas vezes integrais impróprias devem ser calculadas. Nesta seção nos concentraremos em um tipo de integral imprópria — ou seja, a que tem um extremo inferior $-\infty$ ou um extremo superior $+\infty$.

Tais integrais usualmente podem ser calculadas fazendo uma mudança de variáveis que transforma o intervalo infinito em um que é finito. A seguinte identidade serve a esse propósito e funciona para qualquer função que decresça a zero pelo menos tão rapidamente quanto $1/x^2$ quando x tende a infinito:

$$\int_a^b f(x)\,dx = \int_{1/b}^{1/a} \frac{1}{t^2} f\left(\frac{1}{t}\right) dt \tag{22.27}$$

para $ab > 0$. Logo, pode ser usada apenas quando a for positivo e b for ∞ ou quando a for $-\infty$ e b for negativo. Nos casos em que os extremos forem de $-\infty$ a um valor positivo ou de um valor negativo a ∞, a integral pode ser implementada em duas etapas. Por exemplo.

$$\int_{-\infty}^b f(x)\,dx = \int_{-\infty}^{-A} f(x)\,dx + \int_{-A}^b f(x)\,dx \tag{22.28}$$

em que $-A$ é escolhido como um valor negativo suficientemente grande para que a função tenha começado a se aproximar assintoticamente de zero pelo menos tão rapidamente quanto $1/x^2$. Depois de a integral ter sido dividida em duas partes, a primeira pode ser calculada pela Equação (22.27) e a segunda com uma fórmula fechada de Newton-Cotes, tal como a regra 1/3 de Simpson.

Um problema com o uso da Equação (22.27) para o cálculo da integral é que a função transformada será singular em um dos extremos. As fórmulas de integração abertas podem ser usadas para contornar esse dilema, já que permitem o cálculo da integral sem usar dados nas extremidades do intervalo de integração. Para permitir uma flexibilidade máxima, é necessária uma versão com aplicações múltiplas de uma das fórmulas da Tabela 21.4.

Versões com aplicações múltiplas de fórmulas abertas podem ser obtidas usando-se fórmulas fechadas para os segmentos interiores e fórmulas abertas para os das extremidades. Por exemplo, a regra do trapézio com segmentos múltiplos e a regra do ponto médio podem ser combinadas para dar

$$\int_{x_0}^{x_n} f(x)\,dx = h\left[\frac{3}{2}f(x_1) + \sum_{i=2}^{n-2} f(x_i) + \frac{3}{2}f(x_{n-1})\right]$$

Além disso, fórmulas semi-abertas podem ser desenvolvidas para os casos nos quais uma ou outra extremidade do intervalo for fechada. Por exemplo, uma fórmula que é aberta na extremidade inferior e fechada na superior é dada por

$$\int_{x_0}^{x_n} f(x)\,dx = h\left[\frac{3}{2}f(x_1) + \sum_{i=2}^{n-1} f(x_i) + \frac{1}{2}f(x_n)\right]$$

Embora essas relações possam ser usadas, uma fórmula preferível é (Press et al., 1992)

$$\int_{x_0}^{x_n} f(x)\,dx = h[f(x_{1/2}) + f(x_{3/2}) + \cdots + f(x_{n-3/2}) + f(x_{n-1/2})] \tag{22.29}$$

que é chamada de *regra do ponto médio estendida*. Observe que essa fórmula é baseada em extremos de integração que estão $h/2$ depois e antes do primeiro e último pontos dados (Figura 22.8).

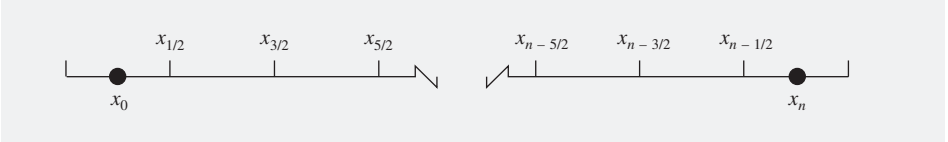


FIGURA 22.8
Disposição dos pontos dados em relação aos extremos de integração na regra do ponto médio estendida.

EXEMPLO 22.6 Cálculo de uma Integral Imprópria

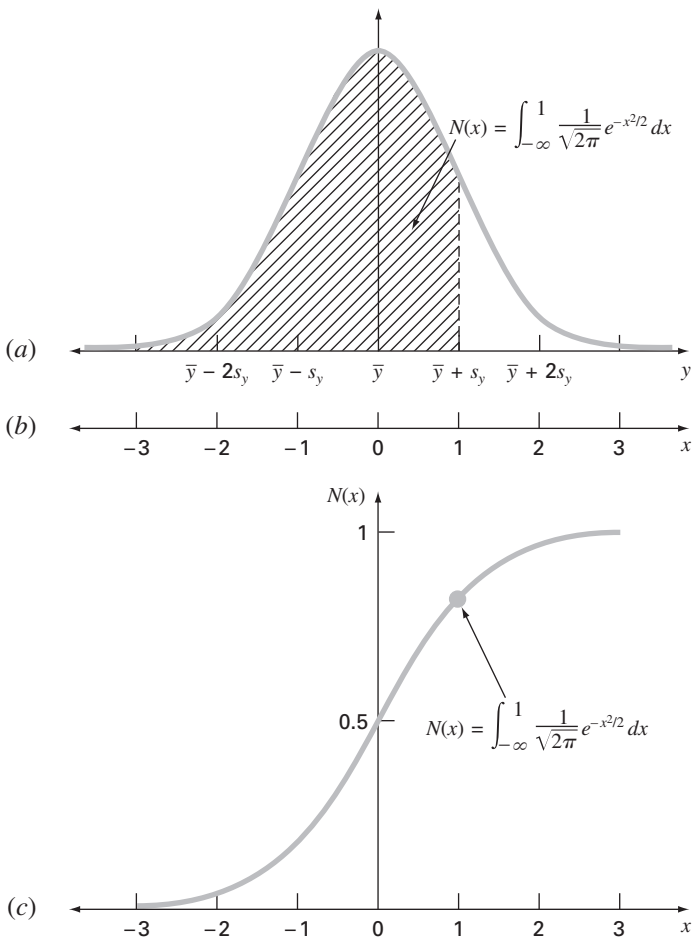
Enunciado do Problema. A *distribuição normal cumulativa* é uma fórmula importante na estatística (veja a Figura 22.9):

$$N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \tag{E22.6.1}$$

em que $x = (y - \bar{y})/s_y$ é chamado de *desvio normal padrão*. Ele representa uma mudança de variáveis para mudar a escala da distribuição normal de modo que ela esteja centrada no zero e que a distância ao longo da abscissa seja medida em múltiplos do desvio padrão (Figura 22.9*b*).

A Equação (E22.6.1) representa a probabilidade de que um evento seja menor do que x. Por exemplo, se $x = 1$, a Equação (E22.6.1) pode ser usada para determinar que a probabilidade de ocorrer um evento que seja menor do que um desvio padrão acima da média é $N(1) = 0{,}8413$. Em outras palavras, se ocorrerem 100 eventos, aproximadamente 84 serão menores do que a média mais um desvio padrão. Como a Equação (E22.6.1) não pode ser expressa em uma forma funcional simples, ela é resolvida numericamente e listada em tabelas de estatística. Use a Equação (22.28) em conjunto com a regra 1/3 de Simpson e a regra do ponto médio estendida para determinar $N(1)$ numericamente.

FIGURA 22.9
(*a*) A distribuição normal, (*b*) a abscissa transformada em termos do desvio normal padrão e (*c*) a distribuição normal cumulativa. A área hachurada em (*a*) e o ponto em (*c*) representam a probabilidade de que um evento aleatório seja menor do que a média mais um desvio padrão.



Solução. A Equação (E22.6.1) pode ser reescrita em termos da Equação (22.28) como

$$N(x) = \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{-2} e^{-x^2/2}\,dx + \int_{-2}^{1} e^{-x^2/2}\,dx\right)$$

A primeira integral pode ser calculada pela aplicação da Equação (22.27), fornecendo

$$\int_{-\infty}^{-2} e^{-x^2/2}\,dx = \int_{-1/2}^{0} \frac{1}{t^2} e^{-1/(2t^2)}\,dt$$

Então, a regra do ponto médio estendida com $h = 1/8$ pode ser usada para obter uma estimativa de

$$\begin{aligned}\int_{-1/2}^{0} \frac{1}{t^2} e^{-1/(2t^2)}\,dt &\cong \frac{1}{8}[f(x_{-7/16}) + f(x_{-5/16}) + f(x_{-3/16}) + f(x_{-1/16})] \\ &= \frac{1}{8}[0{,}3833 + 0{,}0612 + 0 + 0] = 0{,}0556\end{aligned}$$

A regra 1/3 de Simpson com $h = 0{,}5$ pode ser usada para obter uma estimativa da segunda integral

$$\begin{aligned}&\int_{-2}^{1} e^{-x^2/2}\,dx \\ &= [1 - (-2)]\frac{0{,}1353 + 4(0{,}3247 + 0{,}8825 + 0{,}8825) + 2(0{,}6065 + 1) + 0{,}6065}{3(6)} \\ &= 2{,}0523\end{aligned}$$

Portanto, o resultado final pode ser calculado como sendo

$$N(1) \cong \frac{1}{\sqrt{2\pi}}(0{,}0556 + 2{,}0523) = 0{,}8409$$

o que representa um erro de $\varepsilon_t = 0{,}046\%$.

Os cálculos anteriores podem ser melhorados de diversas formas. Primeiro, fórmulas de ordem superior poderiam ser usadas — por exemplo, empregar uma integração de Romberg. Segundo, mais pontos poderiam ser usados. Press et al. (1986) exploraram ambas as opções em profundidade.

Além de extremos infinitos, há muitas outras formas pelas quais uma integral pode ser imprópria. Exemplos comuns incluem casos nos quais a integral é singular em ambos os extremos de integração ou em um ponto no interior do intervalo de integração. Press et al. (1986) fornecem uma discussão interessante de maneiras de lidar com tais situações.

PROBLEMAS

22.1 Use a integração de Romberg para calcular

$$I = \int_1^2 \left(2x + \frac{3}{x}\right)^2 dx$$

com uma acurácia de $\varepsilon_s = 0{,}5\%$. Seus resultados devem ser apresentados na forma da Figura 22.3. Use a solução analítica da integral para determinar o erro relativo porcentual do resultado obtido com a integração de Romberg. Verifique que ε_t é menor do que o critério de parada ε_s.

22.2 Use a integração de Romberg da ordem de h^8 para calcular

$$\int_0^3 xe^x\,dx$$

Compare ε_a e ε_t.

22.3 Use a integração de Romberg para calcular

$$\int_0^2 \frac{e^x \operatorname{sen} x}{1 + x^2}\,dx$$

com uma acurácia de $\varepsilon_s = 0{,}5\%$. Seus resultados devem ser apresentados na forma da Figura 22.3.

22.4 Obtenha uma estimativa da integral do Problema 22.1 usando as fórmulas de Gauss-Legendre com dois, três e quatro pontos. Calcule ε_t para cada caso, com base na solução analítica.

22.5 Obtenha uma estimativa da integral do Problema 22.2 usando as fórmulas de Gauss-Legendre com dois, três e quatro pontos. Calcule ε_t para cada caso, com base na solução analítica.

22.6 Obtenha uma estimativa da integral do Problema 22.3 usando a fórmula de Gauss-Legendre com cinco pontos.

22.7 Faça os cálculos dos Exemplos 21.3 e 22.5 para o pára-quedista em queda livre, mas use a integração de Romberg ($\varepsilon_s = 0{,}05\%$).

22.8 Empregue as fórmulas de Gauss-Legendre com dois até seis pontos para resolver

$$\int_{-3}^{3} \frac{1}{1+x^2}\,dx$$

Interprete seus resultados à luz da Equação (22.26).

22.9 Use integração numérica para calcular o seguinte:

(a) $\displaystyle\int_2^\infty \frac{dx}{x(x+2)}$ **(b)** $\displaystyle\int_0^\infty e^{-y}\operatorname{sen}^2 y\,dy$

(c) $\displaystyle\int_0^\infty \frac{1}{(1+y^2)(1-y^2/2)}\,dy$ **(d)** $\displaystyle\int_{-2}^\infty ye^{-y}\,dy$

(e) $\displaystyle\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-x^2}dx$

Observe que (**e**) é a distribuição normal (lembre-se da Figura 22.9).

22.10 Desenvolva um programa de computador amigável ao usuário para a regra (**a**) do trapézio e (**b**) 1/3 de Simpson com segmentos múltiplos baseado na Figura 22.1. Teste-o integrando

$$\int_0^1 x^{0,1}(1{,}2-x)\left(1-e^{20(x-1)}\right)dx$$

Use o valor verdadeiro 0,602298 para calcular ε_t para $n = 4$.

22.11 Desenvolva um programa de computador amigável ao usuário para a integração de Romberg baseado na Figura 22.4. Teste-o reproduzindo os resultados dos Exemplos 22.3 e 22.4 e fazendo a integral no Problema 22.10.

22.12 Desenvolva um programa de computador amigável ao usuário para a quadratura de Gauss. Teste-o reproduzindo os resultados dos Exemplos 22.3 e 22.4 e fazendo a integral no Problema 22.10.

22.13 Não existe uma expressão em forma fechada para a função

$$\operatorname{erf}(a) = \frac{2}{\sqrt{\pi}}\int_0^a e^{-x^2}dx$$

Use a abordagem da quadratura de Gauss de dois pontos para obter uma estimativa de erf (1,5). Observe que o valor exato é 0,966105.

22.14 A quantidade de massa que pode ser transportada por um tubo durante um período de tempo pode ser calculada por

$$M = \int_{t_1}^{t_2} Q(t)c(t)dt$$

em que M é a massa (mg), t_1 é o instante inicial (min), t_2 é o instante final (min), $Q(t)$ é a vazão (m^3/min) e $c(t)$ é a concentração (mg/m^3). As seguintes representações funcionais definem a variação temporal da vazão e da concentração:

$$Q(t) = 9 + 4\cos^2(0{,}4t)$$
$$c(t) = 5e^{-0{,}5t} + 2e^{0{,}15t}$$

Determine a massa transportada entre $t_1 = 2$ e $t_2 = 8$ min com a integração de Romberg e uma tolerância de 0,1%.

22.15 As profundidades de um rio, H, foram medidas em distâncias igualmente espaçadas através de um canal como tabulado a seguir. A área da seção transversal do rio pode ser determinada por integração como em

$$A_c = \int_0^x H(x)\,dx$$

Use a integração de Romberg para efetuar a integral com um critério de parada de 1%.

x, m	0	2	4	6	8	10	12	14	16
H, m	0	1,9	2	2	2,4	2.6	2,25	1,12	0

CAPÍTULO 23

Derivação Numérica

Já introduzimos a noção de derivação numérica no Capítulo 4. Lembre-se de que usamos a expansão em série de Taylor para deduzir as aproximações por diferenças divididas finitas de derivadas. No Capítulo 4, desenvolvemos aproximações por diferenças progressiva, regressiva e centrada para a primeira derivada e para derivadas de ordem superior. Lembre-se de que, no melhor caso, essas estimativas tinham erros que eram $O(h^2)$ — isto é, seus erros eram proporcionais ao quadrado do tamanho do passo. Esse nível de precisão decorre do número de termos da série de Taylor que foram mantidos durante a dedução dessas fórmulas. Vamos agora ilustrar como desenvolver fórmulas mais acuradas mantendo mais termos.

23.1 FÓRMULAS DE DERIVAÇÃO DE ALTA ACURÁCIA

Como observado, fórmulas de diferenças divididas de alta acurácia podem ser geradas incluindo termos adicionais na expansão em série de Taylor. Por exemplo, a expansão em série de Taylor progressiva pode ser escrita como [Equação (4.21)]

$$f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{f''(x_i)}{2}h^2 + \cdots \tag{23.1}$$

o que pode ser resolvido por

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} - \frac{f''(x_i)}{2}h + O(h^2) \tag{23.2}$$

No Capítulo 4, truncamos esse resultado excluindo os termos da segunda derivada e de derivadas de ordem superior e ficamos com o resultado final

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} + O(h) \tag{23.3}$$

Em contraste a essa abordagem, agora manteremos o termo da segunda derivada substituindo a seguinte aproximação da segunda derivada [lembre-se da Equação (4.24)]

$$f''(x_i) = \frac{f(x_{i+2}) - 2f(x_{i+1}) + f(x_i)}{h^2} + O(h) \tag{23.4}$$

na Equação (23.2) para obter

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} - \frac{f(x_{i+2}) - 2f(x_{i+1}) + f(x_i)}{2h^2}h + O(h^2)$$

ou, agrupando os termos,

$$f'(x_i) = \frac{-f(x_{i+2}) + 4f(x_{i+1}) - 3f(x_i)}{2h} + O(h^2) \tag{23.5}$$

Observe que a inclusão do termo da segunda derivada melhorou a precisão para $O(h^2)$. Versões melhoradas análogas podem ser desenvolvidas para as fórmulas regressiva e centrada, bem como aproximações para as derivadas de ordem superior. As fórmulas estão resumidas nas Figuras 23.1 a 23.3 junto com os resultados do Capítulo 4. O seguinte exemplo ilustra a utilidade dessas fórmulas na estimativa de derivadas.

Primeira Derivada	Erro
$f'(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h}$	$O(h)$
$f'(x_i) = \frac{-f(x_{i+2}) + 4f(x_{i+1}) - 3f(x_i)}{2h}$	$O(h^2)$
Segunda Derivada	
$f''(x_i) = \frac{f(x_{i+2}) - 2f(x_{i+1}) + f(x_i)}{h^2}$	$O(h)$
$f''(x_i) = \frac{-f(x_{i+3}) + 4f(x_{i+2}) - 5f(x_{i+1}) + 2f(x_i)}{h^2}$	$O(h^2)$
Terceira Derivada	
$f'''(x_i) = \frac{f(x_{i+3}) - 3f(x_{i+2}) + 3f(x_{i+1}) - f(x_i)}{h^3}$	$O(h)$
$f'''(x_i) = \frac{-3f(x_{i+4}) + 14f(x_{i+3}) - 24f(x_{i+2}) + 18f(x_{i+1}) - 5f(x_i)}{2h^3}$	$O(h^2)$
Quarta derivada	
$f''''(x_i) = \frac{f(x_{i+4}) - 4f(x_{i+3}) + 6f(x_{i+2}) - 4f(x_{i+1}) + f(x_i)}{h^4}$	$O(h)$
$f''''(x_i) = \frac{-2f(x_{i+5}) + 11f(x_{i+4}) - 24f(x_{i+3}) + 26f(x_{i+2}) - 14f(x_{i+1}) + 3f(x_i)}{h^4}$	$O(h^2)$

FIGURA 23.1
Fórmulas de diferença dividida finita progressiva: são apresentadas duas versões para cada derivada. A segunda versão incorpora mais termos da expansão em série de Taylor e é, conseqüentemente, mais acurada.

FIGURA 23.2
Fórmulas de diferença dividida finita regressiva: são apresentadas duas versões para cada derivada. A segunda versão incorpora mais termos da expansão em série de Taylor e é, conseqüentemente, mais acurada.

Primeira Derivada	Erro
$f'(x_i) = \frac{f(x_i) - f(x_{i-1})}{h}$	$O(h)$
$f'(x_i) = \frac{3f(x_i) - 4f(x_{i-1}) + f(x_{i-2})}{2h}$	$O(h^2)$
Segunda Derivada	
$f''(x_i) = \frac{f(x_i) - 2f(x_{i-1}) + f(x_{i-2})}{h^2}$	$O(h)$
$f''(x_i) = \frac{2f(x_i) - 5f(x_{i-1}) + 4f(x_{i-2}) - f(x_{i-3})}{h^2}$	$O(h^2)$
Terceira Derivada	
$f'''(x_i) = \frac{f(x_i) - 3f(x_{i-1}) + 3f(x_{i-2}) - f(x_{i-3})}{h^3}$	$O(h)$
$f'''(x_i) = \frac{5f(x_i) - 18f(x_{i-1}) + 24f(x_{i-2}) - 14f(x_{i-3}) + 3f(x_{i-4})}{2h^3}$	$O(h^2)$
Quarta Derivada	
$f''''(x_i) = \frac{f(x_i) - 4f(x_{i-1}) + 6f(x_{i-2}) - 4f(x_{i-3}) + f(x_{i-4})}{h^4}$	$O(h)$
$f''''(x_i) = \frac{3f(x_i) - 14f(x_{i-1}) + 26f(x_{i-2}) - 24f(x_{i-3}) + 11f(x_{i-4}) - 2f(x_{i-5})}{h^4}$	$O(h^2)$

Primeira Derivada — Erro

$$f'(x_i) = \frac{f(x_{i+1}) - f(x_{i-1})}{2h} \qquad O(h^2)$$

$$f'(x_i) = \frac{-f(x_{i+2}) + 8f(x_{i+1}) - 8f(x_{i-1}) + f(x_{i-2})}{12h} \qquad O(h^4)$$

Segunda Derivada

$$f''(x_i) = \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1})}{h^2} \qquad O(h^2)$$

$$f''(x_i) = \frac{-f(x_{i+2}) + 16f(x_{i+1}) - 30f(x_i) + 16f(x_{i-1}) - f(x_{i-2})}{12h^2} \qquad O(h^4)$$

Terceira Derivada

$$f'''(x_i) = \frac{f(x_{i+2}) - 2f(x_{i+1}) + 2f(x_{i-1}) - f(x_{i-2})}{2h^3} \qquad O(h^2)$$

$$f'''(x_i) = \frac{-f(x_{i+3}) + 8f(x_{i+2}) - 13f(x_{i+1}) + 13f(x_{i-1}) - 8f(x_{i-2}) + f(x_{i-3})}{8h^3} \qquad O(h^4)$$

Quarta Derivada

$$f''''(x_i) = \frac{f(x_{i+2}) - 4f(x_{i+1}) + 6f(x_i) - 4f(x_{i-1}) + f(x_{i-2})}{h^4} \qquad O(h^2)$$

$$f''''(x_i) = \frac{-f(x_{i+3}) + 12f(x_{i+2}) + 39f(x_{i+1}) + 56f(x_i) - 39f(x_{i-1}) + 12f(x_{i-2}) + f(x_{i-3})}{6h^4} \qquad O(h^4)$$

FIGURA 23.3
Fórmulas de diferença dividida finita centrada: são apresentadas duas versões para cada derivada. A segunda versão incorpora mais termos da expansão em série de Taylor e é, conseqüentemente, mais acurada.

EXEMPLO 23.1 Fórmulas de Derivação de Alta Precisão

Enunciado do Problema. Lembre-se de que, no Exemplo 4.4, obtivemos estimativas da derivada de

$$f(x) = -0{,}1x^4 - 0{,}15x^3 - 0{,}5x^2 - 0{,}25x + 1{,}2$$

em $x = 0{,}5$ usando diferença dividida finita e tamanho de passo de $h = 0{,}25$,

	Progressiva $O(h)$	**Regressiva $O(h)$**	**Centrada $O(h^2)$**
Estimativa	−1,155	−0,714	−0,934
ε_t (%)	−26,5	21,7	−2,4

em que os erros foram calculados com base no valor verdadeiro de −0,9125. Repita esses cálculos, mas use as fórmulas de alta precisão das Figuras 23.1 a 23.3.

Solução. Os dados necessários para esse exemplo são

$$x_{i-2} = 0 \qquad f(x_{i-2}) = 1{,}2$$
$$x_{i-1} = 0{,}25 \qquad f(x_{i-1}) = 1{,}103516$$
$$x_i = 0{,}5 \qquad f(x_i) = 0{,}925$$
$$x_{i+1} = 0{,}75 \qquad f(x_{i+1}) = 0{,}6363281$$
$$x_{i+2} = 1 \qquad f(x_{i+2}) = 0{,}2$$

A diferença progressiva de acurácia $O(h^2)$ é calculada por (Figura 23.1)

$$f'(0{,}5) = \frac{-0{,}2 + 4(0{,}6363281) - 3(0{,}925)}{2(0{,}25)} = -0{,}859375 \qquad \varepsilon_t = 5{,}82\%$$

A diferença regressiva de acurácia $O(h^2)$ é calculada por (Figura 23.2)

$$f'(0{,}5) = \frac{3(0{,}925) - 4(1{,}035156) + 1{,}2}{2(0{,}25)} = -0{,}878125 \qquad \varepsilon_t = 3{,}77\%$$

A diferença centrada de acurácia $O(h^4)$ é calculada por (Figura 23.3)

$$f'(0,5) = \frac{-0,2 + 8(0,6363281) - 8(1,035156) + 1,2}{12(0,25)} = -0,9125 \qquad \varepsilon_t = 0\%$$

Como esperado, os erros para as diferenças progressiva e regressiva são consideravelmente mais acurados do que os resultados do Exemplo 4.4. Entretanto, surpreendentemente, a diferença centrada fornece um resultado perfeito. Isso ocorre porque as fórmulas baseadas na série de Taylor são equivalentes a passar polinômios pelos pontos dados.

23.2 EXTRAPOLAÇÃO DE RICHARDSON

Até este ponto, vimos que existem duas maneiras de melhorar as estimativas da derivada quando usando diferença dividida finita: (1) diminuir o tamanho do passo ou (2) usar uma fórmula de ordem mais alta que utiliza mais pontos. Uma terceira abordagem, baseada na extrapolação de Richardson, usa duas estimativas da derivada para calcular uma terceira aproximação, mais acurada.

Lembre-se, da Seção 22.2.1, que a extrapolação de Richardson fornece uma maneira de obter uma estimativa melhorada da integral I pela fórmula [Equação (22.4)]

$$I \cong I(h_2) + \frac{1}{(h_1/h_2)^2 - 1}[I(h_2) - I(h_1)] \tag{23.6}$$

em que $I(h_1)$ e $I(h_2)$ são estimativas da integral usando dois tamanhos de passos, h_1 e h_2. Por causa de sua conveniência quando expressa como um algoritmo computacional, essa fórmula geralmente é escrita para o caso em que $h_2 = h_1/2$, como

$$I \cong \frac{4}{3} I(h_2) - \frac{1}{3} I(h_1) \tag{23.7}$$

De modo análogo, a Equação (23.7) pode ser escrita para derivadas como

$$D \cong \frac{4}{3} D(h_2) - \frac{1}{3} D(h_1) \tag{23.8}$$

Para aproximações por diferenças centradas com $O(h^2)$, a aplicação dessa fórmula produzirá uma nova estimativa da derivada de $O(h^4)$.

EXEMPLO 23.2 Extrapolação de Richardson

Enunciado do Problema. Usando a mesma função do Exemplo 23.1, obtenha uma estimativa da primeira derivada em $x = 0,5$ usando tamanhos de passo $h_1 = 0,5$ e $h_2 = 0,25$. Então, utilize a Equação (23.8) para calcular uma estimativa melhorada com a extrapolação de Richardson. Lembre-se de que o valor verdadeiro é $-0,9125$.

Solução. A estimativa da primeira derivada pode ser calculada com diferença centrada por

$$D(0,5) = \frac{0,2 - 1,2}{1} = -1,0 \qquad \varepsilon_t = -9,6\%$$

e

$$D(0,25) = \frac{0,6363281 - 1,103516}{0,5} = -0,934375 \qquad \varepsilon_t = -2,4\%$$

A estimativa melhorada pode ser determinada aplicando a Equação (23.8) para obter

$$D = \frac{4}{3}(-0,934375) - \frac{1}{3}(-1) = -0,9125$$

o que, no caso presente, é um resultado exato.

O exemplo anterior fornece um resultado perfeito porque a função sendo analisada era um polinômio de quarto grau. O resultado perfeito decorreu do fato de que a extrapolação de Richardson é, na realidade, equivalente a ajustar um polinômio de grau mais alto aos dados e então calcular a derivada por diferenças divididas centradas. Assim, no caso presente, ajustou a derivada do polinômio de quarto grau exatamente. É claro que, para a maioria das outras funções, isso não ocorreria e nossa estimativa da derivada seria melhorada, mas não perfeita. Conseqüentemente, como foi o caso na aplicação da extrapolação de Richardson, a abordagem pode ser aplicada iterativamente usando um algoritmo de Romberg até que o resultado caia abaixo de um critério de erro aceitável.

23.3 DERIVADAS DE DADOS DESIGUALMENTE ESPAÇADOS

As abordagens discutidas até agora foram desenvolvidas principalmente para determinar a derivada de uma dada função. Para as aproximações por diferenças divididas finitas da Seção 23.1, os dados precisavam estar igualmente espaçados. Para a técnica de extrapolação de Richardson da Seção 23.2, os dados precisavam estar igualmente espaçados e ser gerados por divisões sucessivas na metade dos intervalos. Tal controle do espaçamento dos dados, em geral, estará disponível apenas quando usarmos uma função para gerar uma tabela de valores.

Em contraste, informações obtidas experimentalmente — isto é, dados de experiências ou estudos de campo — são, em geral, coletadas em intervalos desiguais. Tais informações não podem ser analisadas com as técnicas discutidas até este ponto.

Uma forma de tratar dados desigualmente espaçados é ajustar um polinômio interpolador de Lagrange de segundo grau [lembre-se da Equação (18.23)] para cada conjunto de três pontos adjacentes. Lembre-se de que esse polinômio não exige que os pontos sejam igualmente espaçados. O polinômio de segundo grau pode ser deduzido analiticamente, o que dá

$$f'(x) = f(x_{i-1})\frac{2x - x_i - x_{i+1}}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} + f(x_i)\frac{2x - x_{i-1} - x_{i+1}}{(x_i - x_{i-1})(x_i - x_{i+1})} + f(x_{i+1})\frac{2x - x_{i-1} - x_i}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)} \tag{23.9}$$

em que x é o valor no qual se quer estimar a derivada. Embora essa equação seja certamente mais complicada do que a aproximação da primeira derivada das Figuras 23.1 a 23.3, tem algumas vantagens importantes. Em primeiro lugar, pode ser usada para se obter uma estimativa da derivada em qualquer lugar do intervalo prescrito por esses três pontos. Em segundo lugar, os pontos propriamente ditos não precisam ser igualmente espaçados. Em terceiro, a estimativa da derivada tem a mesma ordem de acurácia que a diferença centrada [Equação (4.22)]. Na realidade, para pontos igualmente espaçados, a Equação (23.9) calculada em $x = x_i$ se reduz à Equação (4.22).

EXEMPLO 23.3 Derivando Dados Desigualmente Espaçados

Enunciado do Problema. Como na Figura 23.4, pode ser medido um gradiente de temperatura no solo. O fluxo de calor na interface solo-ar pode ser calculado pela lei de Fourier,

$$q(z = 0) = -k\rho C \left.\frac{dT}{dz}\right|_{z=0}$$

em que q é o fluxo de calor (W/m^2), k é o coeficiente de difusão térmica do solo ($\cong 3{,}5 \times 10^{-7}$ m^2/s), ρ é a densidade do solo ($\cong 1.800$ kg/m^3) e C é o calor específico do solo ($\cong 840$ J/(kg $\cdot$ °C)). Observe que um valor positivo para o fluxo significa que é transferido calor do ar para o solo. Use derivação numérica para calcular o gradiente na interface soloar e empregue essa estimativa para determinar o fluxo de calor para dentro do solo.

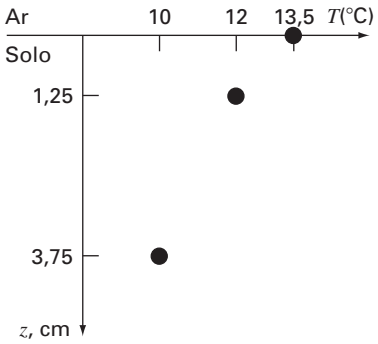


FIGURA 23.4
Temperatura em função da profundidade do solo.

Solução. A Equação (23.9) pode ser usada para calcular a derivada como sendo

$$f'(x) = 13{,}5\frac{2(0)-1{,}25-3{,}75}{(0-1{,}25)(0-3{,}75)} + 12\frac{2(0)-0-3{,}75}{(1{,}25-0)(1{,}25-3{,}75)} + 10\frac{2(0)-0-1{,}25}{(3{,}75-0)(3{,}75-1{,}25)}$$
$$= -14{,}4 + 14{,}4 - 1{,}333333 = -1{,}333333^\circ\text{C/cm}$$

o que pode ser usado para calcular (observe que 1 W = 1 J/s),

$$q(z=0) = -3{,}5\times 10^{-7}\frac{\text{m}^2}{\text{s}}\left(1.800\frac{\text{kg}}{\text{m}^3}\right)\left(840\frac{\text{J}}{\text{kg}\cdot{}^\circ\text{C}}\right)\left(-133{,}3333\frac{^\circ\text{C}}{\text{m}}\right)$$
$$= 70{,}56\ \text{W/m}^2$$

23.4 DERIVADAS E INTEGRAIS PARA DADOS COM ERROS

Além de espaçamento desigual, um outro problema relacionado com a derivação de dados empíricos é que eles, em geral, incluem erros nas medidas. Uma deficiência da derivação numérica é que ela tende a amplificar os erros nos dados. A Figura 23.5*a* mostra dados lisos, sem erros, que, quando derivados, fornecem um resultado liso (Figura 23.5*c*). Em contraste, a Figura 23.5*b* usa os mesmos dados, mas com alguns pontos ligeiramente levantados e alguns ligeiramente abaixados. Essa pequena modificação quase não é percebida a partir da Figura 23.5*b*. Entretanto, o efeito resultante na Figura 23.5*d* é significativo, porque o processo de derivação amplifica erros.

Como poderia ser esperado, a abordagem principal para determinar derivadas para dados imprecisos é usar a regressão por mínimos quadrados para ajustar uma função diferenciável (lisa) aos dados. Na ausência de outras informações, uma regressão polinomial de grau baixo poderia ser uma boa primeira escolha. Obviamente, se a relação funcional verdadeira entre as variáveis independente e dependente for conhecida, deve formar a base para o ajuste por mínimos quadrados.

23.4.1 Derivação *versus* Integração de Dados Incertos

Da mesma maneira como as técnicas de ajuste de curvas, como a regressão, podem ser usadas para derivar dados incertos, um processo similar pode ser usado para a integração. Entretanto, por causa da diferença de estabilidade entre a diferenciação e a integração, isso raramente é feito.

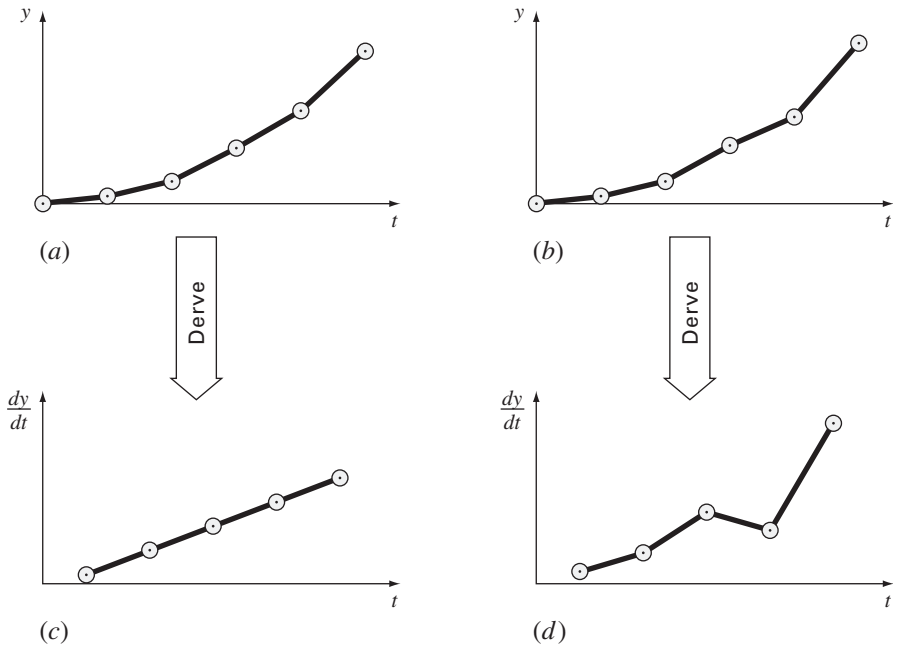


FIGURA 23.5
Ilustração de como pequenos erros nos dados são amplificados pela derivação numérica: (*a*) dados sem erros, (*b*) dados ligeiramente modificados, (*c*) a derivação numérica resultante da curva (*a*), e (*d*) a derivação resultante da curva (*b*) manifestando um aumento na variabilidade. Em contraste, a operação inversa de integração [movendo-se de (*d*) para (*b*) tomando a área sob (*d*)] tende a atenuar ou alisar os erros nos dados.

Como descrito na Figura 23.5, a derivação tende a ser instável — ou seja, amplifica os erros. Em contraste, o fato de a integração ser um processo de soma tende a torná-la muito tolerante em relação à incerteza nos dados. Em essência, como os pontos, em geral, são somados para formar uma integral, erros aleatórios positivos e negativos tendem a se cancelar. Por outro lado, como a derivação envolve subtração, erros aleatórios positivos e negativos tendem a se somar.

23.5 INTEGRAÇÃO E DERIVAÇÃO NUMÉRICAS COM BIBLIOTECAS E PACOTES

As bibliotecas e os pacotes de software têm grandes recursos para integração e derivação numéricas. Nesta seção, será feita uma pequena apresentação dos mais úteis deles.

23.5.1 MATLAB

O software MATLAB tem diversas funções de biblioteca que permitem que funções e dados sejam integrados e derivados. O exemplo a seguir ilustra como algumas delas podem ser usadas.

EXEMPLO 23.4 Usando o MATLAB para Integração e Derivação

Enunciado do Problema. Explore como o MATLAB pode ser usado para integrar e derivar a função

$$f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

de $a = 0$ a $b = 0{,}8$. Lembre-se, dos Capítulos 21 e 22, que o valor verdadeiro da integral pode ser determinado analiticamente e é 1,640533.

Solução. Primeiro, vamos usar a função **quad** do MATLAB para integrar a função. Para usar o quad, antes é preciso desenvolver um M-file para armazenar a função. Usando um editor de texto, podemos criar o seguinte arquivo:

```
function y=fx(x)
y=0.2+25*x-200*x.^2+675*x.^3-900*x.^4+400*x.^5;
```

Ele, então, pode ser armazenado no diretório MATLAB como fx.m.

Depois de entrar no MATLAB, podemos chamar o quad digitando

```
>> Q=quad('fx',0,.8)
```

em que as duas últimas entradas são os extremos de integração. O resultado é

```
Q=
  1.6405
```

Assim, o MATLAB fornece uma estimativa acurada da integral.

A seguir, vamos investigar como o MATLAB trata integrais de dados tabulados. Para fazê-lo, vamos repetir o Exemplo 21.7, em que calculamos a função em intervalos desiguais (lembre-se da Tabela 21.3). Podemos gerar a mesma informação no MATLAB primeiro definindo os valores da variável independente,

```
>> x=[0 .12 .22 .32 .36 .4 .44 .54 .64 .7 .8];
```

Então, geramos um vetor *y* contendo os valores correspondentes das variáveis dependentes chamando fx;

```
>> y=fx(x)

y =
  Columns 1 through 7
    0.2000   1.3097   1.3052   1.7434   2.0749   2.4560
    2.8430
  Columns 8 through 11
    3.5073   3.1819   2.3630   0.2320
```

Podemos integrar esses valores chamando a função **trapz**,

```
>> integral=trapz(x,y)

integral =
    1.5948
```

Assim, como o nome indica, a trapz usa a regra do trapézio em cada intervalo e soma os resultados para obter a integral total.

Finalmente, podemos derivar os dados desigualmente espaçados em *x* e *y*. Para isso, usamos a função **diff**, que simplesmente determina as diferenças entre os elementos adjacentes de um vetor, por exemplo,

```
>> diff(x)

ans =
  Columns 1 through 7
    0.1200   0.1000   0.1000   0.0400   0.0400   0.0400
    0.1000
  Columns 8 through 10
    0.1000   0.0600   0.1000
```

O resultado representa a diferença entre cada par de elementos de *x*. Para calcular as aproximações por diferenças divididas da derivada, simplesmente fazemos a divisão vetorial das diferenças de *y* pelas diferenças de *x* entrando

```
>> d=diff(y)./diff(x)
```

o que fornece

```
d =
  Columns 1 through 7
    9.2477   -0.0449    4.3815    8.2877    9.5274    9.6746
    6.6431
  Columns 8 through 10
   -3.2537   -13.6488   -21.3100
```

Estes resultados representam estimativas grosseiras das derivadas em cada intervalo. Essa abordagem poderia ser refinada usando um espaçamento mais fino.

23.5.2 IMSL

O IMSL tem diversas rotinas para a integração e derivação (Tabela 23.1). Na discussão presente, nos concentraremos na rotina **QDAG**, que integra uma função usando um esquema adaptativo global baseado nas regras de Gauss-Kronrod.

TABELA 23.1 Rotinas IMSL para integrar e derivar.

Categoria	Rotinas	Recursos
Quadratura unidimensional		
	QDAGS	Adaptativo de propósito geral, com singularidade nas extremidades
	QDAG	Adaptativo de propósito geral
	QDAGP	Adaptativo de propósito geral com pontos de singularidade
	QDAGI	Adaptativo de propósito geral, para intervalo infinito
	QDAWO	Adaptativo oscilatório com pesos (trigonométrico)
	QDAWF	Adaptativo de Fourier com pesos (trigonométrico)
	QDAWS	Adaptativo algébrico com pesos e singularidades nas extremidades
	QDAWC	Adaptativo com pesos para o valor principal de Cauchy
	QDNG	Não-adaptativo de caráter geral
Quadratura multidimensional		
	TWODQ	Quadratura bidimensional (integral iterada)
	QAND	Quadratura N-dimensional adaptativa em um hiper-retângulo
Regras de Gauss e recorrências de três termos		
	GQRUL	Regra de quadratura de Gauss para pesos clássicos
	GQRCF	Regra de quadratura de Gauss a partir de coeficientes de recorrência
	RECCF	Coeficientes de recorrências para pesos clássicos
	RECQR	Coeficientes de recorrência a partir de regra de quadratura
	FQRUL	Regra de quadratura de Fejer
Derivação		
	DERIV	Aproximações para a primeira, segunda e terceira derivadas

O QDAG é implementado pelo seguinte comando de chamada:

```
CALL QDAG (F, A, B, ERRABS, ERRREL, IRULE, RESULT, ERREST)
```

em que

F = FUNCTION fornecida pelo usuário a ser integrada. A forma é F(X), em que X é a variável independente. Observe que F deve ser declarada EXTERNAL no programa de chamada.

A = Extremo inferior de integração. (Entrada)

B = Extremo superior de integração. (Entrada)

ERRABS = Acurácia absoluta desejada. (Entrada)

ERRREL = Acurácia relativa desejada. (Entrada)

IRULE = Escolha da regra de quadratura. (Entrada). IRULE = 2 é recomendado para a maioria das funções. Se a função tiver uma singularidade em pico, use IRULE =1. Se a função for oscilante, use IRULE = 6.

RESULT = Estimativa da integral de A a B de F. (Saída)

ERREST = Estimativa do valor absoluto do erro. (Saída)

EXEMPLO 23.5 Usando IMSL para Integrar uma Função

Enunciado do Problema. Use QDAG para determinar a integral de

$$f(x) = 0{,}2 + 25x - 200x^2 + 675x^3 - 900x^4 + 400x^5$$

de $a = 0$ a $b = 0{,}8$. Lembre-se, a partir dos Capítulos 21 e 22, de que o valor verdadeiro da integral pode ser determinado analiticamente e é 1,640533.

Solução. Um exemplo de um programa principal em Fortran 90 e função usando QDAG para resolver esse problema pode ser escrito como

```
PROGRAM Integrate
USE mimsl

IMPLICIT NONE
INTEGER::irule=1
REAL::a=0.,b=0.8,errabs=0.0.errrel=0.001
REAL::errest,res,f
EXTERNAL f

CALL QDAG (f,a,b,errabs,errrel,irule,res,errest)

PRINT '('' Calculado ='',F8.4)', res
PRINT '('' Estimativa do erro ='',1PE10.3)', errest

END PROGRAM

FUNCTION f(x)
IMPLICIT NONE
REAL::x,f
f=0.2+25.*X-200.*X**2+675.*X**3.-900.*X**4+400.*X**5
END FUNCTION
```

Saída:

```
Calculado = 1.6405
Estimativa do erro = 5.000E-05
```

PROBLEMAS

23.1 Calcule aproximações por diferenças progressivas e regressivas de $O(h)$ e de $O(h^2)$ e aproximações por diferenças centradas de $O(h^2)$ e de $O(h^4)$ para a primeira derivada de $y = \cos x$ em $x = \pi/4$ usando um valor de $h = \pi/12$. Faça uma estimativa do erro relativo porcentual verdadeiro ε_t para cada aproximação.

23.2 Repita o Problema 23.1, mas para $y = \log x$ calculado em $x = 25$ com $h = 2$.

23.3 Use aproximações por diferenças centradas para obter estimativas para a primeira e a segunda derivadas de $y = e^x$ em $x = 2$ com $h = 0{,}1$. Use ambas as fórmulas de $O(h^2)$ e $O(h^4)$ nas suas estimativas.

23.4 Use a extrapolação de Richardson para obter uma estimativa da primeira derivada de $y = \cos x$ em $x = \pi/4$ utilizando tamanhos de passo de $h_1 = \pi/3$ e $h_2 = \pi/6$. Use diferenças centradas de $O(h^2)$ para a estimativa inicial.

23.5 Repita o Problema 23.4, mas para a primeira derivada de $\ln x$ em $x = 5$ usando $h_1 = 2$ e $h_2 = 1$.

23.6 Utilize a Equação (23.9) para determinar a primeira derivada de $y = 2x^4 - 6x^3 - 12x - 8$ em $x = 0$ com base nos valores em $x_0 = -0{,}5$, $x_1 = 1$, e $x_2 = 2$. Compare esse resultado com o valor verdadeiro e com a estimativa obtida usando uma aproximação por diferença centrada com base em $h = 1$.

23.7 Demonstre que, para pontos igualmente espaçados, a Equação (23.9) se reduz à Equação (4.22) em $x = x_i$.

23.8 Calcule as aproximações por diferença centrais de primeira ordem de $O(h^4)$ para cada uma das seguintes funções nas posições especificadas e para o tamanho de passo especificado:

(a) $y = x^3 + 4x - 15$ em $x = 0$, $h = 0{,}25$
(b) $y = x^2 \cos x$ em $x = 0{,}4$, $h = 0{,}1$
(c) $y = \text{tg}(x/3)$ em $x = 3$, $h = 0{,}5$
(d) $y = \text{sen}(0{,}5\sqrt{x})/x$ em $x = 1$, $h = 0{,}2$
(e) $y = e^x + x$ em $x = 2$, $h = 0{,}2$

Compare seus resultados com as soluções analíticas.

23.9 Foram coletados os seguintes dados para a distância percorrida em função do tempo para um foguete:

t, s	0	25	50	75	100	125
y, km	0	32	58	78	92	100

Use derivação numérica para obter estimativas da velocidade e da aceleração em cada instante.

23.10 Desenvolva um programa amigável ao usuário para aplicar o algoritmo de Romberg para obter uma estimativa da derivada de uma dada função.

23.11 Desenvolva um programa amigável ao usuário para obter estimativas da primeira derivada para dados desigualmente espaçados. Teste seu programa com os seguintes dados:

x	1	1,5	1,6	2,5	3,5
$f(x)$	0,6767	0,3734	0,3261	0,08422	0,01596

em que $f(x) = 5e^{-2x}x$. Compare seus resultados com as derivadas verdadeiras.

23.12 Lembre-se de que, para o problema do pára-quedista em queda livre, a velocidade é dada por

$$v(t) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{P23.12a}$$

e a distância percorrida pode ser obtida por

$$d(t) = \frac{gm}{c}\int_0^t \left(1 - e^{-(c/m)t}\right) dt \tag{P23.12b}$$

Dados $g = 9{,}81$, $m = 70$ e $c = 12$,

(a) Use o MATLAB para integrar a Equação (P23.12*a*) de $t = 0$ a 10.
(b) Integre analiticamente a Equação (P23.12*b*) com a condição inicial que $d = 0$ em $t = 0$. Calcule o resultado em $t = 10$ para confirmar (**a**).
(c) Use o MATLAB para derivar a Equação (P23.12*a*) em $t = 10$.
(d) Derive analiticamente a Equação (P23.12*a*) em $t = 10$ para confirmar (**c**).

23.13 A distribuição normal é definida por

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

(a) Use o MATLAB para integrar essa função de $x = -1$ a 1 e de -2 a 2.
(b) Use o MATLAB para determinar os pontos de inflexão dessa função.

23.14 Os seguintes dados foram gerados a partir de uma distribuição normal:

x	−2	−1,5	−1	−0,5	0	0,5	1	1,5	2
$f(x)$	0,05399	0,12952	0,24197	0,35207	0,39894	0,35207	0,24197	0,12952	0,05399

(a) Use o MATLAB para integrar esses dados de $x = -1$ a 1 e de -2 a 2 com a função `trap`.
(b) Use o MATLAB para obter estimativas para os pontos de inflexão desses dados.

23.15 Use o IMSL para integrar a distribuição normal (veja Problema 23.13) de $x = -1$ a 1, de -2 a 2 e de -3 a 3.

23.16 Escreva um problema em MATLAB para integrar

$$\int_0^{\pi/2} \cos(\cos x)\, dx$$

23.17 Escreva um programa em MATLAB para integrar

$$\int_0^{2\pi} \frac{\text{sen } t}{t}\, dt$$

usando as funções **quad** e **quadL**. Para saber mais sobre o **quadL**, digite

```
help quadL
```

no *prompt* do MATLAB.

23.18 Use o comando **diff(y)** no MATLAB e calcule a aproximação por diferença finita para a primeira e a segunda derivadas em cada valor de x na tabela a seguir, excluindo as duas extremidades. Use aproximações por diferenças finitas que sejam corretas até segunda ordem, $O(\Delta x^2)$.

x	0	1	2	3	4	5	6	7	8	9	10
y	1,4	2,1	3,3	4,8	6,8	6,6	8,6	7,5	8,9	10,9	10

23.19 O objetivo desse problema é comparar aproximações por diferenças finitas progressiva, regressiva e centrada acuradas até segunda ordem de uma função com o valor real da derivada. Isso será feito para

$$f(x) = e^{-2x} - x$$

(a) Use o cálculo para determinar o valor correto da derivada em $x = 2$.
(b) Para calcular aproximações por diferenças finitas centradas, comece com $\Delta x = 0{,}5$. Assim, para o primeiro cálculo, os valores de x para a aproximação por diferença centrada serão $x = 2 \pm 0{,}5$ ou $x = 1{,}5$ e 2,5. Então, diminua em incrementos de 0,01 até o valor mínimo de $\Delta x = 0{,}01$.
(c) Repita a parte (**b**) para diferenças progressivas e regressivas de segunda ordem. (Observe que isso pode ser feito ao mesmo tempo em que a diferença centrada é calculada no laço.)
(d) Faça os gráficos dos resultados de (**b**) e (**c**) em função de x. Inclua o resultado exato no gráfico para comparação.

23.20 Use uma expansão em série de Taylor para deduzir uma aproximação por diferença finita centrada para a segunda derivada que seja acurada até segunda ordem. Para isso, você deve usar quatro expansões diferentes para os pontos x_{i-2}, x_{i-1}, x_{i+1} e x_{i+2}. Em cada caso, a expansão será em torno do ponto x_i. O intervalo Δx será usado no caso de $i - 1$ e $i + 1$ e $2\Delta x$ será usado no caso de $i - 2$ e $i + 2$. As quatro equações devem ser combinadas de forma a eliminar a primeira e a segunda derivadas. Mantenha termos suficientes em cada expansão para calcular o primeiro termo que será truncado para determinar a ordem da aproximação.

23.21 Use os seguintes dados para determinar a velocidade e a aceleração em $t = 10$ segundos:

Tempo, t, s	0	2	4	6	8	10	12	14	16
Posição, x, m	0	0,7	1,8	3,4	5,1	6,3	7,3	8,0	8,4

Use métodos de (**a**) diferença finita centrada, (**b**) diferença finita progressiva e (**c**) diferença finita regressiva corrigidos de segunda ordem.

23.22 Um avião está sendo monitorado por um radar e os dados foram tomados a cada segundo em coordenadas polares θ e r.

t, s	200	202	204	206	208	210
θ, (rad)	0,75	0,72	0,70	0,68	0,67	0,66
r, m	5.120	5.370	5.560	5.800	6.030	6.240

Em 206 segundos, use diferença finita centrada (corrigida de segunda ordem) para determinar as expressões vetoriais da velocidade $\vec{v}$ e da aceleração $\vec{a}$. A velocidade e a aceleração dadas em coordenadas polares são:

$$\vec{v} = \dot{r}\vec{e}_r + r\dot{\theta}\vec{e}_\theta \qquad \text{e} \qquad \vec{a} = (\ddot{r} - r\dot{\theta}^2)\vec{e}_r + (r\ddot{\theta} + 2\dot{r}\dot{\theta})\vec{e}_\theta$$

23.23 Desenvolva um programa macro VBA em Excel para ler em colunas adjacentes os valores de x e y de uma planilha. Calcule as derivadas em cada ponto usando a Equação (23.9) e mostre os resultados em uma terceira coluna adjacente aos valores de x e y na planilha. Teste seu programa aplicando-o para calcular as velocidades para os valores de tempo e posição do Problema 23.21.

23.24 Use regressão para obter uma estimativa da aceleração para cada instante de tempo para os seguintes dados com polinômios de segundo, terceiro e quarto graus. Faça um gráfico de seus resultados.

t	1	2	3,25	4,5	6	7	8	8,5	9,3	10
v	10	12	11	14	17	16	12	14	14	10

23.25 Você deve medir a vazão de água através de um pequeno tubo. Para isso, coloca um balde na saída do tubo e mede o volume no balde como uma função do tempo, como tabulado a seguir. Faça uma estimativa da vazão em $t = 7$ s.

Tempo, s	0	1	5	8
Volume, cm^3	0	1	8	16,4

23.26 A velocidade v (m/s) do ar escoando por uma superfície plana é medida a diversas distâncias y (m) da superfície. Determine a tensão de cisalhamento τ (N/m^2) na superfície ($y = 0$),

$$\tau = \mu \frac{dv}{dy}$$

Suponha o valor $\mu = 1{,}8 \times 10^{-5}$ N · s/m^2 para a viscosidade dinâmica.

y, m	0	0,002	0,006	0,012	0,018	0,024
v, m/s	0	0,287	0,899	1,915	3,048	4,299

23.27 As reações químicas freqüentemente seguem o modelo:

$$\frac{dc}{dt} = -kc^n$$

em que c é a concentração, t é o tempo, k é a taxa de reação e n é a ordem da reação. Dados valores de c e dc/dt, k e n podem ser calculados por uma regressão linear do logaritmo desta equação:

$$\log\left(-\frac{dc}{dt}\right) = \log k + n \log c$$

Use essa abordagem junto com os seguintes dados para obter estimativas de k e n:

t	10	20	30	40	50	60
c	3,52	2,48	1,75	1,23	0,87	0,61

CAPÍTULO 24

Estudo de Casos: Integração e Derivação Numéricas

O propósito deste capítulo é aplicar os métodos de integração e derivação numéricas discutidos na Parte Seis a problemas práticos de engenharia. Duas situações são encontradas mais freqüentemente. No primeiro caso, a função sob estudo pode ser expressa na forma analítica, mas é muito complicada para ser calculada de modo simples usando os métodos do cálculo. Os métodos numéricos são aplicados a situações desse tipo usando-se a expressão analítica para gerar uma tabela de argumentos e valores da função. No segundo caso, a função a ser analisada é inerentemente tabular, por natureza. Esse tipo de função usualmente representa uma série de medidas, observações ou alguma outra informação empírica. Os dados em ambos os casos são diretamente compatíveis com diversos esquemas discutidos nesta parte do livro.

A Seção 24.1, que trata de cálculos de calor na engenharia química, envolve equações. Nessa aplicação, uma função na forma analítica é integrada numericamente para determinar o calor necessário para elevar a temperatura de um material.

As Seções 24.2 e 24.3 também envolvem funções que estão disponíveis na forma de equação. A Seção 24.2, que é tirada da engenharia civil, usa integração numérica para determinar a força total do vento que age no mastro de um veleiro de corrida. A Seção 24.3 determina a corrente eficaz para um circuito elétrico. Esse exemplo é utilizado para mostrar a utilidade da integração de Romberg e da quadratura de Gauss.

A Seção 24.4 se concentra na análise de informações tabuladas para determinar o trabalho necessário para mover um bloco. Embora essa aplicação tenha uma ligação direta com a engenharia mecânica, é relevante para todas as outras áreas da engenharia. Entre outras coisas, usamos esse exemplo para ilustrar a integração de dados desigualmente espaçados.

24.1 INTEGRAÇÃO PARA DETERMINAR A QUANTIDADE TOTAL DE CALOR (ENGENHARIA QUÍMICA/BIOENGENHARIA)

Fundamentos. Cálculos de calor são empregados rotineiramente na engenharia química e na bioengenharia, bem como em muitos outros campos da engenharia. Essa aplicação fornece um exemplo simples, mas útil, de tais cálculos.

Um problema freqüentemente encontrado é a determinação da quantidade de calor necessária para aumentar a temperatura de um material. A característica necessária para efetuar esses cálculos é a capacidade calorífica c. Esse parâmetro representa a quantidade de calor necessária para elevar uma unidade de massa de uma unidade de temperatura. Se c for constante no intervalo de temperaturas sendo examinado, o calor necessário ΔH (em calorias) pode ser calculado por

$$\Delta H = mc\ \Delta T \tag{24.1}$$

em que c tem unidades de cal/(g · °C), m é a massa e ΔT é a variação de temperatura (°C). Por exemplo, a quantidade de calor necessária para elevar a temperatura de 20 g de água de 5 para 10 °C é igual a

$$\Delta H = 20(1)(10 - 5) = 100 \text{ cal}$$

em que a capacidade calorífica da água é aproximadamente 1 cal/(g · °C). Tal cálculo é adequado quando ΔT é pequeno. Entretanto, para grandes variações de temperatura, a capacidade calorífica não é constante e, de fato, varia como função da temperatura. Por exemplo, a capacidade calorífica de um material poderia aumentar de acordo com uma relação como

$$c(T) = 0{,}132 + 1{,}56 \times 10^{-4}T + 2{,}64 \times 10^{-7}T^2 \tag{24.2}$$

Nesse caso, deve-se calcular a quantidade de calor necessária para elevar a temperatura de 1.000 g desse material de −100 a 200 °C.

Solução. A Equação (PT6.4) fornece uma maneira de calcular o valor médio de $c(T)$:

$$\bar{c}(T) = \frac{\int_{T_1}^{T_2} c(T)\,dT}{T_2 - T_1} \tag{24.3}$$

o que pode ser substituído na Equação (24.1) para fornecer

$$\Delta H = m \int_{T_1}^{T_2} c(T)\,dT \tag{24.4}$$

em que $\Delta T = T_2 - T_1$. Agora, como nesse caso $c(T)$ é uma função quadrática simples, ΔH pode ser determinado analiticamente. A Equação (24.2) é substituída na Equação (24.4) e o resultado integrado para fornecer um valor exato de $\Delta H = 42.732$ cal. É útil e instrutivo comparar esse resultado com os métodos numéricos desenvolvidos no Capítulo 21. Para isso, é necessário gerar uma tabela de valores de c para diversos valores de T:

T, °C	c, cal/(g · °C)
−100	0,11904
−50	0,12486
0	0,13200
50	0,14046
100	0,15024
150	0,16134
200	0,17376

Esses pontos podem ser usados em conjunto com a regra 1/3 de Simpson com seis segmentos para calcular uma estimativa da integral de 42.732. O resultado pode ser substituído na Equação (24.4) para fornecer um valor de $\Delta H = 42.732$, que coincide exatamente com a solução analítica. Tal coincidência exata ocorreria não importando o número de segmentos usados. Isso deveria ser esperado, já que c é uma função quadrática e a regra de Simpson é exata para polinômios de grau menor ou igual a três (ver Seção 21.2.1).

TABELA 24.1 Resultados usando a regra do trapézio com diversos tamanhos de passos.

Tamanho do Passo, °C	ΔH	ε_t (%)
300	96.048	125
150	43.029	0,7
100	42.864	0,3
50	42.765	0,07
25	42.740	0,018
10	42.733,3	<0,01
5	42.732,3	<0,01
1	42.732,01	<0,01
0,05	42.732,00003	<0,01

Os resultados usando a regra do trapézio estão listados na Tabela 24.1. Vemos que a regra do trapézio também é capaz de fornecer uma estimativa do calor total bem acurada. Entretanto, é necessário um tamanho de passo bem pequeno ($< 10\ ^\circ$C) para uma precisão de cinco casas. Esse exemplo é uma boa ilustração de por que a regra de Simpson é tão popular. Ela é fácil de aplicar tanto com uma calculadora manual quanto com um computador pessoal. Além disso, usualmente é suficientemente acurada para tamanhos de passo relativamente grandes e é exata para polinômios de grau menor ou igual a três.

24.2 FORÇA EFETIVA NO MASTRO DE UM VELEIRO DE CORRIDA (ENGENHARIA CIVIL/AMBIENTAL)

Fundamentos. Uma seção transversal de um veleiro de corrida é mostrada na Figura 24.1*a*. A força do vento (f) exercida por pé do mastro da vela varia como uma função da distância acima do convés do barco (z), como na Figura 24.1*b*. Calcule a força de tensão T no cabo de suporte esquerdo do mastro supondo que o cabo de suporte direito está completamente frouxo e que o mastro está ligado ao convés de modo que transmite forças horizontais e verticais mas não transmite torques. Suponha que o mastro permaneça vertical.

Solução. Para prosseguir com esse problema, é necessário que a força distribuída f seja convertida em uma força total equivalente F e que sua posição efetiva d acima do convés seja calculada (Figura 24.2). Esse cálculo é complicado pelo fato de que a força exercida por pé do mastro varia com a distância acima do mastro. A força total exercida no mastro pode ser expressa como a integral de uma função contínua:

$$F = \int_0^{30} 200\left(\frac{z}{5+z}\right)e^{-2z/30}\,dz \tag{24.5}$$

Essa integral não-linear é difícil de calcular analiticamente. Portanto, é conveniente empregar abordagens numéricas como a regra de Simpson e a regra do trapézio para esse problema. Pode-se fazê-lo calculando $f(z)$ para diversos valores de z e então usando a Equação (21.10) ou (21.18). Por exemplo, a Tabela 24.2 tem valores de $f(z)$ para um tamanho de passo de 3 pés que fornece dados para a regra 1/3 de Simpson ou para a regra do trapézio. Resultados para diversos tamanhos de passo são dados na Tabela 24.3. Observa-se que ambos os métodos dão um valor de $F = 1.480{,}6$ lb quando o tamanho do passo se torna pequeno. Nesse caso, tamanhos de passo de 0,05 pé para a regra do trapézio e 0,5 pé para a regra de Simpson fornecem bons resultados.

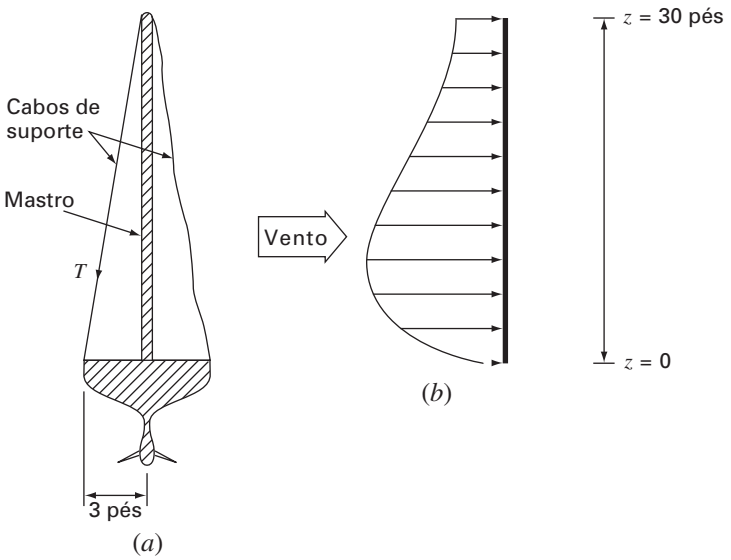


FIGURA 24.1
(*a*) Seção transversal de um veleiro de corrida. (*b*) Força do vento f exercida por pé do mastro como uma função da distância z acima do convés do barco.

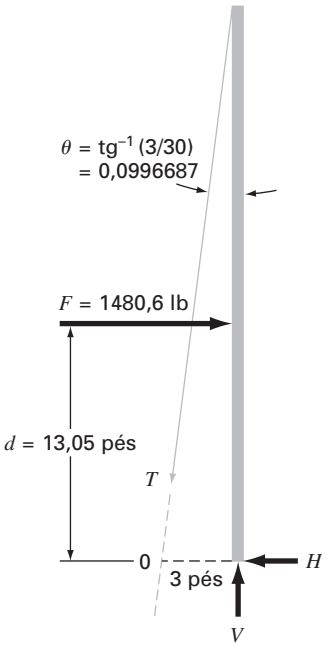


FIGURA 24.2
Diagrama de corpo livre das forças exercidas no mastro de um veleiro.

TABELA 24.2 Valores de $f(z)$ para um tamanho de passo de 3 pés que fornecem dados para a regra do trapézio e a regra 1/3 de Simpson.

z, pés	0	3	6	9	12	15	18	21	24	27	30
$f(z)$, lb/pés	0	61,40	73,13	70,56	63,43	55,18	47,14	39,83	33,42	27,89	23,20

TABELA 24.3 Valores de F calculados com base em diversas versões da regra do trapézio e da regra 1/3 de Simpson.

Técnica	Tamanho do Passo, pés	Segmentos	F, lb
Regra do trapézio	15	2	1.001,7
	10	3	1.222,3
	6	5	1.372,3
	3	10	1.450,8
	1	30	1.477,1
	0,5	60	1.479,7
	0,25	120	1.480,3
	0,1	300	1.480,5
	0,05	600	1.480,6
Regra 1/3 de Simpson	15	2	1.219,6
	5	6	1.462,9
	3	10	1.476,9
	1	30	1.480,5
	0,5	60	1.480,6

A posição de ação efetiva de F (Figura 24.2) pode ser obtida pelo cálculo da integral

$$d = \frac{\int_0^{30} zf(z)\,dz}{\int_0^{30} f(z)\,dz} \tag{24.6}$$

ou

$$d = \frac{\int_0^{30} 200z[z/(5+z)]e^{-2z/30}\,dz}{1.480,6} \tag{24.7}$$

Essa integral pode ser calculada usando métodos similares aos usados acima. Por exemplo, a regra 1/3 de Simpson com um tamanho de passo de 0,5 fornece $d = 19.326,9/1.480,6 = 13,05$ pés.

Com F e d determinados a partir de métodos numéricos, é usado um diagrama de corpo livre para desenvolver as equações de balanço de força e balanço de torque. Esse diagrama de corpo livre é mostrado na Figura 24.2. Somando as forças na direção horizontal e na vertical e tomando os torques em relação ao ponto 0, obtemos

$$\Sigma F_H = 0 = F - T \operatorname{sen}\theta - H \tag{24.8}$$

$$\Sigma F_V = 0 = V - T\cos\theta \tag{24.9}$$

$$\Sigma M_0 = 0 = 3V - Fd \tag{24.10}$$

em que T é a tensão no cabo e H e V são as reações desconhecidas no mastro transmitidas pelo convés. O sentido e o módulo de H e V são desconhecidos. A Equação (24.10) pode ser resolvida diretamente para determinar V, pois F e d são conhecidos.

$$V = \frac{Fd}{3} = \frac{(1.480,6)(13,05)}{3} = 6.440,6 \text{ lb}$$

Portanto, da Equação (24.9),

$$T = \frac{V}{\cos\theta} = \frac{6.440{,}6}{0{,}995} = 6.473 \text{ lb}$$

e, da Equação (24.8),

$$H = F - T\,\text{sen}\,\theta = 1.480{,}6 - (6.473)(0{,}0995) = 836{,}54 \text{ lb}$$

Essas forças agora permitem prosseguir para outros aspectos do projeto estrutural do barco, como os cabos e o sistema de suporte do mastro no convés. Esse problema ilustra bem duas utilidades da integração numérica que podem ser encontradas durante o projeto de estruturas em engenharia. Parece que a regra do trapézio e a regra 1/3 de Simpson são fáceis de aplicar e que são ferramentas práticas para a resolução de problemas. A regra 1/3 de Simpson é mais acurada do que a regra do trapézio para o mesmo tamanho de passo, e assim, em geral, é preferida.

24.3 CORRENTE EFICAZ POR INTEGRAÇÃO NUMÉRICA (ENGENHARIA ELÉTRICA)

Fundamentos. O valor médio de uma corrente elétrica oscilante em um período pode ser zero. Por exemplo, suponha que a corrente seja descrita por uma única função senoidal: $i(t) = \text{sen}(2\pi/T)$, onde T é o período. O valor médio dessa função pode ser determinado pela seguinte equação:

$$i = \frac{\displaystyle\int_0^T \text{sen}\left(\frac{2\pi t}{T}\right) dt}{T - 0} = \frac{-\cos(2\pi) + \cos 0}{T} = 0$$

Apesar de o resultado médio ser zero, tal corrente é capaz de realizar trabalho e produzir calor. Portanto, os engenheiro elétricos, em geral, caracterizam tal corrente por

$$I_{\text{RMS}} = \sqrt{\frac{1}{T}\int_0^T i^2(t)\,dt} \tag{24.11}$$

em que $i(t)$ é a corrente instantânea. Calcule a corrente eficaz ou RMS (de *root-mean-square*, em inglês) da onda com a forma mostrada na Figura 24.3 usando a regra do trapézio, a regra l/3 de Simpson, a integração de Romberg e a quadratura de Gauss para $T = 1$ s.

Solução. As estimativas da integral para diversas aplicações da regra do trapézio e da regra 1/3 de Simpson estão listadas na Tabela 24.4. Observe que a regra de Simpson é mais acurada que a regra do trapézio.

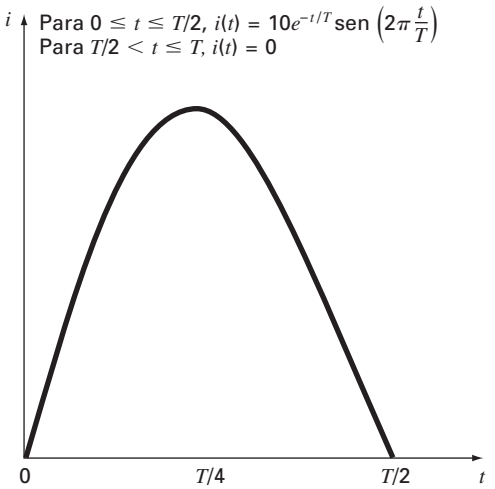


FIGURA 24.3
Uma corrente elétrica variando periodicamente.

O valor exato da integral é 15,41261. Esse resultado é obtido usando uma regra do trapézio com 128 segmentos ou uma regra de Simpson com 32 segmentos. A mesma estimativa também é obtida usando a integração de Romberg (Figura 24.4).

Além disso, a quadratura de Gauss pode ser usada para fazer a mesma estimativa. A determinação da corrente eficaz envolve o cálculo da integral ($T = 1$)

$$I = \int_0^{1/2} (10e^{-t} \operatorname{sen} 2\pi t)^2 \, dt \tag{24.12}$$

Inicialmente, é feita uma mudança de variáveis aplicando as Equações (22.23) e (22.24) para obter

$$t = \frac{1}{4} + \frac{1}{4}t_d \qquad dt = \frac{1}{4}\, dt_d$$

Essas relações podem ser substituídas na Equação (24.12) para fornecer

$$I = \int_{-1}^{1} \left[10e^{-[1/4+(1/4)t_d]} \operatorname{sen} 2\pi \left(\frac{1}{4} + \frac{1}{4}t_d \right) \right]^2 \frac{1}{4}\, dt_d \tag{24.13}$$

Para a fórmula de Gauss-Legendre de dois pontos, essa função é calculada em $t_d = -1/\sqrt{3}$ e $1/\sqrt{3}$, com os resultados sendo 7,684096 e 4,313728, respectivamente. Esses valores podem ser substituídos na Equação (22.17) para fornecer uma estimativa da integral de 11,99782, o que representa um erro de $\varepsilon_t = 22{,}1\%$.

A fórmula de três pontos é (Tabela 22.1)

$$I = 0{,}5555556(1{,}237449) + 0{,}8888889(15{,}16327) + 0{,}5555556(2{,}684915)$$
$$= 15{,}65755 \qquad |\varepsilon_t| = 1{,}6\%$$

TABELA 24.4 Valores para a integral calculados usando diversos esquemas numéricos. O erro relativo percentual ε_t é baseado no valor verdadeiro de 15,41261.

Técnica	Segmentos	Integral	ε_t (%)
Regra do trapézio	1	0,0	100
	2	15,16327	1,62
	4	15,40143	0,0725
	8	15,41196	$4{,}21 \times 10^{-3}$
	16	15,41257	$2{,}59 \times 10^{-4}$
	32	15,41261	$1{,}62 \times 10^{-5}$
	64	15,41261	$1{,}30 \times 10^{-6}$
	128	15,41261	0
Regra 1/3 de Simpson	2	20,21769	−31,2
	4	15,48082	−0,443
	8	15,41547	−0,0186
	16	15,41277	$1{,}06 \times 10^{-3}$
	32	15,41261	0

FIGURA 24.4
Resultados da utilização da integração de Romberg para obter uma estimativa da corrente eficaz.

$O(h^2)$	$O(h^4)$	$O(h^6)$	$O(h^8)$	$O(h^{10})$	$O(h^{12})$
0	20,21769	15,16503	15,41502	15,41261	15,41261
15,16327	15,48082	15,41111	15,41262	15,41261	
15,40143	15,41547	15,41225	15,41261		
15,41196	15,41277	15,41261			
15,41257	15,41262				
15,41261					

Os resultados da utilização de fórmulas com mais pontos estão resumidos na Tabela 24.5.

TABELA 24.5 Resultados da utilização de fórmulas de quadratura de Gauss com diversos números de pontos para obter aproximações da integral.

Pontos	Estimativa	ε_t (%)
2	11,9978243	22.1
3	15,6575502	−1.59
4	15,4058023	4.42×10^{-2}
5	15,4126391	-2.01×10^{-4}
6	15,4126109	-1.82×10^{-5}

A estimativa da integral de 15,41261 pode ser substituída na Equação (24.12) para calcular uma I_{RMS} de 3,925890 A. Esse resultado poderia ser utilizado para guiar outros aspectos de projeto e operação do circuito.

24.4 INTEGRAÇÃO NUMÉRICA PARA CALCULAR O TRABALHO (ENGENHARIA MECÂNICA/AEROESPACIAL)

Fundamentos. Muitos problemas de engenharia envolvem o cálculo do trabalho. A fórmula geral é

$$\text{Trabalho} = \text{força} \times \text{distância}$$

Quando esse conceito lhe foi apresentado na física do ensino médio, foram apresentadas aplicações simples usando forças que permaneciam constantes ao longo do deslocamento. Por exemplo, se uma força de 10 lb era usada para puxar um bloco por uma distância de 15 pés, o trabalho seria calculado como 150 pés · lb.

Embora um cálculo tão simples seja útil para introduzir o conceito, problemas em situações realísticas geralmente são muito mais complexos. Por exemplo, suponha que a força varie no decorrer do movimento. Em tais casos, a equação do trabalho é reescrita como

$$W = \int_{x_0}^{x_n} F(x)\, dx \tag{24.14}$$

em que W é o trabalho (pés · lb), x_0 e x_n são as posições inicial e final, respectivamente, e $F(x)$ é uma força que varia como função da posição. Se $F(x)$ for fácil de integrar, a Equação (24.14) pode ser calculada analiticamente. Entretanto, em uma situação realística de problema, a força pode não ser expressa de tal forma. De fato, quando forem analisados dados medidos, a força poderá estar disponível apenas de forma tabular. Para esses casos, a integração numérica é a única opção viável para os cálculos.

Mais complexidade é introduzida se o ângulo entre a força e a direção do movimento também variar como uma função da posição (Figura 24.5). A equação do trabalho pode ser ainda mais modificada para levar em conta esse efeito, como em

$$W = \int_{x_0}^{x_n} F(x) \cos\left[\theta(x)\right] dx \tag{24.15}$$

Novamente, se $F(x)$ e $\theta(x)$ forem funções simples, a Equação (24.15) poderá ser resolvida analiticamente. No entanto, como na Figura 24.5, agora é mais provável que a relação funcional seja complicada. Para essas situações, os métodos numéricos fornecem a única alternativa para determinar a integral.

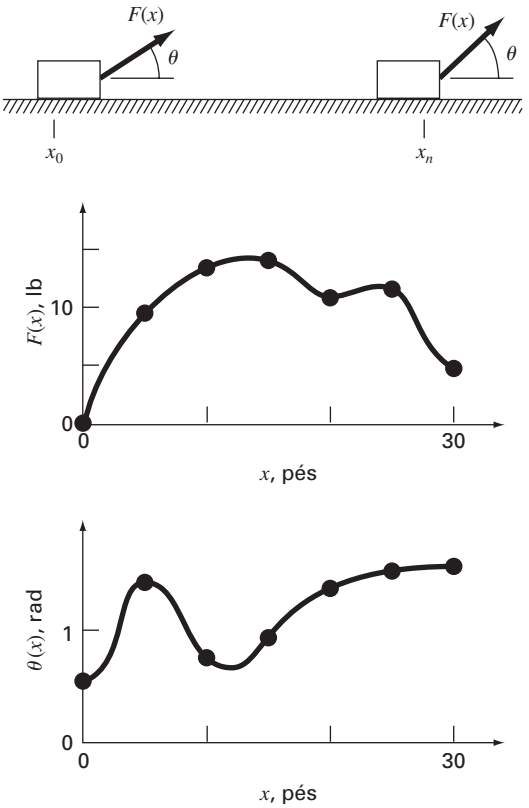


FIGURA 24,5
Uma força variável agindo em um bloco. Nesse caso, o ângulo e o valor absoluto da força variam.

TABELA 24.6 Dados para a força $F(x)$ e os ângulos $\theta(x)$ como uma função da posição x.

x, pés	$F(x)$, lb	θ, rad	$F(x) \cos \theta$
0	0,0	0,50	0,0000
5	9,0	1,40	1,5297
10	13,0	0,75	9,5120
15	14,0	0,90	8,7025
20	10,5	1,30	2,8087
25	12,0	1,48	1,0881
30	5,0	1,50	0,3537

Suponha que você deva fazer os cálculos para a situação descrita na Figura 24.5. Embora a figura mostre valores de $F(x)$ e de $\theta(x)$ no contínuo, suponha que, por causa de restrições experimentais, foram fornecidas apenas medidas discretas em intervalos de $x = 5$ pés (Tabela 24.6). Use versões com uma única aplicação e com aplicações múltiplas da regra do trapézio e das regras 1/3 e 3/8 de Simpson para calcular o trabalho para esses dados.

Solução. Os resultados da análise estão resumidos na Tabela 24.7. Foi calculado um erro relativo porcentual ε_t com relação ao valor verdadeiro da integral de 129,52 que foi estimado com base nos valores tirados da Figura 24.5 em intervalos de 1 pé.

Os resultados são interessantes porque a saída mais acurada ocorre para a simples regra do trapézio com dois segmentos. Estimativas mais refinadas usando mais segmentos, bem como as regras de Simpson, fornecem resultados menos acurados.

A razão para esse resultado aparentemente antiintuitivo é que o espaçamento grosseiro dos pontos não é adequado para capturar as variações da força e dos ângulos. Isso é particularmente evidente na Figura 24.6, em que traçamos uma curva contínua para o produto de $F(x)$ e cos $[\theta(x)]$. Observe como o uso de sete pontos para caracterizar uma função que varia no contínuo faz que os dois picos em $x = 2{,}5$ e 12,5 pés sejam perdidos. A omissão desses dois pontos efetivamente limita a acurácia das estimativas por integração numérica na Tabela 24.7. O fato de a regra do trapézio com dois segmentos fornecer o resultado mais acurado decorre de um acaso nas posições dos pontos para esse problema particular (Figura 24.7).

TABELA 24.7 Estimativa do trabalho calculada usando a regra do trapézio e as regras de Simpson. O erro relativo percentual ε_t é calculado com relação a um valor verdadeiro da integral (129,52 pés.lb) que foi obtida com base em valores em intervalos de 1 pé.

Técnica	Segmentos	Trabalho	ε_t, %
Trapézio	1	5,31	95,9
	2	133,19	2,84
	3	124,98	3,51
	6	119,09	8,05
Regra 1/3 de Simpson	2	175,82	−35,75
	6	117,13	9,57
Regra 3/8 de Simpson	3	139,93	−8,04

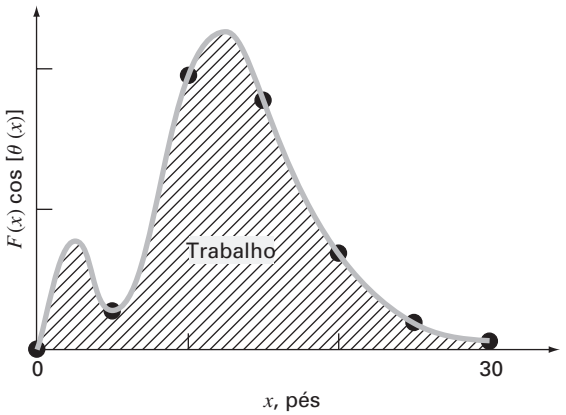


FIGURA 24.6
Um gráfico no contínuo de $F(x)$ cos $[\theta(x)]$ em função da posição com os sete pontos discretos usados para desenvolver estimativas por integração numérica na Tabela 24.7. Observe como o uso desses sete pontos para caracterizar essa função variando no contínuo perde os dois picos em $x = 2{,}5$ e 12,5 pés.

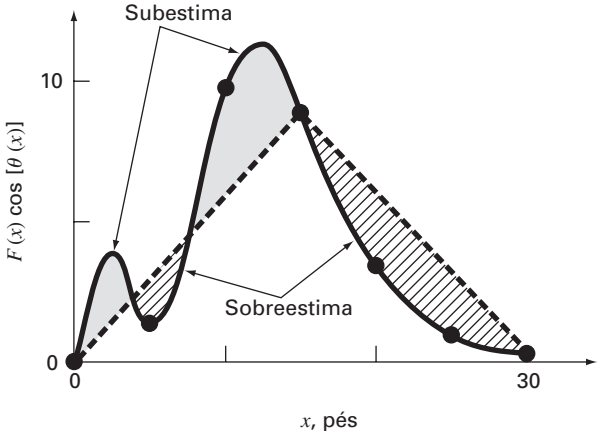
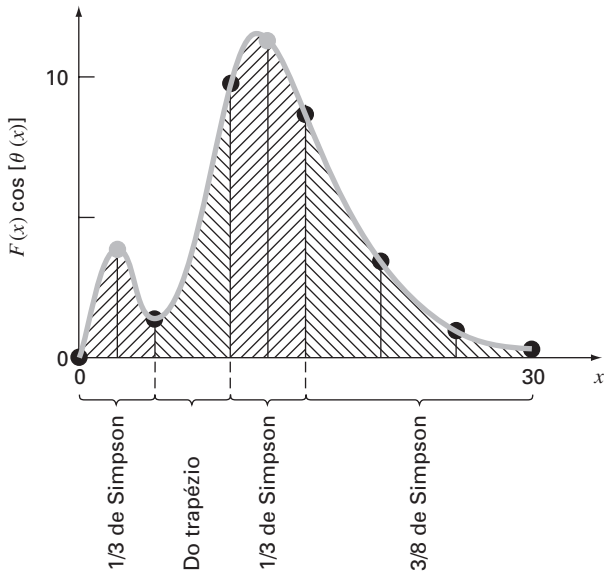


FIGURA 24.7
Descrição gráfica de por que a regra do trapézio com dois segmentos fornece uma boa estimativa da integral para esse caso. Por acaso, o uso de dois trapézios leva a um balanceamento entre erros positivos e negativos.

A conclusão a ser tirada da Figura 24.6 é que um número adequado de medidas deve ser feito para calcular integrais de forma acurada. No caso presente, se os dados estivessem disponíveis em $F(2,5)\cos[\theta(2,5)] = 4{,}3500$ e $F(12,5)\cos[\theta(12,5)] = 11{,}3600$, poderíamos determinar uma estimativa da integral usando o algoritmo para dados desigualmente espaçados descrito previamente na Seção 21.3. A Figura 24.8 ilustra a divisão em segmentos desiguais para esse caso. Incluindo esses dois pontos adicionais, obtemos uma estimativa melhorada da integral de 126,9 ($\varepsilon_t = 2{,}02\%$). Logo, a inclusão de dados adicionais incorporaria os picos que estavam faltando anteriormente e, como conseqüência, levaria a melhores resultados.

FIGURA 24.8
O esquema de divisão em segmentos desiguais que resulta da inclusão de dois pontos adicionais em $x = 2{,}5$ e 12,5 aos dados na Tabela 24.6. São indicadas as fórmulas de integração numérica aplicadas a cada conjunto de segmentos.



PROBLEMAS

Engenharia química/Bioengenharia

24.1 Faça os mesmos cálculos que na Seção 24.1, mas calcule a quantidade de calor necessária para elevar a temperatura de 1.200 g de material de −150 a 100 °C. Use a regra de Simpson para seu cálculo, com valores de T em incrementos de 50 °C.

24.2 Repita o Problema 24.1, mas use integração de Romberg até um erro de $\varepsilon_s = 0{,}01\%$.

24.3 Repita o Problema 24.1, mas use fórmulas de Gauss-Legendre para dois e três pontos. Interprete seus resultados.

24.4 A integração fornece um meio de calcular quanta massa entra em ou deixa um reator em um período especificado de tempo, como em

$$M = \int_{t_1}^{t_2} Qc\,dt$$

em que t_1 e t_2 são os instantes inicial e final, respectivamente. Essa fórmula faz sentido do ponto de vista intuitivo se você lembrar a analogia entre integral e soma. Logo, a integral representa a somatória do produto da vazão vezes a concentração para fornecer a massa total entrando ou saindo entre t_1 e t_2. Se a vazão for constante, Q pode ser movido para fora da integral:

$$M = Q\int_{t_1}^{t_2} c\,dt \qquad \text{(P24.4)}$$

Use integração numérica para calcular essa expressão para os dados listados a seguir. Observe que $Q = 4$ m^3/min.

t, min	0	10	20	30	35	40	45	50
c, mg/m^3	10	35	55	52	40	37	32	34

24.5 A concentração química do fluxo de saída para um reator de mistura completa é medida como

t, min	0	1	4	6	8	12	16	20
c, mg/m^3	12	22	32	45	58	75	70	48

Para um fluxo de saída $Q = 0{,}3$ m^3/s, obtenha uma estimativa da massa de composto químico, em gramas, que sai do reator entre $t = 0$ e 20 min.

24.6 A *primeira lei de difusão de Flick* afirma que

$$\text{Fluxo de massa} = -D\frac{dc}{dx} \qquad \text{(P24.6)}$$

em que o fluxo de massa é a quantidade de massa que passa por unidade de área por unidade de tempo (g/cm^2/s), D é o coeficiente de difusão (cm^2/s), c é a concentração e x é a distância (cm). Um engenheiro ambiental mede as seguintes concentrações de um poluente no sedimento no fundo de um lago ($x = 0$ na interface sedimento-água e aumentando para baixo):

x, cm	0	1	3
c, 10^{-6} g/cm^3	0,06	0,32	0,6

Use a melhor técnica de derivação disponível para obter uma estimativa da derivada em $x = 0$. Empregue essa estimativa em conjunto com a Equação (P24.6) para calcular o fluxo de massa do poluente saindo do sedimento e indo para a água no fundo do lago ($D = 1{,}52 \times 10^{-6}$ cm^2/s). Para um lago com $3{,}6 \times 10^6$ m^2 de sedimentos, quanto poluente seria transportado para o lago em um período de um ano?

24.7 Os seguintes dados foram coletados quando um grande tanque de óleo estava sendo carregado:

t, min	0	10	20	30	45	60	75
V, 10^6 barris	0,4	0,7	0,77	0,88	1,05	1,17	1,35

Calcule a vazão Q (isto é, dV/dt) para cada instante, com ordem de h^2.

24.8 Você está interessado em medir a velocidade de um fluido em um canal retangular estreito e aberto carregando dejetos de petróleo entre locais em uma refinaria de petróleo. Você sabe que, por causa do atrito no fundo, a velocidade varia com a profundidade do canal. Se os seus técnicos tiverem tempo de fazer apenas duas medidas de velocidade, em que profundidades você diria que eles fizessem essas medidas para obter a melhor estimativa da velocidade média? Faça sua recomendação em termos da percentagem da profundidade total d medida da superfície do fluido. Por exemplo, medir na superfície seria $0\%d$, enquanto no fundo seria $100\%d$.

24.9 O tecido mole segue um comportamento de deformação exponencial sob tensão uniaxial enquanto estiver no intervalo de elongação fisiológico ou normal. Isso pode ser expresso como

$$\sigma = \frac{E_o}{A}(e^{a\varepsilon} - 1)$$

em que σ é a tensão, ε é a deformação e E_o e a são constantes do material que são determinadas experimentalmente. Para calcular as duas constantes do material, a equação anterior é derivada com relação a ε, o que é uma relação fundamental para tecidos moles

$$\frac{d\sigma}{d\varepsilon} = E_o + a\sigma$$

Para calcular E_o e a, são usados dados de tensão-deformação para traçar o gráfico de $d\sigma/d\varepsilon$ em função de σ, e a inclinação e a intersecção com o eixo vertical desse gráfico são as duas constantes do material. A tabela contém dados de tensão-deformação para o *chordae tendineae* do coração (pequenos tendões usados para manter as válvulas do coração fechadas durante a contração do músculo cardíaco). Esses dados são da aplicação de carga ao tecido, enquanto curvas diferentes são produzidas quando as cargas são retiradas.

(a) Calcule a derivada $d\sigma/d\varepsilon$ usando diferenças finitas que sejam acuradas até segunda ordem. Marque os dados e elimine os pontos próximos do zero que parecem não seguir a relação linear. O erro nesses dados vem da inabilidade dos instrumentos de lerem pequenos valores nessa região. Faça uma análise de regressão dos dados remanescentes para determinar os valores de E_o e de a. Trace os dados de tensão em função da deformação junto com a curva analítica expressa pela primeira equação. Isso indicará quão bem a curva analítica ajusta os dados.

(b) Com freqüência, a análise anterior não funciona bem porque o valor de E_o é difícil de calcular. Para resolver esse problema, E_o não é usado. É escolhido um ponto dado $(\bar{\sigma}, \bar{\varepsilon})$ que esteja no meio do intervalo usado na análise de regressão. Esses valores são substituídos na primeira equação e um valor de E_o/a é determinado e substituído na primeira equação:

$$\sigma = \left(\frac{\bar{\sigma}}{e^{a\bar{\varepsilon}} - 1}\right)(e^{ax} - 1)$$

Usando essa abordagem, dados experimentais que sejam bem definidos produzirão uma boa concordância entre os dados e a curva analítica. Use essa nova relação e trace novamente os pontos dados de tensão em função dos pontos dados de deformação e a nova curva analítica.

24.10 A técnica padrão para determinar a saída cardíaca é o método do indicador de diluição desenvolvido por Hamilton. Uma extremidade de um pequeno cateter é inserida em uma artéria radial e a outra extremidade é conectada a um densitômetro, o qual registra automaticamente a concentração do corante no sangue. Uma quantidade conhecida de corante, 5,6 mg, é injetada rapidamente e os seguintes dados são obtidos:

Tempo, s	Concentração, mg/L	Tempo, s	Concentração, mg/L
5	0	21	2,3
7	0,1	23	1,1
9	0,11	25	0,9
11	0,4	27	1,75
13	4,1	29	2,06
15	9,1	31	2,25
17	8	33	2,32
19	4,2	35	2,43

Traçando esses dados, obtemos a curva de diluição do corante na Figura P24.10*a*. A concentração atinge um valor máximo em cerca de 15 segundos e a seguir decresce rapidamente e depois se eleva por causa da recirculação do corante. A curva é redesenhada em um gráfico semilog na Figura P24.10*b*. Observe que uma reta aproxima o ramo descendente da curva de diluição. Para separar o efeito de recirculação, os analistas estendem a parte reta. A saída cardíaca pode então ser calculada a partir de

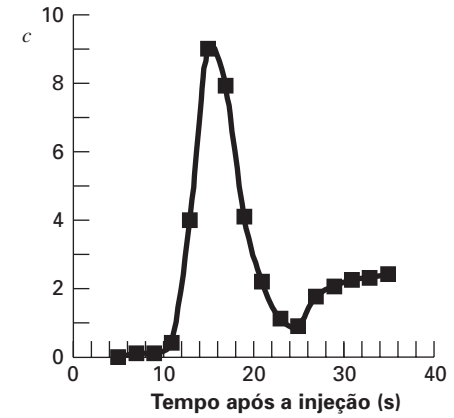
$$C = \frac{M}{A} \times 60 \text{ s/min}$$

em que C é a saída cardíaca [L/min], M é a quantidade de corante injetado (mg) e A é a área sob a curva com a correção linear. Calcule a saída cardíaca desse paciente usando a regra do trapézio com um tamanho de passo de 0,5 s.

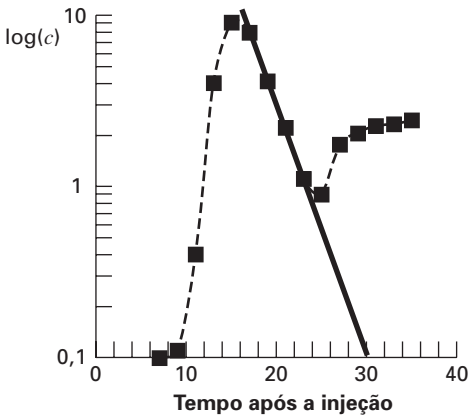
24.11 Glaucoma é a segunda causa principal da perda de visão em todo o mundo. Alta pressão intra-ocular (pressão dentro do olho) quase sempre acompanha a perda de visão. É postulado que a alta pressão danifica um subconjunto de células no olho que é responsável pela visão. Um pesquisador teorizou que a relação entre a perda de visão e a pressão pode ser descrita por

$$VL = A \exp\left(k \int_{25}^{t} (P - 13)\, dt\right)$$

$\sigma \times 10^3$ N/m^2	87,8	96,6	176	263	350	569	833	1.227	1.623	2.105	2.677	3.378	4.257
$\varepsilon \times 10^{-3}$ m/m	153	198	270	320	355	410	460	512	562	614	664	716	766



(*a*)



(*b*)

Figura P24.10

em que *VL* é a perda de visão porcentual, *P* é a pressão intra-ocular (mm Hg), *t* é o tempo (em anos) e *k* e *A* são constantes. Usando os dados a seguir de três pacientes, faça estimativas das constantes *k* e *A*.

Paciente	**A**		**B**		**C**	
Idade na ocasião do diagnóstico	65		43		80	
VL	60		40		30	
	Idade, anos	***P*, mm Hg**	**Idade, anos**	***P*, mm Hg**	**Idade, anos**	***P*, mm Hg**
	25	13	25	11	25	13
	40	15	40	30	40	14
	50	22	41	32	50	15
	60	23	42	33	60	17
	65	24	43	35	80	19

24.12 Uma de suas colegas inventou um emplastro transdérmico para administrar insulina através da pele a pacientes diabéticos de uma forma controlada, eliminando a necessidade de injeções dolorosas. Ela coletou os seguintes dados sobre o fluxo de massa de insulina sendo liberado através do emplastro (e da pele) como uma função do tempo:

Fluxo, mg/cm²/h	**Tempo, h**	**Fluxo, mg/cm²/h**	**Tempo, h**
15	0	8	5
14	1	5	10
12	2	2,5	15
11	3	2	20
9	4	1	24

Lembre-se de que o fluxo de massa é a vazão por uma área ou $(1/A)\,dm/dt$. Forneça a melhor estimativa que puder para a quantidade de droga liberada através da pele em 24 horas usando um emplastro de 12 cm^2.

24.13 A *videoangiografia* é usada para medir o escoamento do sangue e determinar a situação da função circulatória. Para quantificar as videoangiografias, o diâmetro do vaso sanguíneo e a velocidade do sangue são necessários para que o escoamento total de sangue seja determinado. A seguir se encontra um perfil densitométrico tomado de uma videoangiografia de um vaso sanguíneo. Uma forma de determinar onde está a borda do vaso sanguíneo a partir de um angiográfico é determinar onde a primeira derivada do perfil é um valor extremo. Usando os dados fornecidos, encontre a fronteira do vaso sanguíneo e faça uma estimativa do seu diâmetro. Use fórmulas de diferenças centradas tanto de $O(h^2)$ quando de $O(h^4)$ e compare os resultados.

Distância	**Densidade**	**Distância**	**Densidade**	**Distância**	**Densidade**	**Distância**	**Densidade**
0	26,013	28	38,273	56	39,124	84	37,331
4	26,955	32	39,103	60	38,813	88	35,980
8	26,351	36	39,025	64	38,925	92	31,936
12	28,343	40	39,432	68	38,804	96	28,843
16	31,100	44	39,163	72	38,806	100	26,309
20	34,667	48	38,920	76	38,666	104	26,146
24	37,251	52	38,631	80	38,658		

Engenharia Civil/Ambiental

24.14 Faça os mesmos cálculos que na Seção 24.2, mas use integração de Romberg de $O(h^8)$ para calcular a integral.

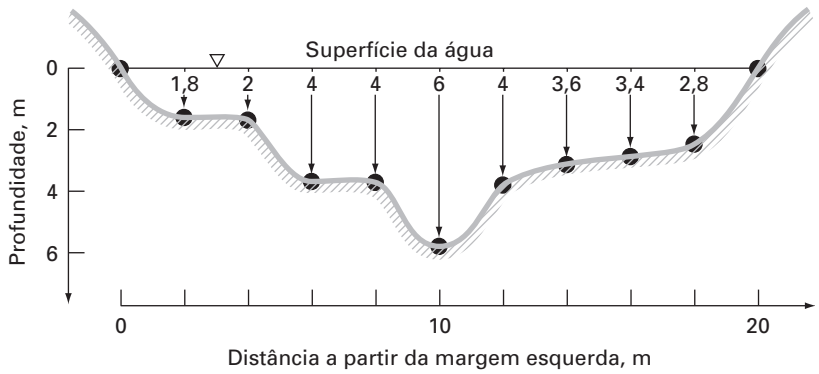


Figura P24.17
A seção transversal de um ribeirão.

24.15 Faça os mesmos cálculos que na Seção 24.2, mas use quadratura de Gauss para calcular a integral.

24.16 Como na Seção 24.2, calcule F usando a regra do trapézio e as regras 1/3 e 3/8 de Simpson, mas use a força a seguir. Divida o mastro em intervalos de 5 pés.

$$F = \int_0^{30} \frac{250z}{6+z} e^{-z/10}\, dz$$

24.17 As áreas das seções transversais (A) são necessárias para diversas tarefas em engenharia de recursos hídricos, incluindo previsão de enchentes e projetos de represas. A menos que estejam disponíveis aparelhos sonoros eletrônicos para obter perfis contínuos do fundo do canal, o engenheiro deve basear-se em medidas de profundidade discretas para calcular A. Um exemplo de uma seção transversal típica de um ribeirão é mostrado na Figura P24.17. Os pontos dados representam posições nas quais um barco era ancorado e a leitura da profundidade tomada. Use duas aplicações da regra do trapézio ($h = 4$ e 2 m) e a regra 1/3 de Simpson ($h = 2$ m) para obter uma estimativa da área da seção transversal a partir desses dados.

24.18 Como descrito no Problema 24.17, a área da seção transversal de um canal pode ser calculada por

$$A_c = \int_0^B H(y)\, dy$$

em que B é a largura total do canal (m), H é a profundidade (m) e y é a distância da margem (m). De uma forma parecida, a vazão média Q (m^3/s) pode ser calculada por

$$Q = \int_0^B U(y)H(y)\, dy$$

em que U é a velocidade da água (m/s). Use essas relações e um método numérico para determinar A_c e Q para os seguintes dados:

y, m	0	2	4	5	6	9
H, m	0,5	1,3	1,25	1,7	1	0,25
U, m/s	0,03	0,06	0,05	0,12	0,11	0,02

24.19 Durante um exame, foi pedido que você calculasse a área do campo mostrado na Figura P24.19. Use as regras de Simpson para determinar a área.

24.20 Um estudo em engenharia de transporte exige o cálculo do número total de carros que passa por uma intersecção em um período de 24 h. Um indivíduo visita a intersecção em diversos instantes durante o dia e conta o número de carros que passam pela intersecção em um minuto. Utilize os dados resumidos na Tabela P24.20 para obter uma estimativa do número total de carros que passam pela intersecção por dia. (Tome cuidado com as unidades.)

24.21 A força do vento distribuída contra o lado de um edifício é medida como sendo

Altura, l, m	0	30	60	90	120	150	180	210	240
Força, $F(l)$, N/m	0	340	1.200	1.600	2.700	3.100	3.200	3.500	3.800

Calcule a força líquida e a reta da ação que resulta dessa distribuição do vento.

24.22 A água exerce pressão na face a montante de um dique como mostrado na Figura P24.22. A pressão pode ser caracterizada por

$$p(z) = \rho g(D - z) \qquad \text{(P24.22)}$$

em que $p(z)$ é a pressão em pascais (ou N/m^2) exercida a uma elevação de z metros acima do fundo da represa; ρ é a densidade da água, a qual nesse problema supõe-se ser constante e igual a 10^3 kg/m^3; g é a aceleração da gravidade (9,8 m/s^2); e D é a elevação (em m) da superfície da água acima do fundo da represa. De acordo com a Equação (P24.22), a pressão aumenta linearmente com a profundidade, como descrito na Figura P24.22*a*. Omitindo a pressão atmosférica (pois ela age contra os dois lados do dique e essencialmente se cancela), a força total f_t pode ser determinada multiplicando-se a pressão pela área do lado do dique (como mostrado na Figura P24.22*b*). Como tanto a pressão quanto a área variam com o nível de profundidade, a força total é obtida calculando

$$f_t = \int_0^D \rho g w(z)(D - z)\, dz$$

em que $w(z)$ é igual à largura do dique (m) na profundidade z (Figura P24.22*b*). A reta de ação também pode ser determinada pelo cálculo de

$$d = \frac{\displaystyle\int_0^D \rho g z w(z)(D - z)\, dz}{\displaystyle\int_0^D \rho g w(z)(D - z)\, dz}$$

Use a regra de Simpson para determinar f_t e d. Verifique seus resultados usando seu programa de computador para a regra do trapézio.

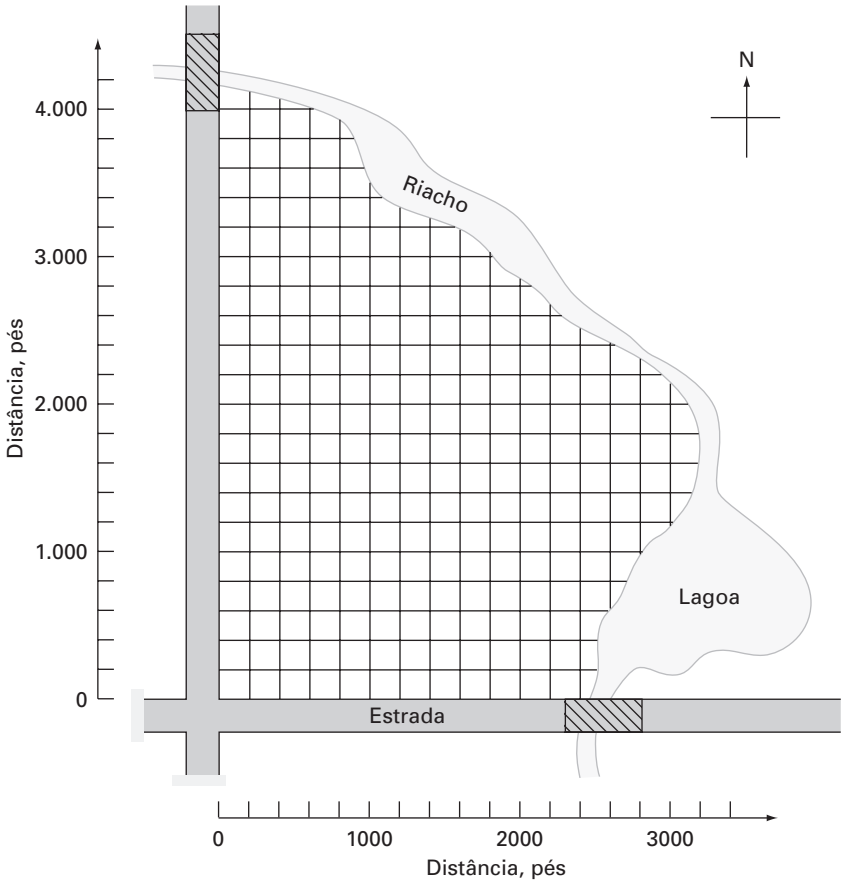
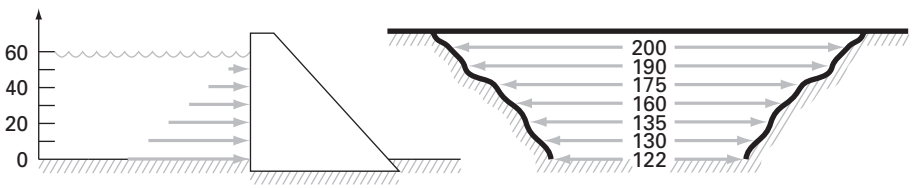


Figura P24.19
Um campo limitado por duas estradas e um riacho.

TABELA P24.20 Taxa de fluxo de tráfego (carros/min) em uma intersecção medida em diversos instantes dentro de um período de 24 horas.

Tempo	Taxa	Tempo	Taxa	Tempo	Taxa
12:00 (meia-noite)	2	9:00 A.M.	11	6:00 P.M.	20
2:00 A.M.	2	10:30 A.M.	4	7:00 P.M.	10
4:00 A.M.	0	11:30 A.M.	11	8:00 P.M.	8
5:00 A.M.	2	12:30 P.M.	12	9:00 P.M.	10
6:00 A.M.	6	2:00 P.M.	8	10:00 P.M.	8
7:00 A.M.	7	4:00 P.M.	7	11:00 P.M.	7
8:00 A.M.	23	5:00 P.M.	26	12:00 (meia-noite)	3

Figura P24.22
Água exercendo pressão na face a montante de um dique: (*a*) visão lateral mostrando a força linearmente crescente com a profundidade; (*b*) visão frontal mostrando a largura da represa em metros.



(*a*) (*b*)

24.23 Para obter uma estimativa do tamanho de uma nova represa, você deve determinar o volume total de água (m^3) que escoa por um rio no período de um ano. Estão disponíveis os seguintes dados médios a longo prazo para o rio:

Dados em meados de	Jan.	Fev.	Mar.	Abr.	Jun.	Set.	Out.	Nov.	Dez.
Vazão, m^3/s	30	38	82	125	95	20	22	24	35

Determine o volume. Tome cuidado com as unidades e para fazer estimativas corretas nas extremidades.

24.24 Os dados listados na tabela a seguir dão medidas horárias do fluxo de calor q (cal/cm^2/h) na superfície de um coletor solar. Como um engenheiro arquitônico, você deve obter uma estimativa do calor total absorvido por um painel coletor de 150.000 cm^2 durante um período de 14 horas. O painel tem uma eficiência de absorção e_{ab} de 45%. O calor total absorvido é dado por

$$h = e_{ab} \int_0^t q A \, dt$$

em que A é a área e q é o fluxo de calor.

t	0	2	4	6	8	10	12	14
q	0,10	5,32	7,80	8,00	8,03	6,27	3,54	0,20

24.25 O fluxo de calor q é a quantidade de calor fluindo por uma unidade de área de um material por unidade de tempo. Ele pode ser calculado com a lei de Fourier,

$$J = -k \frac{dT}{dx}$$

em que J tem unidades de J/m^2/s ou W/m^2 e k é um coeficiente de condutividade térmica que parametriza as propriedades de condução de calor do material e tem unidades de W/(°C · m). T é a temperatura (°C) e x é a distância (m) ao longo da trajetória do fluxo de calor. A lei de Fourier é usada rotineiramente pelos engenheiros arquitetônicos para determinar o fluxo de calor através de paredes. As seguintes temperaturas foram medidas a partir da superfície ($x = 0$) de uma parede de pedra:

x, m	0	0,08	0,16
T, °C	20	17	15

Se o fluxo em $x = 0$ for 60 W/m^2, calcule k.

24.26 A área da superfície horizontal A_s (m^2) de um lago a uma profundidade particular pode ser calculada a partir do volume por derivação,

$$A_s(z) = -\frac{dV}{dz}(z)$$

em que V é o volume (m^3) e z é a profundidade (m) medida da superfície até o fundo. A concentração média de uma substância que varia com a profundidade $\bar{c}$ (g/m^3) pode ser calculada por integração

$$\bar{c} = \frac{\int_0^Z c(z) A_s(z) \, dz}{\int_0^Z A_s(z) \, dz}$$

em que Z é a profundidade total (m). Calcule a concentração média com base nos seguintes dados:

z, m	0	4	8	12	16
V, 10^6 m^3	9,8175	5,1051	1,9635	0,3927	0,0000
c, g/m^3	10,2	8,5	7,4	5,2	4,1

Engenharia Elétrica

24.27 Faça os mesmos cálculos que na Seção 24.3, mas para a corrente especificada por

$$i(t) = 5e^{-1,25t} \operatorname{sen} 2\pi t \qquad \text{para } 0 \le t \le T/2$$
$$i(t) = 0 \qquad \text{para } T/2 < t \le T$$

em que $T = 1$ s. Use a quadratura de Gauss com cinco pontos para obter uma estimativa da integral.

24.28 Repita o Problema 24.27, mas use a regra 1/3 de Simpson com cinco segmentos.

24.29 Repita o Problema 24.27, mas use integração de Romberg até $\varepsilon_s = 1\%$.

24.30 A lei de Faraday caracteriza a queda de voltagem através de um indutor como

$$V_L = L \frac{di}{dt}$$

em que V_L é a queda de voltagem(V), L é a indutância (em henrys; 1 H = 1 V · s/A), i é a corrente (A) e t é o tempo (s). Determine a queda de voltagem como função do tempo a partir dos seguintes dados para uma indutância de 4 H.

t	0	0,1	0,2	0,3	0,5	0,7
i	0	0,16	0,32	0,56	0,84	2,0

24.31 Com base na lei de Faraday (Problema 24.30), use os seguintes dados de voltagem para obter uma estimativa da indutância em henrys se uma corrente de 2 A passar pelo indutor por um período de 400 milissegundos.

t, ms	0	10	20	40	60	80	120	180	280	400
V, volts	0	18	29	44	49	46	35	26	15	7

24.32 Suponha que a corrente através de um resistor é descrita pela função

$$i(t) = (60 - t)^2 + (60 - t) \operatorname{sen}(\sqrt{t})$$

e a resistência é uma função da corrente,

$$R = 12i + 2i^{2/3}$$

Calcule a voltagem média entre $t = 0$ e 60 usando a regra 1/3 de Simpson para segmentos múltiplos.

24.33 Se um capacitor inicialmente não tem nenhuma carga, a voltagem através dele como uma função do tempo pode ser calculada por

$$V(t) = \frac{1}{C} \int_0^t i(t) \, dt$$

Se $C = 10^{-5}$ farad, use os seguintes dados de corrente para desenvolver um gráfico da voltagem em função do tempo:

t, s	0	0,2	0,4	0,6	0,8	1	1,2
i, 10^{-3} A	0,2	0,3683	0,3819	0,2282	0,0486	0,0082	0,1441

Engenharia Mecânica/Aeroespacial

24.34 Faça os mesmos cálculos que na Seção 24.4, mas use a seguinte equação:

$$F(x) = 1,6x - 0,045x^2$$

Use os valores de θ da Tabela 24.6.

24.35 Faça os mesmos cálculos que na Seção 24.4, mas use a seguinte equação:

$$\theta(x) = 0,8 + 0,125x - 0,009x^2 + 0,0002x^3$$

Use a equação do Problema 24.34 para $F(x)$. Use regras do trapézio com 4, 8 e 16 segmentos para calcular a integral.

24.36 Repita o Problema 24.35, mas use a regra 1/3 de Simpson.

24.37 Repita o Problema 24.35, mas use integração de Romberg até $\varepsilon_s = 0,5\%$.

24.38 Repita o Problema 24.35, mas use a quadratura de Gauss.

24.39 O trabalho feito em um objeto é igual à força vezes a distância percorrida na direção da força. A velocidade de um objeto na direção da força é dada por

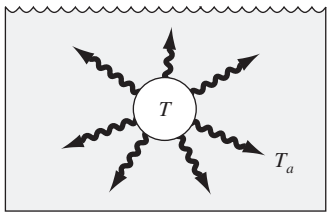
$$v = 4t \qquad 0 \le t \le 4$$
$$v = 16 + (4 - t)^2 \qquad 4 \le t \le 14$$

em que v está em m/s. Use a regra de Simpson com aplicações múltiplas para determinar o trabalho se uma força constante de 200 N for aplicada para todos os instantes t.

24.40 A taxa de resfriamento de um corpo (Figura P24.40) pode ser expressa por

$$\frac{dT}{dt} = -k(T - T_a)$$

Figura P24.40



em que T é a temperatura do corpo (°C), T_a é a temperatura do meio ao redor (°C) e k é uma constante de proporcionalidade (por minuto). Logo, essa equação (chamada de *lei de resfriamento de Newton*) especifica que a taxa de resfriamento é proporcional à diferença entre a temperatura do corpo e a do meio ao redor. Se uma bola de metal aquecida a 80 °C for jogada na água à temperatura mantida constante em $T_a = 20$ °C, a temperatura da bola muda como em

Tempo, min	0	5	10	15	20	25
T, °C	80	44,5	30,0	24,1	21,7	20,7

Utilize derivação numérica para determinar dT/dt em cada instante de tempo. Trace dT/dt em função de $T - T_a$ e empregue regressão linear para calcular k.

24.41 Uma haste sujeita a uma carga axial (Figura P24.41*a*) será deformada, como mostrado na curva de tensão-deformação na Figura P24.41*b*. A área sob a curva da tensão zero até o ponto de ruptura é chamada de *módulo de rigidez* do material. Ele fornece uma medida da energia por unidade de volume necessário para causar a ruptura do material. Assim, ele é representativo da habilidade de suportar uma carga de impacto. Use integração numérica para calcular o módulo de rigidez para a curva de tensão-deformação vista na Figura P24.41*b*.

24.42 Se a distribuição de um fluido escoando por um tubo for conhecida (Figura P24.42), a vazão Q (isto é, o volume de água passando pelo tubo por unidade de tempo) pode ser calculada por $Q = \int v\,dA$, em que v é a velocidade e A é a área da seção transversal do tubo. (Para perceber o significado dessa relação fisicamente, lembre-se da conexão íntima entre soma e integração.) Para um tubo circular, $A = \pi r^2$ e $dA = 2\pi r\,dr$. Portanto,

$$Q = \int_0^r v(2\pi r)\,dr$$

em que r é a distância radial medida a partir do centro do tubo. Se a distribuição de velocidade for dada por

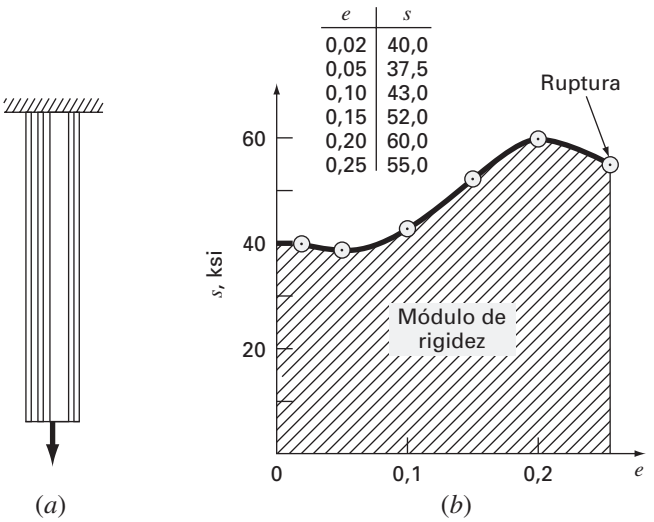
$$v = 2\left(1 - \frac{r}{r_0}\right)^{1/6}$$

em que r_0 é o raio total (nesse caso, 3 cm), calcule Q usando a regra do trapézio com aplicações múltiplas. Discuta o resultado.

24.43 Usando os seguintes dados, calcule o trabalho feito ao se esticar uma mola que tem constante de mola $k = 300$ N/m até $x = 0{,}35$ m:

F, 10^3N	0	0,01	0,028	0,046	0,063	0,082	0,11	0,13
x, m	0	0,05	0,10	0,15	0,20	0,25	0,30	0,35

Figura P24.41
(*a*) Uma haste sob carga axial e (*b*) a curva de tensão-deformação em que a tensão está em *kips* por polegada quadrada (10^3 lb/pol^2) e a deformação é adimensional.



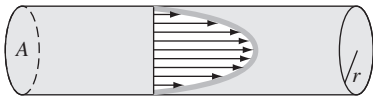


Figura P24.42

24.44 A posição de ataque de um jato em uma pista de aterrissagem de aviões de transporte foi cronometrada durante a aterrissagem:

t, s	0	0,52	1,04	1,75	2,37	3,25	3,83
x, m	153	185	210	249	261	271	273

em que x é a distância a partir da extremidade do avião de carga. Estime (**a**) a velocidade (dx/dt) e (**b**) a aceleração (dv/dt) usando derivação numérica.

24.45 Empregue a regra de Simpson com aplicação múltipla para calcular a distância vertical percorrida por um foguete se a velocidade vertical for dada por

$$\begin{aligned} v &= 11t^2 - 5t & 0 \le t \le 10 \\ v &= 1100 - 5t & 10 \le t \le 20 \\ v &= 50t + 2(t-20)^2 & 20 \le t \le 30 \end{aligned}$$

24.46 A velocidade para cima de um foguete pode ser calculada seguindo a fórmula:

$$v = u \ln\left(\frac{m_0}{m_0 - qt}\right) - gt$$

em que v é a velocidade para cima, u é a velocidade relativa ao foguete na qual o fluido é expelido, m_0 é a massa inicial do foguete no instante $t = 0$, q é a taxa de consumo de combustível e g é a aceleração para baixo da gravidade (suposta constante e igual a 9,8 m/s^2). Se $u = 1.800$ m/s, $m_0 = 160.000$ kg e $q = 2.500$ kg/s, use a regra do trapézio com seis segmentos e a regra 1/3 de Simpson, a quadratura de Gauss com seis pontos e métodos de Romberg de $O(h^8)$ para determinar quão alto o foguete irá voar em 30 s.

24.47 Com relação aos dados do Problema 20.57, encontre a taxa de tensão usando métodos de diferenças finitas. Use aproximações de derivadas acuradas até primeira ordem e trace o gráfico de seus resultados. Olhando para o gráfico, é fácil perceber que existe algum erro experimental inicial. Encontre a média e o desvio-padrão da taxa de tensão depois de eliminar os dados representando o erro experimental inicial.

24.48 O escoamento completamente desenvolvido, movendo-se por um tubo de 40 cm de diâmetro, tem um perfil de velocidade ilustrado na tabela abaixo. Encontre a vazão em volume Q usando a relação $Q = \int_0^R 2\pi r v\, dr$, em que r é a distância ao eixo axial do tubo, R é o raio do tubo e v é a velocidade. Resolva o problema usando duas abordagens diferentes.

(a) Ajuste uma curva polinomial aos dados de velocidade e integre analiticamente.
(b) Use a regra 1/3 de Simpson com aplicações múltiplas para integrar.
(c) Encontre o erro percentual usando a integral do ajuste polinomial como o valor mais correto.

Raio, *r*, cm	0,0	2,5	5,0	7,5	10,0	12,5	15,0	17,5	20,0
Velocidade, *v*, m/s	0,914	0,890	0,847	0,795	0,719	0,543	0,427	0,204	0

24.49 O escoamento completamente desenvolvido de um fluido plástico de Bingham se movendo em um tubo de 12 pol de diâmetro tem o perfil de velocidade dado. O escoamento de um fluido de Bingham não sofre cisalhamento no núcleo central, produzindo escoamento plug na região em torno do eixo central.

Raio, *r*, pol	0	1	2	3	4	5	6
Velocidade, *v*, pés/s	5,00	5,00	4,62	4,01	3,42	1,69	0,00

Encontre a vazão em volume Q usando a relação $Q = \int_{r_1}^{r_2} 2\pi r v\, dr + v_c A_c$, em que r é a distância ao eixo axial do tubo, R é o raio do tubo, v é a velocidade, v_c é a velocidade no centro e A_c é a área da seção transversal do plug. Resolva o problema usando duas abordagens diferentes.

(a) Ajuste uma curva polinomial aos dados fora do núcleo e integre analiticamente.
(b) Use a regra 1/3 de Simpson com aplicações múltiplas para integrar.
(c) Encontre o erro porcentual usando a integral do ajuste polinomial como o valor mais correto.

24.50 A entalpia de um gás real é uma função da pressão como descrito a seguir. Os dados foram tomados para um fluido real. Faça uma estimativa da entalpia do fluido a 400 K e a 50 atm (calcule a integral de 0 atm a 50 atm).

$$H \int_0^P \left(V - T\left(\frac{\partial V}{\partial T}\right)_P\right) dP$$

	***V*, L**		
***P*, atm**	***T* = 350 K**	***T* = 400 K**	***T* = 450 K**
0,1	220	250	282,5
5	4,1	4,7	5,23
10	2,2	2,5	2,7
20	1,35	1,49	1,55
25	1,1	1,2	1,24
30	0,90	0,99	1,03
40	0,68	0,75	0,78
45	0,61	0,675	0,7
50	0,54	0,6	0,62

24.51 Fornecidos os dados a seguir, encontre o trabalho isotérmico feito no gás quando ele é comprimido de 23 L até 3 L (lembre-se de que $W = -\int_{V_1}^{V_2} P\, dV$).

V, L	3	8	13	18	23
P, atm	12,5	3,5	1,8	1,4	1,2

(a) Calcule o trabalho feito sobre o gás numericamente, usando a regra do trapézio com 1, 2 e 4 segmentos.
(b) Calcule as razões dos erros nessas estimativas e relacione-as à análise de erro da regra do trapézio com aplicações múltiplas discutida no Capítulo 21.

24.52 A Equação de Rosin-Rammler-Bennet (RRB) é usada para descrever a distribuição de tamanho na poeira fina. $F(x)$ representa a massa acumulada das partículas de diâmetro inferior ou igual a x. x' e n' são constantes iguais a 30 μm e 1,44, respectivamente. A distribuição de densidade de massa $f(x)$ ou a massa das partículas

de poeira de diâmetro x é encontrada tomando-se a derivada da distribuição acumulada

$$F(x) = 1 - e^{-(x/x')^{n'}} \qquad f(x) = \frac{dF(x)}{dx}$$

(a) Calcule numericamente a distribuição de densidade de massa $f(x)$ e trace tanto $f(x)$ quanto a distribuição acumulada $F(x)$.
(b) Usando seus resultados da parte (**a**), calcule o tamanho modal da distribuição de densidade de massa — isto é, o tamanho no qual a derivada de $f(x)$ é igual a zero.
(c) Encontre a área da superfície por massa de poeira S_m (cm^2/g) usando

$$S_m = \frac{6}{\rho}\int_{d_{\min}}^{\infty} \frac{f(x)}{x}\,dx$$

A equação é válida apenas para partículas esféricas. Suponha uma densidade $\rho = 1$ g cm^{-3} e um diâmetro mínimo de poeira incluído na distribuição $d_{\min}$ de 1 μm.

24.53 Para o escoamento de fluido, o fluxo de calor para a superfície pode ser calculado por

$$J = -k\frac{dT}{dy}$$

em que J é o fluxo de calor (W/m^2), k é a condutividade térmica (W/m · K), T é a temperatura (K) e y é a distância normal à superfície (m). As seguintes medidas foram feitas para ar escoando sobre uma placa plana que tem 200 cm de comprimento e 50 cm de largura:

y, cm	0	1	3	5
T, K	900	480	270	200

Se $k = 0{,}028$ J/s · m · K, **(a)** determine o fluxo na superfície e (**b**) o calor transferido em watts. Observe que 1 J = 1 W · s.

24.54 O gradiente de pressão para o escoamento laminar através de um tubo de raio constante é dado por

$$\frac{dp}{dx} = -\frac{8\mu Q}{\pi r^4}$$

em que p é a pressão (N/m^2), x é a distância ao longo do eixo central do tubo (m), μ é a viscosidade dinâmica (N · s/m^2), Q é o escoamento (m^3/s) e r é o raio (m).

(a) Determine a queda de pressão para um tubo de 10 cm de comprimento para um líquido viscoso ($\mu = 0{,}005$ N · s/m^2, densidade $\rho = 1 \times 10^3$ kg/m^3) com um escoamento de 10×10^{-6} m^3/s e os seguintes raios variáveis ao longo de seu comprimento:

x, cm	0	2	4	5	6	7	10
r, mm	2	1,35	1,34	1,6	1,58	1,42	2

(b) Compare seu resultado com a queda de pressão que teria ocorrido se o tubo tivesse um raio constante igual ao raio médio.
(c) Determine o número de Reynolds médio para o tubo para verificar que o escoamento é realmente laminar (Re $= \rho v D/\mu <$ 2.100, em que v é a velocidade).

24.55 Dados de velocidade do ar foram coletados em diferentes raios do eixo central de um tubo circular de 16 cm de diâmetro, como tabulado a seguir:

r, cm	0	1,60	3,20	4,80	6,40	7,47	7,87	7,95	8
v, m/s	10	9,69	9,30	8,77	7,95	6,79	5,57	4,89	0

Use integração numérica para determinar a vazão em massa, que pode ser calculada por

$$\int_0^R \rho v 2\pi r\,dr$$

em que ρ é a densidade (1,2 kg/m^3). Expresse seu resultado em kg/s.



EPÍLOGO: PARTE SEIS

PT6.4 PRÓS E CONTRAS

A Tabela PT6.4 fornece um resumo dos prós e contras envolvidos na integração numérica ou quadratura. A maioria desses métodos é baseada na interpretação física simples de que uma integral é a área sob uma curva. Essas técnicas foram desenvolvidas para calcular a integral em dois casos diferentes: (1) uma função matemática (dada analiticamente) ou (2) dados discretos na forma tabular.

As fórmulas de Newton-Cotes são os métodos principais discutidos no Capítulo 21. Elas são aplicáveis a funções definidas tanto no discreto quanto no contínuo. Estão disponíveis as versões aberta e fechada dessas fórmulas. As formas abertas, que têm extremos de integração que se estendem além dos intervalos dos dados, raramente são usadas para o cálculo de integrais definidas. Entretanto, elas são úteis na solução de equações diferenciais ordinárias e no cálculo de integrais impróprias.

As fórmulas de Newton-Cotes fechadas são baseadas na substituição da função definida analiticamente ou dos dados tabulados por um polinômio interpolador que é fácil de integrar. A versão mais simples é a regra do trapézio, que é baseada em tomar a área abaixo de uma reta ligando valores adjacentes da função. Uma forma de melhorar a precisão da regra do trapézio é dividir o intervalo de integração de a a b em diversos segmentos e aplicar o método a cada segmento.

Além de aplicar a regra do trapézio com uma divisão em segmentos menores, outra forma de obter estimativas mais acuradas da integral é usar polinômios de grau mais alto para ligar os pontos. Se for usada uma equação quadrática, o resultado será a regra 1/3 de Simpson. Se for usada uma equação cúbica, o resultado será a regra 3/8 de Simpson. Como elas são muito mais precisas que a regra do trapézio, em geral, essas fórmulas são preferidas e estão disponíveis versões com aplicações múltiplas. Para situações com um número par de segmentos, é recomendada a aplicação múltipla da regra 1/3. Para um número ímpar de segmentos, a regra 3/8 pode ser aplicada aos últimos três segmentos e a regra 1/3 aos segmentos restantes.

TABELA PT6.4 Comparação das características dos métodos alternativos para integração numérica. As comparações são baseadas na experiência geral e não representam o comportamento de funções especiais.

Método	Pontos Dados Necessários para uma Aplicação	Pontos Dados Necessários para *n* Aplicações	Erro de Truncamento	Aplicação	Esforço de Programação	Comentário
Regra do trapézio	2	$n+1$	$\simeq h^3 f''(\xi)$	Ampla	Fácil	
Regra 1/3 de Simpson	3	$2n+1$	$\simeq h^5 f^{(4)}(\xi)$	Ampla	Fácil	
Regra de Simpson (1/3 e 3/8)	3 ou 4	≥ 3	$\simeq h^5 f^{(4)}(\xi)$	Ampla	Fácil	
Newton-Cotes de ordem mais alta	≥ 5	N/A	$\simeq h^7 f^{(6)}(\xi)$	Rara	Fácil	
Integração de Romberg	3			Exige que $f(x)$ seja conhecida	Moderado	Inapropriado para dados tabulares
Quadratura de Gauss	≥ 2	N/A		Exige que $f(x)$ seja conhecida	Fácil	Inapropriado para dados tabulares

Estão disponíveis também fórmulas de Newton-Cotes de ordem superior. No entanto, elas raramente são usadas na prática. Quando for necessária uma acurácia alta, estão disponíveis as fórmulas de integração de Romberg e de quadratura de Gauss. Deve ser observado que tanto a integração de Romberg como a quadratura de Gauss têm utilidade prática apenas nos casos nos quais a função está disponível. Essas técnicas não são apropriadas para dados tabulados.

PT6.5 RELAÇÕES E FÓRMULAS IMPORTANTES

A Tabela PT6.5 resume informações importantes que foram apresentadas na Parte Seis. Essa tabela pode ser consultada para acesso rápido às relações e às fórmulas importantes.

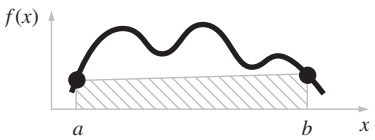
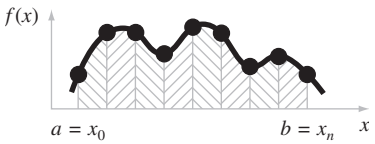
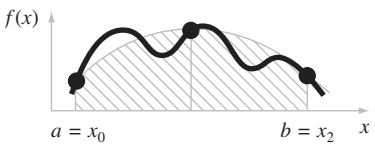
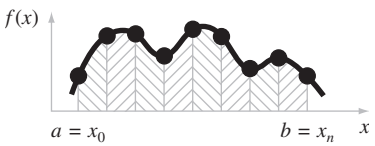
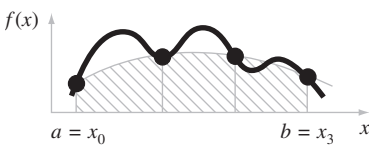
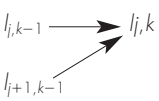
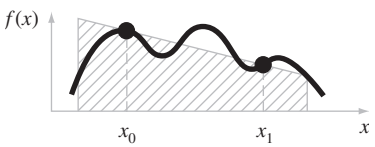
PT6.6 MÉTODOS AVANÇADOS E REFERÊNCIAS ADICIONAIS

Embora tenhamos revisado diversas técnicas de integração numérica, existem outros métodos que têm utilidade na prática da engenharia. Por exemplo, a *integração de Simpson adaptativa* é baseada na divisão do intervalo de integração em uma série de subintervalos de largura h. Então, a regra 1/3 de Simpson é utilizada para calcular a integral em cada subintervalo dividindo o tamanho do passo por dois, de uma forma iterativa, isto é, com tamanho de passo h, $h/2$, $h/4$, $h/8$, e assim por diante. Continuam-se as iterações em cada subintervalo até que uma estimativa de erro aproximada caia abaixo de um critério de parada pré-especificado ε_s. A seguir, a integral total é calculada como a somatória das estimativas da integral nos subintervalos. Essa técnica é especialmente útil para funções complicadas que têm regiões exibindo tanto pequenas quanto grandes ordens de variação. Discussões sobre integração adaptativa podem ser encontradas em Gerald e Wheatley (1989) e em Rice (1983). Além disso, os esquemas adaptativos para a resolução de equações diferenciais ordinárias podem ser usados para calcular integrais complicadas, como mencionado em PT6.1 e como será discutido no Capítulo 25.

Outro método de obter integrais é ajustar um *spline cúbico* aos dados. As equações cúbicas resultantes podem ser facilmente integradas (Forsythe et al., 1977). Uma abordagem parecida às vezes é usada também para a derivação. Finalmente, além das fórmulas de Gauss-Legendre, discutidas na Seção 22.3, existem também diversas outras fórmulas de quadratura. Carnahan, Luther e Wilkes (1969) e Ralston e Rabinowitz (1978) resumem muitas dessas abordagens.

Em síntese, a intenção do material anterior é lhe fornecer avenidas para a exploração mais profunda do assunto. Além do mais, todas as referências anteriores descrevem as técnicas básicas cobertas na Parte Seis. Sugerimos que você consulte essas fontes alternativas para ampliar seu entendimento dos métodos numéricos para integração.

TABELA PT6.5 Resumo das informações importantes apresentadas na Parte Seis.

Método	Formulação	Interpretação Gráfica	Erro
Regra do trapézio	$I \simeq (b-a)\dfrac{f(a)+f(b)}{2}$		$-\dfrac{(b-a)^3}{12}f''(\xi)$
Aplicação múltipla da regra do trapézio	$I \simeq (b-a)\dfrac{f(x_0)+2\sum_{i=1}^{n-1} f(x_i)+f(x_n)}{2n}$		$-\dfrac{(b-a)^3}{12n^2}\bar{f}''$
Regra 1/3 de Simpson	$I \simeq (b-a)\dfrac{f(x_0)+4f(x_1)+f(x_2)}{6}$		$-\dfrac{(b-a)^5}{2880}f^{(4)}(\xi)$
Aplicação múltipla da Regra 1/3 de Simpson	$I \simeq (b-a)\dfrac{f(x_0)+4\sum_{i=1,3}^{n-1} f(x_i)+2\sum_{j=2,4}^{n-2} f(x_j)+f(x_n)}{3n}$		$-\dfrac{(b-a)^5}{180n^4}f^{(4)}$
Regra 3/8 de Simpson	$I \simeq (b-a)\dfrac{f(x_0)+3f(x_1)+3f(x_2)+f(x_3)}{8}$		$-\dfrac{(b-a)^5}{6480}f^{(4)}(\xi)$
Integração de Romberg	$I_{j,k} = \dfrac{4^{k-1}I_{j+1,k-1}-I_{j,k-1}}{4^{k-1}-1}$		$O(h^{2k})$
Quadratura de Gauss	$I \simeq c_0 f(x_0)+c_1 f(x_1)+\cdots+c_{n-1}f(x_{n-1})$		$\simeq f^{(2n+2)}(\xi)$

PARTE SETE



EQUAÇÕES DIFERENCIAIS ORDINÁRIAS

PT7.1 MOTIVAÇÃO

No primeiro capítulo deste livro, deduzimos a seguinte equação baseada na segunda lei de Newton para calcular a velocidade v de um pára-quedista em queda livre como função do tempo [lembre-se da Equação (1.9)]:

$$\frac{dv}{dt} = g - \frac{c}{m}v \tag{PT7.1}$$

em que g é a constante gravitacional, m é a massa e c é o coeficiente de arrasto. Tais equações, que são compostas por uma função desconhecida e suas derivadas, são chamadas *equações diferenciais*. A Equação (PT7.1) é, algumas vezes, denominada *equação de taxa*, porque expressa a taxa de variação de uma variável como uma função das variáveis e parâmetros. Essas equações desempenham papel fundamental na engenharia, já que muitos fenômenos físicos são mais bem formulados matematicamente em termos de sua taxa de variação.

Na Equação (PT7.1), a quantidade que está sendo derivada, v, é chamada *variável dependente*. A quantidade com relação à qual v é derivada, t, é denominada *variável independente*. Quando a função envolve uma única variável independente, a equação é conhecida como *equação diferencial ordinária* (ou EDO). Isso para diferenciá-la de uma *equação diferencial parcial* (ou EDP), que envolve duas ou mais variáveis independentes.

As equações diferenciais também são classificadas em relação à sua ordem. Por exemplo, a Equação (PT7.1) é denominada *equação de primeira ordem*, pois a derivada mais alta é a primeira derivada. Uma *equação de segunda ordem* incluiria uma segunda derivada. Por exemplo, a equação que descreve a posição x de um sistema massa-mola com amortecimento é a equação de segunda ordem (lembre-se da Seção 8.4),

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = 0 \tag{PT7.2}$$

em que c é uma constante de amortecimento e k é a constante da mola. Analogamente, uma equação de ordem n inclui uma derivada n-ésima.

Equações de ordem superior podem ser reduzidas a um sistema de equações de primeira ordem. Para a Equação (PT7.2), isso é feito definindo-se uma nova variável y, por

$$y = \frac{dx}{dt} \tag{PT7.3}$$

a qual pode ser derivada para fornecer

$$\frac{dy}{dt} = \frac{d^2x}{dt^2} \tag{PT7.4}$$

As Equações (PT7.3) e (PT7.4) podem, então, ser substituídas na Equação (PT7.2) para fornecer

$$m\frac{dy}{dt} + cy + kx = 0 \tag{PT7.5}$$

ou

$$\frac{dy}{dt} = -\frac{cy + kx}{m} \tag{PT7.6}$$

Assim, as Equações (PT7.3) e (PT7.6) são um par de equações de primeira ordem que são equivalentes à equação original de segunda ordem. Como as outras equações diferenciais de ordem n podem ser reduzidas de forma análoga, esta parte do livro se concentra na solução de equações de primeira ordem. Algumas das aplicações em engenharia do Capítulo 28 tratam da solução de EDOs de segunda ordem por redução a um par de equações de primeira ordem.

PT7.1.1 Métodos Não Computacionais para Resolução de EDOs

Sem os computadores, as EDOs, em geral, são resolvidas com técnicas de integração analíticas. Por exemplo, a Equação (PT7.1) poderia ser multiplicada por dt e integrada para fornecer

$$v = \int \left(g - \frac{c}{m}v\right) dt \tag{PT7.7}$$

O lado direito dessa equação é chamado *integral indefinida*, porque os extremos de integração não são especificados, em contraste com as integrais definidas discutidas anteriormente, na Parte Seis [compare a Equação (PT7.7) com a Equação (PT6.6)].

Uma solução analítica da Equação (PT7.7) será obtida se a integral indefinida puder ser calculada exatamente na forma de uma equação. Por exemplo, lembre-se de que, para o problema do pára-quedista em queda livre, a Equação (PT7.7) foi resolvida analiticamente pela Equação (1.10) (supondo $v = 0$ em $t = 0$):

$$v(t) = \frac{gm}{c}\left(1 - e^{-(c/m)t}\right) \tag{1.10}$$

A mecânica de dedução de cada solução analítica será discutida na Seção PT7.2. Por enquanto, o fato importante é que soluções exatas para muitas EDOs de importância prática não estão disponíveis. Como é verdade para a maioria das situações discutidas em outras partes deste livro, os métodos numéricos fornecem a única alternativa viável nesses casos. Como esses métodos numéricos, em geral, exigem computadores, os engenheiros na era pré-computador estavam um pouco limitados no escopo de suas pesquisas.

Um método muito importante que os engenheiros e os matemáticos aplicados desenvolveram para superar esse dilema foi a *linearização*. Uma equação diferencial ordinária linear é uma equação que tem a forma geral

$$a_n(x)y^{(n)} + \cdots + a_1(x)y' + a_0(x)y = f(x) \tag{PT7.8}$$

em que $y^{(n)}$ é a n-ésima derivada de y em relação a x e os a's e o f' são funções especificadas de x. Essa equação é chamada *linear*, porque não há produtos ou funções não-lineares da variável dependente y e de suas derivadas. A importância prática das EDOs lineares é que elas pode ser resolvida de forma analítica.[1] Em contraste, a maioria das equações não-lineares não pode ser resolvida exatamente. Assim, na era pré-computador, uma tática para resolver equações não-lineares era linearizá-las.

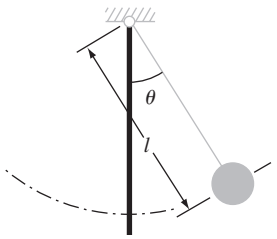


FIGURA PT7.1
O pêndulo oscilante.

Um exemplo simples é a aplicação das EDOs para prever o movimento de um pêndulo oscilante (Figura PT7.1). De maneira similar à dedução do problema do pára-quedista em queda livre, a segunda lei de Newton pode ser usada para deduzir a seguinte equação diferencial (ver Seção 28.4 para a dedução completa):

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\operatorname{sen}\theta = 0 \tag{PT7.9}$$

em que θ é o ângulo de deslocamento do pêndulo, g é a constante gravitacional e l é o comprimento do pêndulo. Essa equação é não-linear por causa do termo $\operatorname{sen}\theta$. Uma forma de obter uma solução analítica é perceber que para pequenos deslocamentos do pêndulo do equilíbrio (isto é, para pequenos valores de θ),

$$\operatorname{sen}\theta \cong \theta \tag{PT7.10}$$

[1] N.R.T: Na verdade, mesmo estas só podem ser resolvidas exatamente em situações particulares.

Logo, se, por hipótese, estivermos interessados apenas no caso em que θ for pequeno, a Equação (PT7.10) pode ser substituída na Equação (PT7.9) para fornecer

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\theta = 0 \tag{PT7.11}$$

Portanto, transformamos a Equação (PT7.9) em uma forma linear que é fácil de resolver analiticamente.

Embora a linearização continue sendo uma ferramenta muito valiosa na resolução de problemas de engenharia, existem casos em que ela não pode ser usada. Por exemplo, suponha que estejamos interessados no estudo do comportamento do pêndulo para grandes deslocamentos do equilíbrio. Em tais casos, os métodos numéricos oferecem uma opção viável para obter soluções. Atualmente, a grande disponibilidade de computadores coloca essa opção ao alcance de todos os engenheiros praticantes.

PT7.1.2 EDOs e a Prática da Engenharia

As leis fundamentais da física, matemática, eletrônica e termodinâmica, em geral, são baseadas em observações empíricas que explicam a variação das propriedades físicas e dos estados dos sistemas. Em vez de descrever os estados de sistemas físicos diretamente, as leis geralmente são dadas em termos de variações espaciais e temporais.

Diversos exemplos estão listados na Tabela PT7.1. Essas leis definem mecanismos de variação. Quando combinadas com as leis de continuidade da energia, da massa e do momento, produzem equações diferenciais. Integração subseqüente dessas equações diferenciais resulta em funções matemáticas que descrevem o estado espacial e temporal de um sistema em termos da variação de energia, massa ou velocidade.

O problema do pára-quedista em queda livre introduzido no Capítulo 1 é um exemplo de dedução de uma equação diferencial ordinária a partir da lei fundamental. Lembre-se de que a segunda lei de Newton era usada para deduzir uma EDO que descrevia a taxa de variação da velocidade de um pára-quedista em queda livre. Integrando essa relação, obtivemos uma equação para prever a velocidade de queda como uma função do tempo (Figura PT7.2). Essa equação poderia ser utilizada de diversas maneiras, inclusive com o propósito de projetar.

De fato, tais relações matemáticas são a base da solução para um grande número de problemas de engenharia. Entretanto, como descrito na seção anterior, muitas das equações diferenciais de significado prático não podem ser resolvidas usando os métodos analíticos do cálculo. Logo, os métodos discutidos nos capítulos a seguir são extremamente importantes em todos os campos da engenharia.

TABELA PT7.1 Exemplos de leis fundamentais que são escritas em termos de taxas de variação (t é o tempo e x, a posição).

Lei	Expressão Matemática	Variáveis e Parâmetros
Segunda lei de movimento de Newton	$\frac{dv}{dt} = \frac{F}{m}$	Velocidade (v), força (F) e massa (m)
Lei de Fourier do calor	$q = -k'\frac{dT}{dx}$	Fluxo de calor (q), condutividade térmica (k') e temperatura (T)
Lei de difusão de Fick	$J = -D\frac{dc}{dx}$	Fluxo de massa (J), coeficiente de difusão (D) e concentração (c)
Lei de Faraday (queda de voltagem em um indutor)	$\Delta V_L = L\frac{di}{dt}$	Queda de voltagem (ΔV_L), indutância (L) e corrente (i)

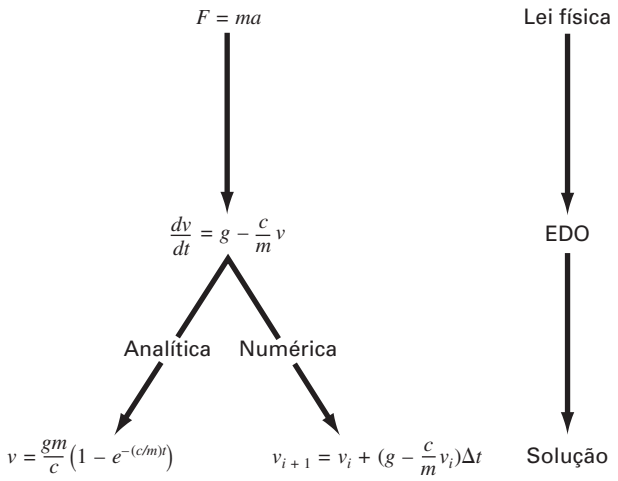


FIGURA PT7.2
A seqüência de eventos na aplicação de EDOs na resolução de problemas de engenharia. O exemplo mostrado é a velocidade de um pára-quedista em queda livre.

PT7.2 FUNDAMENTOS MATEMÁTICOS

Uma solução de uma equação diferencial ordinária é uma função específica da variável independente e dos parâmetros que satisfaz a equação diferencial original. Para ilustrar esse conceito, vamos começar com uma função dada

$$y = -0,5x^4 + 4x^3 - 10x^2 + 8,5x + 1 \tag{PT7.12}$$

que é um polinômio de quarto grau (Figura PT7.3*a*). Agora, se derivarmos a Equação (PT7.12), obtemos uma EDO:

$$\frac{dy}{dx} = -2x^3 + 12x^2 - 20x + 8,5 \tag{PT7.13}$$

Essa equação também descreve o comportamento do polinômio, mas de modo diferente da Equação (PT7.12). Em vez de descrever explicitamente os valores de y para cada valor de x, a Equação (PT7.13) dá a taxa de variação de y com relação a x (isto é, a inclinação) em todos os valores de x. A Figura PT7.3 mostra o gráfico da função e da derivada em função de x. Observe como os valores nulos da derivada correspondem a pontos nos quais a função original é horizontal — isto é, tem inclinação nula. Além disso, os valores absolutos máximos das derivadas estão nas extremidades do intervalo, em que as inclinações da função são maiores.

Embora, como acabamos de ilustrar, possamos determinar uma equação diferencial dada a função original, o objetivo aqui é determinar a função original dada a equação diferencial. A função original, então, representa a solução. No caso presente, podemos determinar essa solução analiticamente integrando a Equação (PT7.13):

$$y = \int (-2x^3 + 12x^2 - 20x + 8,5)\, dx$$

Aplicando a regra de integração (lembre-se da Tabela PT6.2)

$$\int u^n\, du = \frac{u^{n+1}}{n+1} + C \qquad n \neq -1$$

a cada termo da equação obtém-se a solução

$$y = -0,5x^4 + 4x^3 - 10x^2 + 8,5x + C \tag{PT7.14}$$

que é idêntica à função original, exceto por uma diferença importante. Durante a derivação e a posterior integração, perdemos o valor constante 1 na equação original e ganhamos o valor C. Esse C é chamado *constante de integração*. O fato de tal constante arbitrária apare-

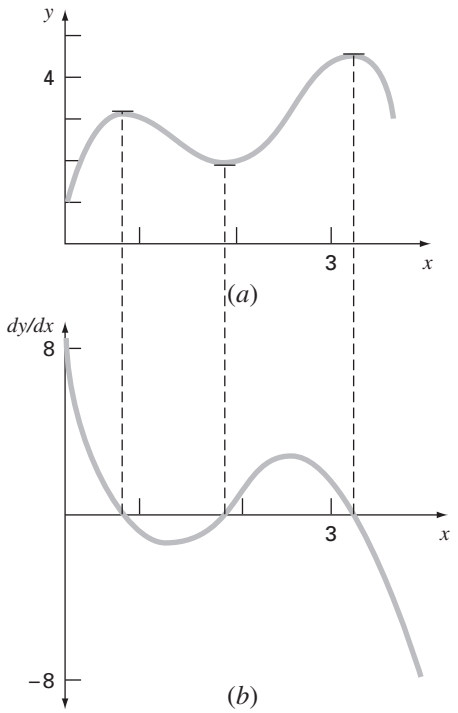


FIGURA PT7.3
Gráficos de (*a*) *y versus x* e (*b*) *dy/dx versus x* para a função $y = -0{,}5x^4 + 4x^3 - 10x^2 + 8{,}5x + 1$.

cer indica que a solução não é única. Na realidade, ela é apenas uma de um número infinito de possíveis funções (correspondendo a um número infinito de possíveis valores de C) que satisfazem a equação diferencial. Por exemplo, a Figura PT7.4 mostra seis funções possíveis que satisfazem a Equação (PT7.14).

Portanto, para especificar completamente a solução, uma equação diferencial está normalmente acompanhada de *condições auxiliares*. Para EDOs de primeira ordem, é necessário um tipo de condição auxiliar denominada *valor inicial* para determinar a constante e obter uma solução única. Por exemplo, a Equação (PT7.13) poderia estar acompanhada pela condição inicial que em $x = 0$, $y = 1$. Esses valores poderiam então ser substituídos na Equação (PT7.14):

$$1 = -0{,}5(0)^4 + 4(0)^3 - 10(0)^2 + 8{,}5(0) + C \tag{PT7.15}$$

para determinar que $C = 1$. Portanto, a única solução que satisfaz tanto a equação diferencial quanto a condição inicial especificada é obtida substituindo $C = 1$ na Equação (PT7.14) para obter

$$y = -0{,}5x^4 + 4x^3 - 10x^2 + 8{,}5x + 1 \tag{PT7.16}$$

Assim, "prendemos" a Equação (PT7.14) forçando-a a passar pela condição inicial e, ao fazer isso, determinamos uma única solução da EDO e completamos o ciclo para a função original [Equação (PT7.12)].

As condições iniciais, em geral, têm interpretação física tangível nas equações diferenciais provenientes de situações de problema físico. Por exemplo, no problema do pára-quedista em queda livre, a condição inicial era um reflexo do fato físico de que, no instante zero, a velocidade vertical era nula. Se o pára-quedista já estivesse em movimento vertical no instante zero, a solução teria sido modificada para levar em conta essa velocidade inicial.

Quando estivermos lidando com uma equação diferencial de ordem n, serão necessárias n condições para obter uma solução única. Se todas as condições forem especificadas

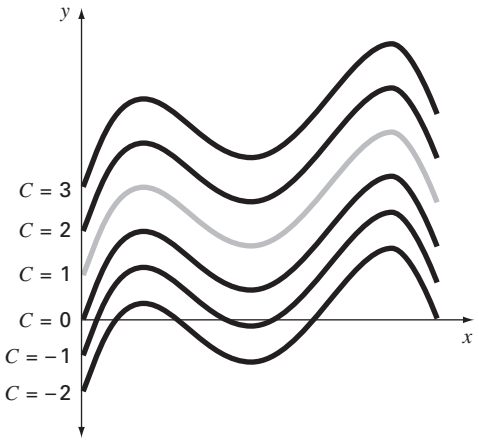


FIGURA PT7.4
Seis soluções possíveis para a integral de $-2x^3 + 12x^2 - 20x + 8{,}5$. Cada uma corresponde a um valor diferente da constante de integração C.

em um mesmo valor da variável independente (por exemplo, em $x = 0$ ou $t = 0$), então o problema é chamado *problema de valor inicial*. Isso para diferenciá-lo dos *problemas de contorno*, nos quais as especificações das condições ocorrem em valores diferentes da variável independente. Os Capítulos 25 e 26 se concentrarão em problemas de valor inicial. Os problemas de contorno serão estudados no Capítulo 27, juntamente com autovalores.

PT7.3 ORIENTAÇÃO

Antes de prosseguir para os métodos numéricos para a resolução de equações diferenciais ordinárias, alguma orientação poderia ser útil. Nossa intenção, com o material descrito a seguir, é lhe fornecer uma visão geral do material discutido na Parte Sete. Além disso, formulamos objetivos para focalizar seus estudos do assunto.

PT7.3.1 Escopo e Visão Geral

A Figura PT7.5 fornece uma visão geral da Parte Sete. Duas categorias amplas de métodos numéricos para problemas de valor inicial serão discutidas nesta parte do livro. Os métodos de passo único, que são cobertos no Capítulo 25, permitem o cálculo de y_{i+1}, dada a equação diferencial e y_i. Os métodos de passo múltiplo, que são cobertos no Capítulo 26, exigem valores adicionais de y além de em i.

A não ser por uma pequena exceção, os *métodos de passo único*, descritos no Capítulo 25, pertencem ao que se chama de técnicas de Runge-Kutta. Embora o capítulo pudesse ter sido organizado em torno dessa noção teórica, optamos por uma abordagem mais gráfica e intuitiva para introduzir os métodos. Assim, começamos o capítulo com o *método de Euler*, o qual tem uma interpretação gráfica muito simples. Então, usamos argumentos visualmente orientados para desenvolver duas versões melhoradas do método de Euler — as técnicas de *Heun* e do *ponto médio*. Depois dessa introdução, deduziremos formalmente o conceito das abordagens de *Runge-Kutta* (ou *RK*) e mostraremos como as técnicas anteriores são de fato métodos de primeira e segunda ordens de RK. Isso é seguido por uma discussão das formulações de alta ordem de RK que são freqüentemente usadas na resolução de problemas de engenharia. Além disso, discutimos as aplicações dos métodos de passo único aos *sistemas de EDOs*. Finalmente, o capítulo termina com uma discussão dos *métodos RK adaptativos*, que automaticamente ajustam o tamanho do passo em resposta ao erro de truncamento do cálculo.

O Capítulo 26 começa com a descrição de *EDOs rígidas*. Estas são tanto EDOs individuais quanto sistemas de EDOs que têm ambas componentes rápidas e lentas em suas

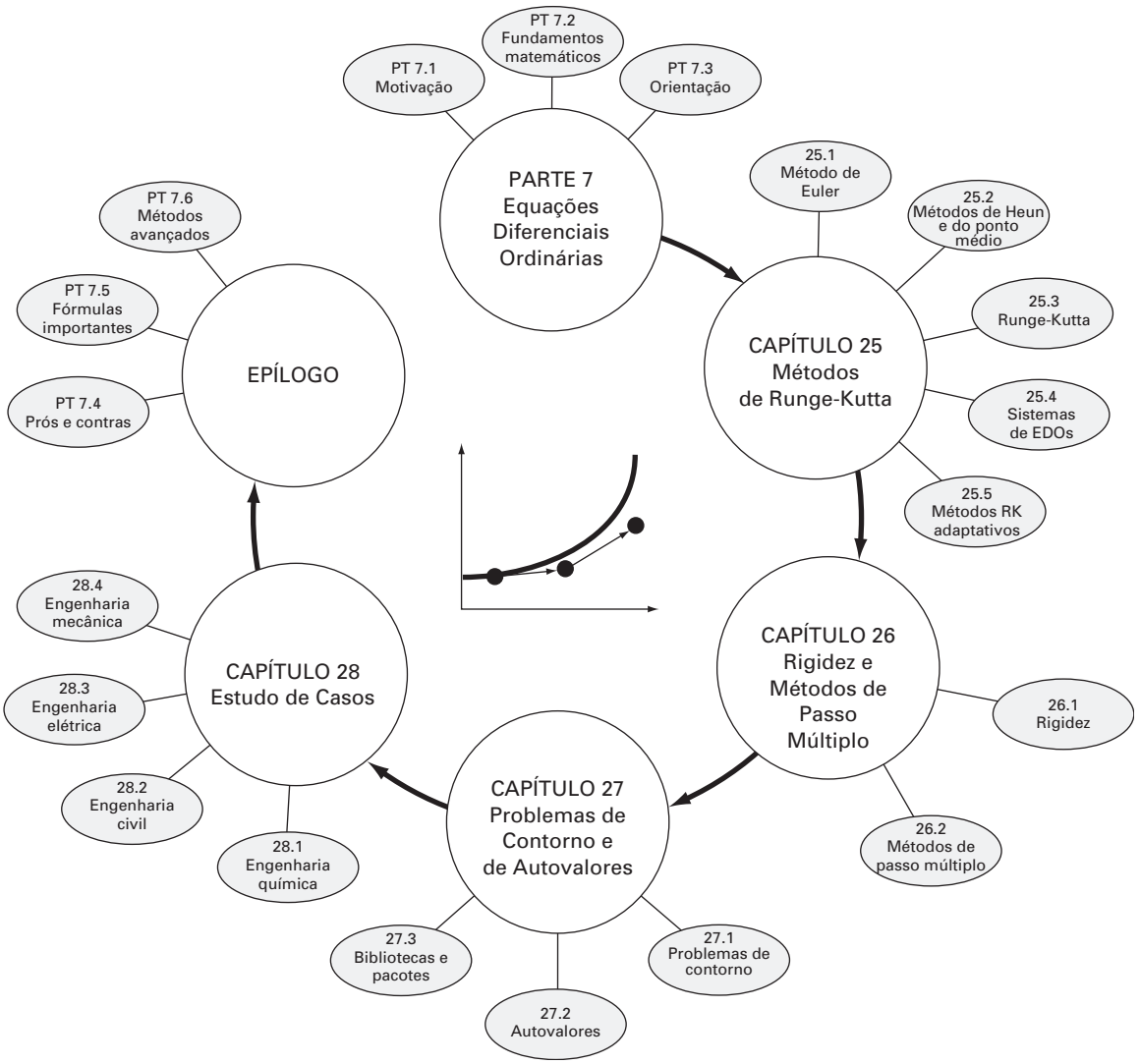


FIGURA PT7.3.1
Organização esquemática da Parte Sete: Equações Diferenciais Ordinárias.

soluções. Introduzimos a idéia de uma técnica de *solução implícita* como comumente usada para remediar esse problema.

A seguir, discutiremos *métodos de passo múltiplo*. Esses algoritmos retêm informações dos passos anteriores para capturar de forma mais eficiente a trajetória da solução. Eles também fornecem estimativas do erro de truncamento que podem ser usadas para implementar o controle do tamanho do passo. Nesta seção, tomamos, inicialmente, uma abordagem visual intuitiva por meio do uso de um método simples — o método de *Heun de início não automático* — para introduzir todas as características essenciais da abordagem de passo múltiplo.

No Capítulo 27, nos voltamos para os problemas de *contorno* e de *autovalores*. Para os primeiros, introduziremos os *métodos shooting* e de *diferenças finitas*. Para os últimos, discutiremos diversas abordagens, incluindo os *métodos polinomial* e *de potência*. Finalmente, o capítulo termina com uma descrição da aplicação de diversos *pacotes* e *bibliotecas de software* para a solução de EDOs e autovalores.

O Capítulo 28 é dedicado a aplicações de todos os campos da engenharia. Finalmente, é incluída uma pequena seção de revisão no final da Parte Sete. Esse epílogo resume e compara fórmulas importantes e conceitos relacionados com as EDOs. A comparação inclui uma discussão dos prós e contras que são relevantes para sua implementação na prática da engenharia. O epílogo também resume fórmulas importantes e inclui referências para tópicos avançados.

PT7.3.2 Metas e Objetivos

Objetivos de Estudo. Depois de completar a Parte Sete, o estudante deve ter aumentado imensamente sua capacidade de confrontar e resolver equações diferenciais ordinárias e problemas de autovalores. Metas gerais de estudo deveriam incluir o domínio das técnicas, bem como a capacidade de avaliar a confiabilidade das respostas e ser capaz de escolher o "melhor" método (ou métodos) para qualquer problema em particular. Além desses objetivos gerais, os objetivos de estudo específicos na Tabela PT7.2 deveriam ser dominados.

Objetivos Computacionais. São fornecidos algoritmos para muitos dos métodos na Parte Sete. Essa informação lhe permitirá expandir sua biblioteca de software. Por exemplo, você pode achar útil, do ponto de vista profissional, ter um software que use o método de Runge-Kutta de quarta ordem para mais que cinco equações e resolver EDOs com uma abordagem adaptativa do tamanho do passo.

Além disso, uma das metas mais importantes deveria ser dominar os diversos pacotes de software de propósito geral que estão amplamente disponíveis. Em particular, você deveria aderir ao uso dessas ferramentas para implementar métodos numéricos para a resolução de problemas de engenharia.

TABELA PT7.2 Objetivos de estudo específicos para a Parte Sete.

1. Entender a representação visual dos métodos de Euler, Heun e do ponto médio.
2. Saber a relação do método de Euler com a expansão em série de Taylor e a percepção do que isso fornece em relação ao erro do método.
3. Compreender a diferença entre os erros de truncamento local e global e como eles se relacionam com o método numérico para um problema particular.
4. Saber a dependência com a ordem e o tamanho do passo com os erros de truncamento globais para todos os métodos descritos na Parte Sete; entender como esses erros agem na acurácia das técnicas.
5. Entender as bases dos métodos preditor-corretor; em particular, perceber que a eficiência do corretor depende altamente da acurácia do preditor.
6. Saber a forma geral dos métodos de Runge-Kutta; entender a dedução do método RK de segunda ordem e como ele se relaciona com a expansão em série de Taylor; perceber que existe um número infinito de versões possíveis para os métodos RK de segunda ordem e de ordem superior.
7. Saber como aplicar qualquer dos métodos RK aos sistemas de equações; ser capaz de reduzir uma EDO de ordem n a um sistema de n EDOs de primeira ordem.
8. Reconhecer o tipo de contexto do problema no qual o ajuste do tamanho do passo é importante.
9. Entender como o controle adaptativo do tamanho do passo é integrado ao método RK de quarta ordem.
10. Perceber como a combinação das componentes lenta e rápida torna a equação ou um sistema de equações rígidos.
11. Entender a distinção entre esquemas de soluções explícitas e implícitas para EDOs; em particular, perceber como os últimos (1) amenizam o problema da rigidez e (2) complicam a mecânica da solução.
12. Compreender a diferença entre problemas de valor inicial e problemas de contorno.
13. Saber a diferença entre métodos de passo múltiplo e de passo único; perceber que todos os métodos de passo múltiplo são preditor-corretor, mas que nem todos os métodos preditor-corretor são de passo múltiplo.
14. Entender a ligação entre fórmulas de integração e métodos preditor-corretor.
15. Reconhecer a diferença fundamental entre as fórmulas de integração de Newton-Cotes e Adams.
16. Saber o raciocínio por trás dos métodos polinomiais e de potência para determinar autovalores; em particular, perceber seus pontos fortes e suas limitações.
17. Entender como a deflação de Hoteller permite que o método de potência seja usado para calcular autovalores intermediários.
18. Saber como usar pacotes e/ou bibliotecas de software para integrar EDOs e calcular autovalores.

CAPÍTULO 25

Métodos de Runge-Kutta

Este capítulo é dedicado à solução de equações diferenciais ordinárias da forma

$$\frac{dy}{dx} = f(x, y)$$

No Capítulo 1, usamos um método numérico para resolver uma equação desse tipo para a velocidade do pára-quedista em queda livre. Lembre-se de que o método tinha a forma geral

$$\text{Valor novo} = \text{valor antigo} + \text{inclinação} \times \text{tamanho do passo}$$

ou, em termos matemáticos,

$$y_{i+1} = y_i + \phi h \tag{25.1}$$

De acordo com essa equação, a estimativa da inclinação ϕ é usada para extrapolar de um valor antigo y_i para um valor novo y_{i+1} em uma distância h (Figura 25.1). Essa fórmula pode ser aplicada passo a passo para cálculos no futuro e, portanto, para percorrer a trajetória da solução.

Todos os métodos de passo único podem ser expressos nessa forma geral, sendo que a única diferença é a maneira como é feita a estimativa da inclinação. Como no problema do pára-quedista em queda livre, a abordagem mais simples é usar a equação diferencial para obter uma estimativa da inclinação na forma da primeira derivada em x_i. Em outras palavras, a inclinação no início do intervalo é tomada como uma aproximação da inclinação média em todo o intervalo. Essa abordagem, chamada de *método de Euler*, é discutida na primeira parte deste capítulo. O assunto é seguido por outros métodos de passo único que usam estimativas alternativas da inclinação que resultam em previsões mais acuradas. Todas essas técnicas são chamadas, em geral, de métodos de *Runge-Kutta*.

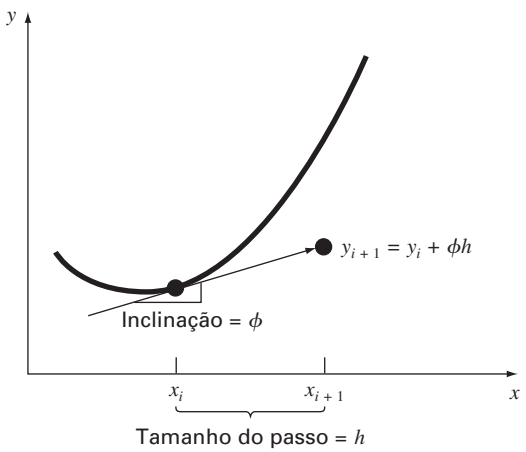


FIGURA 25.1
Descrição gráfica de um método de um passo.

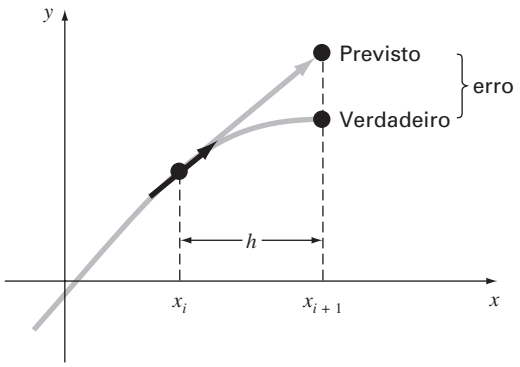


FIGURA 25.2
Método de Euler.

25.1 MÉTODO DE EULER

A primeira derivada fornece uma estimativa direta da inclinação em x_i (Figura 25.2):

$$\phi = f(x_i, y_i)$$

em que $f(x_i, y_i)$ é a equação diferencial calculada em x_i e y_i. Essa estimativa pode ser substituída na Equação (25.1):

$$y_{i+1} = y_i + f(x_i, y_i)h \tag{25.2}$$

Essa fórmula é conhecida como *método de Euler* (ou *Euler-Cauchy* ou *ponto-inclinação*). Um novo valor de y é previsto usando a inclinação (igual à primeira derivada no valor original de x) para extrapolar linearmente sobre um tamanho de passo h (Figura 25.2).

EXEMPLO 25.1 Método de Euler

Enunciado do Problema. Use o método de Euler para integrar numericamente a Equação (PT7.13):

$$\frac{dy}{dx} = -2x^3 + 12x^2 - 20x + 8{,}5$$

de $x = 0$ a $x = 4$ com um tamanho de passo de 0,5. A condição inicial em $x = 0$ é $y = 1$. Lembre-se de que a solução exata é dada pela Equação (PT7.16):

$$y = -0{,}5x^4 + 4x^3 - 10x^2 + 8{,}5x + 1$$

Solução. A Equação (25.2) pode ser usada para implementar o método de Euler:

$$y(0{,}5) = y(0) + f(0, 1)0{,}5$$

em que $y(0) = 1$ e a estimativa da inclinação em $x = 0$ é

$$f(0, 1) = -2(0)^3 + 12(0)^2 - 20(0) + 8{,}5 = 8{,}5$$

Portanto,

$$y(0{,}5) = 1{,}0 + 8{,}5(0{,}5) = 5{,}25$$

A solução verdadeira em $x = 0{,}5$ é

$$y = -0{,}5(0{,}5)^4 + 4(0{,}5)^3 - 10(0{,}5)^2 + 8{,}5(0{,}5) + 1 = 3{,}21875$$

Logo, o erro é

$$E_t = \text{verdadeiro} - \text{aproximado} = 3{,}21875 - 5{,}25 = -2{,}03125$$

ou, expresso como um erro relativo porcentual, $\varepsilon_t = -63{,}1\%$. Para o segundo passo,

$$\begin{aligned} y(1) &= y(0{,}5) + f(0{,}5, 5{,}25)0{,}5 \\ &= 5{,}25 + [-2(0{,}5)^3 + 12(0{,}5)^2 - 20(0{,}5) + 8{,}5]0{,}5 \\ &= 5{,}875 \end{aligned}$$

A solução verdadeira em $x = 1{,}0$ é 3,0 e, portanto, o erro relativo porcentual é $-95{,}8\%$. O cálculo é repetido e os resultados estão resumidos na Tabela 25.1 e na Figura 25.3. Observe que, embora os cálculos capturem a tendência geral da solução verdadeira, o erro é considerável. Como discutido na próxima seção, esse erro pode ser reduzido pelo uso de um tamanho de passo menor.

TABELA 25.1 Comparação dos valores verdadeiro e aproximado da integral de $y' = -2x^3 + 12x^2 - 20x + 8{,}5$, com a condição inicial de que $y = 1$ em $x = 0$. Os valores aproximados foram calculados usando o método de Euler com um tamanho de passo 0,5. O erro local refere-se ao erro provocado por um único passo. É calculado com a expansão em série de Taylor como no Exemplo 25.2. O erro global é a discrepância global por causa dos passos passados, bem como do passo atual.

			Erro Porcentual Relativo	
x	$y_{verdadeiro}$	y_{Euler}	Global	Local
0,0	1,00000	1,00000		
0,5	3,21875	5,25000	−63,1	−63,1
1,0	3,00000	5,87500	−95,8	−28,0
1,5	2,21875	5,12500	131,0	−1,41
2,0	2,00000	4,50000	−125,0	20,5
2,5	2,71875	4,75000	−74,7	17,3
3,0	4,00000	5,87500	46,9	4,0
3,5	4,71875	7,12500	−51,0	−11,3
4,0	3,00000	7,00000	−133,3	−53,0

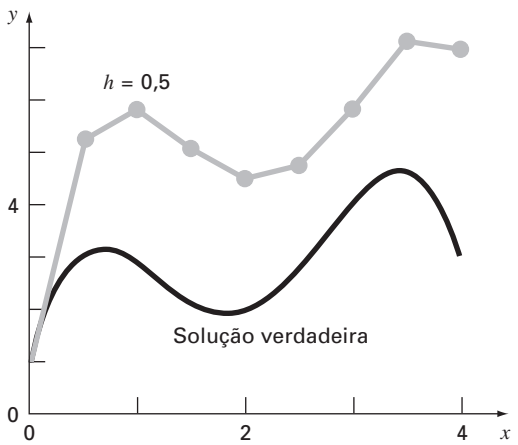


FIGURA 25.3
Comparação da solução verdadeira com a solução numérica usando o método de Euler para a integral de $y' = -2x^3 + 12x^2 - 20x + 8{,}5$ de $x = 0$ a $x = 4$ com um tamanho de passo de 0,5. A condição inicial em $x = 0$ é $y = 1$.

O exemplo anterior usa um polinômio simples para a equação diferencial para facilitar a análise de erro a seguir. Logo,

$$\frac{dy}{dx} = f(x)$$

Obviamente, um caso mais geral (e mais comum) envolveria EDOs que dependam de x e y,

$$\frac{dy}{dx} = f(x, y)$$

Conforme progredirmos por esta parte do texto, nossos exemplos vão envolver cada vez mais EDOs que dependem tanto das variáveis dependentes quanto das independentes.

25.1.1 Análise de Erro para o Método de Euler

A solução numérica de EDOs envolve dois tipos de erros (lembre-se dos Capítulos 3 e 4):

1. Erros de *truncamento* ou de discretização causados pela natureza das técnicas usadas para aproximar os valores de y.
2. Erros de *arredondamento* causados pelo número limitado de algarismos significativos que podem ser representados por um computador.

Os erros de truncamento são compostos de duas partes. A primeira é o *erro de truncamento local*, que resulta da aplicação do método em questão em um único passo. A segunda é o *erro de truncamento propagado*, que resulta das aproximações produzidas durante os passos anteriores. A soma dos dois é o *erro de truncamento global* ou total.

Uma visão do valor absoluto e das propriedades do erro de truncamento pode ser obtida deduzindo-se o método de Euler diretamente da expansão em série de Taylor. Para fazê-lo, lembre-se de que a equação a ser integrada terá a forma geral

$$y' = f(x, y) \tag{25.3}$$

em que $y' = dy/dx$ e x e y são as variáveis independente e dependente, respectivamente. Se a solução — isto é, a função descrevendo o comportamento de y — tiver derivadas contínuas, pode ser representada por uma expansão em série de Taylor em torno de um valor inicial (x_i, y_i), como em [lembre-se da Equação (4.7)]

$$y_{i+1} = y_i + y_i' h + \frac{y_i''}{2!}h^2 + \cdots + \frac{y_i^{(n)}}{n!}h^n + R_n \tag{25.4}$$

em que $h = x_{i+1} - x_i$ e R_n é o resto, dado por

$$R_n = \frac{y^{(n+1)}(\xi)}{(n+1)!}h^{n+1} \tag{25.5}$$

em que ξ está em algum ponto do intervalo entre x_i e x_{i+1}. Uma forma alternativa pode ser desenvolvida substituindo-se a Equação (25.3) nas Equações (25.4) e (25.5) para obter

$$y_{i+1} = y_i + f(x_i, y_i)h + \frac{f'(x_i, y_i)}{2!}h^2 + \cdots + \frac{f^{(n-1)}(x_i, y_i)}{n!}h^n + O(h^{n+1}) \tag{25.6}$$

em que $O(h^{n+1})$ especifica que o erro de truncamento local é proporcional ao tamanho do passo elevado à $(n + 1)$-ésima potência.

Comparando as equações (25.2) e (25.6), pode ser visto que o método de Euler corresponde à série de Taylor até e incluindo o termo $f(x_i, y_i)h$. Adicionalmente, a comparação indica que ocorre um erro de truncamento porque aproximamos a solução verdadeira usando um número finito de termos da série de Taylor. Então, truncamos ou deixamos de fora uma parte da solução verdadeira. Por exemplo, o erro de truncamento no método de Euler é atribuído aos termos restantes na série de Taylor que não foram incluídos na Equação (25.2). Subtraindo a Equação (25.2) da Equação (25.6), obtemos

$$E_t = \frac{f'(x_i, y_i)}{2!}h^2 + \cdots + O(h^{n+1}) \tag{25.7}$$

em que E_t é o erro de truncamento local verdadeiro. Para h suficientemente pequeno, os erros nos termos da Equação (25.7) em geral diminuem conforme a ordem aumenta (lembre-se do Exemplo 4.2 e da discussão que o acompanhava) e o resultado freqüentemente

é representado por

$$E_a = \frac{f'(x_i, y_i)}{2!}h^2 \tag{25.8}$$

ou

$$E_a = O(h^2) \tag{25.9}$$

em que E_a é o erro de truncamento local aproximado.

EXEMPLO 25.2 Estimativa por Série de Taylor para o Erro do Método de Euler

Enunciado do Problema. Use a Equação (25.7) para obter uma estimativa do erro do primeiro passo do Exemplo 25.1. Use-a também para determinar o erro devido a cada termo de ordem mais alta da expansão em série de Taylor.

Solução. Como estamos lidando com um polinômio, podemos usar a série de Taylor para obter estimativas exatas dos erros no método de Euler. A Equação (25.7) pode ser escrita como

$$E_t = \frac{f'(x_i, y_i)}{2!}h^2 + \frac{f''(x_i, y_i)}{3!}h^3 + \frac{f^{(3)}(x_i, y_i)}{4!}h^4 \tag{E25.2.1}$$

em que $f'(x_i, y_i)$ é a primeira derivada da equação diferencial (isto é, a segunda derivada da solução). Para o caso presente, isso é

$$f'(x_i, y_i) = -6x^2 + 24x - 20 \tag{E25.2.2}$$

e $f''(x_i, y_i)$ é a segunda derivada da EDO

$$f''(x_i, y_i) = -12x + 24 \tag{E25.2.3}$$

e $f^{(3)}(x_i, y_i)$ é a terceira derivada da EDO

$$f^{(3)}(x_i, y_i) = -12 \tag{E25.2.4}$$

Podemos omitir os termos adicionais (isto é, derivadas de quarta ordem ou de ordem superior) da Equação (E25.2.1) porque, nesse caso particular, eles são iguais a zero. Deve ser observado que para outras funções (por exemplo, funções transcendentais tais como funções senoidais ou exponenciais) isso não seria necessariamente verdade e os termos de ordem mais alta teriam valores não-nulos. Entretanto, no caso presente, as Equações (E25.2.1) a (E25.2.4) definem completamente o erro de truncamento para uma única aplicação do método de Euler.

Por exemplo, o erro devido ao truncamento do termo de segunda ordem pode ser calculado como

$$E_{t,2} = \frac{-6(0,0)^2 + 24(0,0) - 20}{2}(0,5)^2 = -2,5 \tag{E25.2.5}$$

Para o termo de terceira ordem:

$$E_{t,3} = \frac{-12(0,0) + 24}{6}(0,5)^3 = 0,5$$

e para o termo de quarta ordem:

$$E_{t,4} = \frac{-12}{24}(0,5)^4 = -0,03125$$

Esses três resultados podem ser somados para fornecer o erro de truncamento total:

$$E_t = E_{t,2} + E_{t,3} + E_{t,4} = -2,5 + 0,5 - 0,03125 = -2,03125$$

que é exatamente o erro que foi introduzido no passo inicial do Exemplo 25.1. Observe como $E_{t,2} > E_{t,3} > E_{t,4}$, o que confirma a aproximação representada pela Equação (25.8).

Como ilustrado no Exemplo 25.2, a série de Taylor fornece um meio de quantificar o erro no método de Euler. Mas existem limitações associadas com seu uso para esse propósito:

1. A série de Taylor fornece apenas uma estimativa do erro de truncamento local — isto é, o erro criado durante um único passo do método. Não fornece uma medida do erro de truncamento propagado, e, portanto, global. Na Tabela 25.1, incluímos os erros de truncamento local e global para o Exemplo 25.1. O erro local foi calculado para cada passo no tempo com a Equação (25.2), mas usando o valor verdadeiro de y_i (a segunda coluna da tabela) para calcular y_{i+1} em vez do valor aproximado (a terceira coluna), como é feito no método de Euler. Como esperado, a média do valor absoluto do erro de truncamento local (25%) é menor do que a média do erro global (90%). A única razão pela qual pudemos fazer esses cálculos exatos de erros é que sabemos o valor verdadeiro *a priori*. Esse não seria o caso em um problema real. Portanto, como discutido a seguir, em geral devemos aplicar técnicas como o método de Euler usando tamanhos de passo diferentes para obter uma estimativa indireta dos erros envolvidos.
2. Como mencionado, em problemas reais usualmente lidamos com funções mais complicadas do que simples polinômios. Por isso, as derivadas que são necessárias para calcular a expansão em série de Taylor nem sempre seriam fáceis de obter.

Embora essas limitações impeçam a análise de erro exata na maioria dos problemas práticos, a série de Taylor ainda fornece uma percepção valiosa do comportamento do método de Euler. De acordo com a Equação (25.9), vemos que o erro local é proporcional ao quadrado do tamanho do passo e à primeira derivada da equação diferencial. Pode ser mostrado também que o erro de truncamento global é $O(h)$, ou seja, proporcional ao tamanho do passo (Carnahan et al., 1969). Essas observações levam a algumas conclusões úteis:

1. O erro pode ser reduzido diminuindo-se o tamanho do passo.
2. O método fornecerá previsões livres de erros se a função subjacente (isto é, a solução da equação diferencial) for linear, porque, para uma reta, a segunda derivada seria nula.

A última conclusão faz sentido do ponto de vista intuitivo porque o método de Euler usa segmentos de reta para aproximar a solução. Portanto, o método de Euler é conhecido como um *método de primeira ordem*.

Deve ser observado também que esse padrão geral é válido para os métodos de passo único de ordem superior descritos nas páginas seguintes. Isto é, um método de ordem n fornecerá um resultado perfeito se a solução subjacente for um polinômio de grau n. Além disso, o erro de truncamento local será $O(h^{n+1})$, e o erro global, $O(h^n)$.

EXEMPLO 25.3 Efeito da Redução do Tamanho do Passo no Método de Euler

Enunciado do Problema. Repita os cálculos do Exemplo 25.1, mas use um tamanho de passo de 0,25.

Solução. O cálculo é repetido e os resultados estão compilados na Figura 25.4*a*. A divisão do tamanho do passo por dois reduz o valor absoluto do erro médio global para 40% e o valor absoluto do erro local para 6,4%. Isso é comparado com os erros global e local do Exemplo 25.1, de 90% e 24,8%, respectivamente. Assim, como esperado, o erro local é dividido por quatro e o erro global, por dois.

Observe também como o erro local muda de sinal para valores intermediários ao longo do intervalo. Isso decorre principalmente do fato de que a primeira derivada da equação diferencial é uma parábola que muda de sinal [lembre-se da Equação (E25.2.2) e veja a Figura 25.4*b*]. Como o erro local é proporcional a essa função, o efeito resultante da oscilação do sinal é impedir que o erro global continue crescendo conforme os

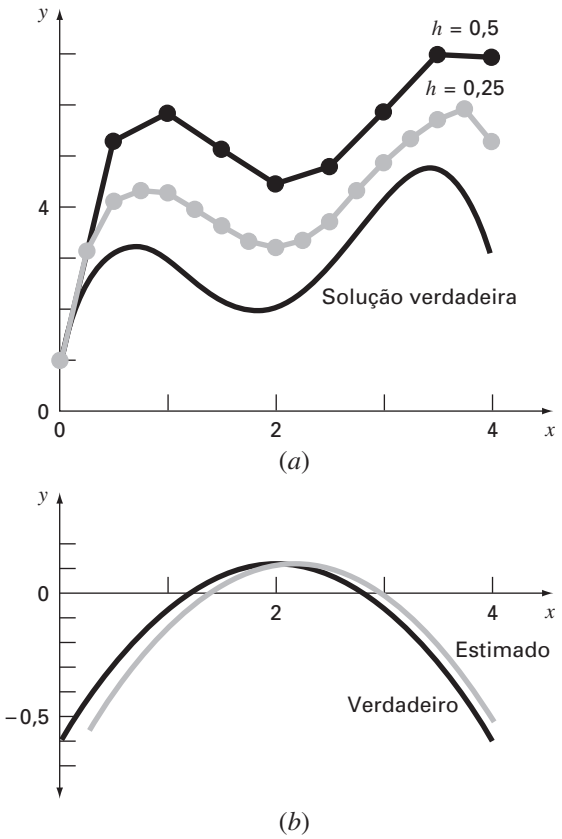


FIGURA 25.4
(*a*) Comparação de duas soluções numéricas com o método de Euler usando tamanhos de passo de 0,5 e 0,25. (*b*) Comparação do erro de truncamento local verdadeiro e estimado para o caso no qual o tamanho do passo é 0,5. Observe que o erro "estimado" é baseado na Equação (E25.2.5).

cálculos continuam. Logo, de $x = 0$ a $x = 1{,}25$, os erros locais são todos negativos e, conseqüentemente, o erro global aumenta nesse intervalo. Na seção intermediária do intervalo, erros locais positivos começam a reduzir o erro global. Perto da extremidade, o processo é revertido e o erro global aumenta de novo. Se o erro local mudar continuamente de sinal no intervalo de cálculo, o efeito resultante será usualmente reduzir o erro global. Entretanto, se os erros locais forem do mesmo sinal, a solução numérica pode divergir para cada vez mais longe da solução verdadeira conforme os cálculos continuam. Tais resultados são ditos *instáveis.*

O efeito de mais reduções no tamanho do passo no erro de truncamento global do método de Euler é ilustrado na Figura 25.5. Esse gráfico mostra o valor absoluto do erro relativo porcentual em $x = 5$ como uma função do tamanho do passo para o problema que examinamos nos Exemplos 25.1.a 25.3. Observe que mesmo quando h é reduzido para 0,001, o erro ainda excede 0,1%. Como esse tamanho de passo corresponde a 5.000 passos para ir de $x = 0$ a $x = 5$, o gráfico sugere que uma técnica de primeira ordem como o método de Euler demanda um grande esforço computacional para obter níveis de erro aceitáveis. Mais tarde neste capítulo, apresentaremos técnicas de ordem superior que alcançam uma acurácia muito melhor para o mesmo esforço computacional. Contudo, deve ser observado que, apesar de sua ineficiência, a simplicidade do método de Euler torna-o uma opção extremamente atrativa para muitos problemas de engenharia. Como ele é muito fácil de programar, a técnica é particularmente útil para análises rápidas. Na próxima seção, será desenvolvido um algoritmo computacional para o método de Euler.

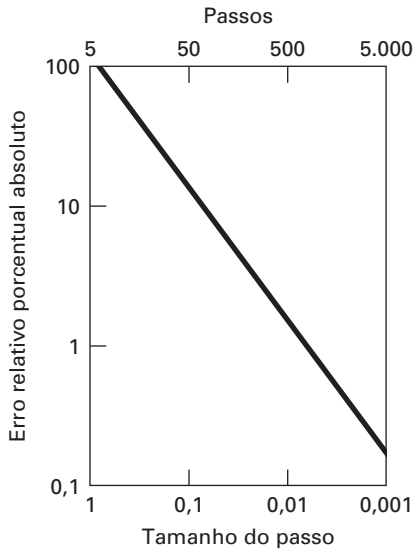


FIGURA 25.5
Efeito do tamanho do passo no erro de truncamento global do método de Euler para a integral de $y' = -2x^3 + 12x^2 - 20x + 8{,}5$. O gráfico mostra o valor absoluto do erro global relativo porcentual em $x = 5$ como uma função do tamanho do passo.

25.1.2 Algoritmo para o Método de Euler

Os algoritmos para métodos de passo único são extremamente simples de programar. Como especificado no início deste capítulo, todos os métodos de passo único têm a forma geral

$$\text{Novo valor} = \text{valor antigo} + \text{inclinação} \times \text{tamanho do passo} \tag{25.10}$$

A única maneira pela qual os métodos diferem é no cálculo da inclinação.

Suponha que você queira fazer os cálculos simples descritos na Tabela 25.1. Ou seja, você gostaria de usar o método de Euler para integrar $y' = -2x^3 + 12x^2 - 20x + 8{,}5$ com a condição inicial que $y = 1$ em $x = 0$. Você gostaria de integrar até $x = 4$, usando um tamanho de passo 0,5 e mostrando todos os resultados. Um pseudocódigo simples para completar essa tarefa poderia ser escrito como na Figura 25.6.

FIGURA 25.6
Pseudocódigo para uma versão "simplória" do método de Euler.

```
'defina o inervalo de integração
xi = 0
xf = 4
'inicialize as variáveis
x = xi
y = 1
'defina o tamanho do passo e o
'número de passos do cálculo
dx = 0.5
nc = (xf − xi)/dx
'imprima as condições iniciais
PRINT x, y
'laço para implementar o método de Euler
'e mostrar os resultados
DOFOR i = 1, nc
  dydx = −2x^3 + 12x^2 − 20x + 8.5
  y = y + dydx · dx
  x = x + dx
  PRINT x, y
END DO
```

Embora esse programa consiga duplicar os resultados da Tabela 25.1, ele não é muito bem projetado. Primeiro, e principalmente, não é muito modular. Ainda que isso não seja muito importante para um programa tão pequeno, seria crítico se quiséssemos modificar e melhorar esse algoritmo.

Além disso, existem diversas questões relacionadas à maneira como implementamos as iterações. Por exemplo, suponha que o tamanho do passo deva ser muito pequeno para obter melhor acurácia. Em tais casos, como todo valor calculado é mostrado, o número de valores de saída poderia ser muito grande. Além disso, o algoritmo pressupõe que o intervalo de cálculo é divisível pelo tamanho do passo. Finalmente, a distribuição de x na reta $x = x + dx$ pode estar sujeito a erros de quantificação do tipo discutido anteriormente na Seção 3.4.1. Por exemplo, se dx for mudado para 0,01 e a representação IEEE padrão em ponto flutuante fosse usada (cerca de sete algarismos significativos), o resultado no final dos cálculos seria 3,999997 em vez de 4. Para $dx = 0{,}001$, seria 3,999892!

Um algoritmo muito mais modular que evita essas dificuldades está mostrado na Figura 25.7. O algoritmo não mostra todos os valores calculados. Em vez disso, o usuário especifica um intervalo de saída, *xout*, que determina o intervalo no qual os resultados calculados estão armazenados em vetores, xp_m e yp_m. Esses valores são armazenados em vetores de modo que possam ser exibidos de diversas formas depois que o cálculo for completado (por exemplo, impresso, ou em um gráfico ou copiado em um arquivo).

O Programa Principal faz passos de saída maiores e chama uma rotina Integrator que faz passos de cálculo menores. Observe que os laços que controlam os passos grandes e pequenos param por condições lógicas. Logo, os intervalos não têm de ser exatamente divisíveis pelos tamanhos de passo.

A rotina Integrator então chama uma rotina de Euler que executa um único passo do método de Euler. A rotina Euler chama uma rotina Derivative que calcula o valor da derivada.

Tal modulação pode parecer um exagero no caso presente, mas facilitará muito a modificação do programa nas seções posteriores. Por exemplo, embora o programa na Figura 25.7 tenha sido projetado especificamente para implementar o método de Euler, o módulo de Euler é a única parte que é específica do método. Logo, tudo que é necessário para aplicar esse algoritmo aos outros métodos de passo único é modificar essa rotina.

FIGURA 25.7
Pseudocódigo para uma versão modular "melhorada" do método de Euler.

(*a*) Programa Principal

```
Designe valores para
y = valor inicial da variável dependente
xi = valor inicial da variável
independente
xf = valor final da variável
independente
dx = tamanho do passo de cálculo
xout = intervalo de saída

x = xi
m = 0
xp_m = x
yp_m = y
DO
  xend = x + xout
  IF (xend > xf) THEN xend = xf
  h = dx
  CALL Integrator (x, y, h, xend)
  m = m + 1
  xp_m = x
  yp_m = y
  IF (x ≥ xf) EXIT
END DO
DISPLAY RESULTS
END
```

(*b*) Rotina para Fazer um Passo de Saída

```
SUB Integrator (x, y, h, xend)
  DO
    IF (xend − x < h) THEN h = xend − x
    CALL Euler (x, y, h, ynew)
    y = ynew
    IF (x ≥ xend) EXIT
  END DO
END SUB
```

(*c*) Método de Euler para uma Única EDO

```
SUB Euler (x, y, h, ynew)
  CALL Derivs(x, y, dydx)
  ynew = y + dydx * h
  x = x + h
END SUB
```

(*d*) Rotina para Determinar a Derivada

```
SUB Derivs (x, y, dydx)
  dydx = ...
END SUB
```

EXEMPLO 25.4 Resolvendo EDOs com o Computador

Enunciado do Problema. É possível desenvolver um programa de computador a partir do pseudocódigo na Figura 25.7. Podemos usar esse software para resolver um outro problema associado ao pára-quedista em queda livre. Você se lembra, da Parte Um, de que nosso modelo matemático para a velocidade era baseado na segunda lei de movimento de Newton na forma

$$\frac{dv}{dt} = g - \frac{c}{m}v \tag{E25.4.1}$$

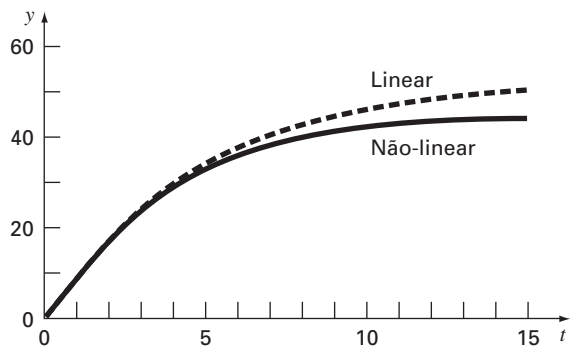
A equação diferencial era resolvida tanto analítica (Exemplo 1.1) quanto numericamente usando o método de Euler (Exemplo 1.2). Essas soluções eram para o caso em que $g = 9{,}8$, $c = 12{,}5$, $m = 68{,}1$, e $v = 0$ em $t = 0$.

O objetivo do exemplo presente é repetir esses cálculos numéricos usando um modelo mais complicado para a velocidade baseado em uma descrição matemática mais completa da força de arrasto causada pela resistência do ar. O modelo é dado por

$$\frac{dv}{dt} = g - \frac{c}{m}\left[v + a\left(\frac{v}{v_{\max}}\right)^b\right] \tag{E25.4.2}$$

em que g, m e c são os mesmos que para a Equação (E25.4.1) e a, b e $v_{\max}$ são constantes empíricas, as quais, para esse caso, são iguais a 8,3, 2,2 e 46, respectivamente. Observe que esse modelo é capaz de ajustar de modo mais acurado medidas empíricas da forças de arrasto em função da velocidade do que o modelo linear simples do Exemplo 1.1. No entanto, esse aumento da flexibilidade é ganho à custa do cálculo de três coeficientes em vez de um. Além disso, o modelo matemático resultante é mais difícil de se resolver analiticamente. Nesse caso, o método de Euler fornece uma alternativa conveniente para obter uma solução numérica aproximada.

FIGURA 25.8
Resultado gráfico da solução da EDO não-linear [Equação (E25.4.2)]. Observe que o gráfico também mostra a solução do modelo linear [Equação (E25.4.1)] para comparação.



Solução. Os resultados para os modelos linear e não-linear estão mostrados na Figura 25.8, com um tamanho de passo de integração de 0,1 s. O gráfico na Figura 25.8 também mostra uma superposição da solução do modelo linear com o propósito de comparação.

Os resultados das duas simulações indicam como o aumento na complexidade da formulação do coeficiente de arrasto afeta a velocidade do pára-quedista. Nesse caso, a velocidade terminal diminui por causa da resistência causada pelos termos de ordem superior na Equação (E25.4.2).

Modelos alternativos poderiam ser testados de maneira parecida. A combinação de uma solução gerada por computador faz que isso seja uma tarefa fácil e eficiente. Tal conveniência permite que você dedique mais tempo à consideração de alternativas criativas e aspectos holísticos do problema do que a cálculos manuais tediosos.

25.1.3 Métodos por Séries de Taylor de Ordem Superior

Uma forma de reduzir o erro no método de Euler é incluir termos de ordem superior da expansão em série de Taylor na solução. Por exemplo, incluindo o termo de segunda ordem da Equação (25.6), obteríamos

$$y_{i+1} = y_i + f(x_i, y_i)h + \frac{f'(x_i, y_i)}{2!}h^2 \tag{25.11}$$

com um erro de truncamento local de

$$E_a = \frac{f''(x_i, y_i)}{6}h^3$$

Embora a incorporação de termos de ordem superior seja suficientemente simples de implementar para polinômios, sua inclusão não é tão trivial quando a EDO é mais complicada. Em particular, EDOs que são uma função tanto da variável dependente quanto da independente exigem a derivação pela regra da cadeia. Por exemplo, a primeira derivada de $f(x, y)$ é

$$f'(x_i, y_i) = \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y}\frac{dy}{dx}$$

A segunda derivada é

$$f''(x_i, y_i) = \frac{\partial[\partial f/\partial x + (\partial f/\partial y)(dy/dx)]}{\partial x} + \frac{\partial[\partial f/\partial x + (\partial f/\partial y)(dy/dx)]}{\partial y}\frac{dy}{dx}$$

As derivadas de ordem superior se tornam cada vez mais complicadas.

Conseqüentemente, como descrito nas próximas seções, foram desenvolvidos métodos de passo único alternativos. Esses esquemas são comparáveis quanto ao desempenho às abordagens por séries de Taylor de ordem superior, mas exigem apenas o cálculo de derivadas de primeira ordem.

25.2 MELHORIAS NO MÉTODO DE EULER

Uma fonte fundamental de erro no método de Euler é que supomos que a derivada no início do intervalo pode ser usada no intervalo todo. Estão disponíveis duas modificações simples para contornar essas deficiências. Como será demonstrado na Seção 25.3, ambas as modificações na verdade pertencem a uma classe maior de técnicas de solução chamadas métodos de Runge-Kutta. Entretanto, como elas têm uma interpretação gráfica muito simples, nós as apresentaremos antes de sua dedução formal como métodos de Runge-Kutta.

25.2.1 Método de Heun

Um método de melhorar a estimativa da inclinação envolve a determinação de duas derivadas para o intervalo — uma no ponto inicial e outra no ponto final. Então, é feita a média das duas derivadas para obter uma estimativa melhorada da inclinação no intervalo todo. Essa abordagem, chamada de *método de Heun*, é descrita graficamente na Figura 25.9.

Lembre-se de que, no método de Euler, a inclinação no início de um intervalo

$$y_i' = f(x_i, y_i) \tag{25.12}$$

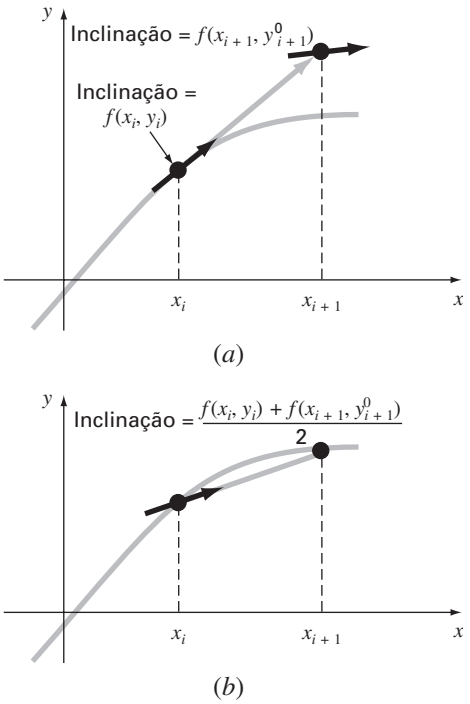


FIGURA 25.9
Descrição gráfica do método de Heun. (*a*) Preditor e (*b*) corretor.

é usada para extrapolar linearmente para y_{i+1}:

$$y^0_{i+1} = y_i + f(x_i, y_i)h \tag{25.13}$$

No método de Euler padrão, pararíamos nesse ponto. No entanto, no método de Heun, o y^0_{i+1} calculado na Equação (25.13) não é a resposta final, mas uma previsão intermediária. É por isso que o distinguimos com o sobrescrito 0. A Equação (25.13) é chamada *equação preditora*. Ela fornece uma estimativa de y_{i+1} que permite o cálculo de uma estimativa da inclinação na extremidade final do intervalo:

$$y'_{i+1} = f\left(x_{i+1}, y^0_{i+1}\right) \tag{25.14}$$

Assim, as duas inclinações [Equações (25.12) e (25.14)] podem ser combinadas para obter uma inclinação média no intervalo:

$$\bar{y}' = \frac{y'_i + y'_{i+1}}{2} = \frac{f(x_i, y_i) + f\left(x_{i+1}, y^0_{i+1}\right)}{2}$$

Essa inclinação média é então usada para extrapolar linearmente de y_i a y_{i+1} usando o método de Euler:

$$y_{i+1} = y_i + \frac{f(x_i, y_i) + f\left(x_{i+1}, y^0_{i+1}\right)}{2}h$$

que é chamada de *equação corretora*.

O método de Heun é uma *abordagem do tipo preditor-corretor*. Todos os métodos de passo múltiplo que serão discutidos no Capítulo 26 são desse tipo. O método de Heun é o único método de passo único preditor-corretor discutido neste livro. Como deduzimos anteriormente, ele pode ser representado sucintamente como

Preditor (Fig. 25.9*a*): $$y^0_{i+1} = y_i + f(x_i, y_i)h \tag{25.15}$$

Corretor (Fig. 25.9*b*): $$y_{i+1} = y_i + \frac{f(x_i, y_i) + f\left(x_{i+1}, y^0_{i+1}\right)}{2}h \tag{25.16}$$

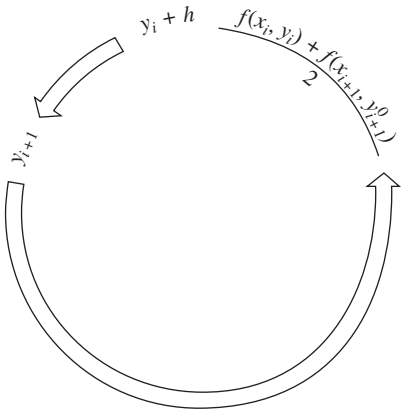


FIGURA 25.10
Representação gráfica da iteração do corretor do método de Heun para obter uma estimativa melhorada.

Observe que, como a Equação (25.16) tem y_{i+1} em ambos os lados do sinal de igual, pode ser aplicada de uma forma iterativa. Isto é, uma estimativa antiga pode ser usada repetidamente para fornecer uma estimativa melhorada de y_{i+1}. O processo é descrito na Figura 25.10. Deve ficar claro que esse processo iterativo não converge necessariamente para a resposta verdadeira, mas convergirá para uma estimativa com um erro de truncamento finito, como ilustrado no próximo exemplo.

Como no caso dos métodos iterativos parecidos discutidos nas seções anteriores do livro, um critério de parada para a convergência do corretor é fornecido por [lembre-se da Equação (3.5)]

$$|\varepsilon_a| = \left| \frac{y^j_{i+1} - y^{j-1}_{i+1}}{y^j_{i+1}} \right| 100\% \tag{25.17}$$

em que y^{j-1}_{i+1} e y^j_{i+1} são os resultados da iteração anterior e da iteração presente do corretor, respectivamente.

EXEMPLO 25.5

Método de Heun

Enunciado do Problema. Use o método de Heun para integrar $y' = 4e^{0,8x} - 0,5y$ de $x = 0$ a $x = 4$ com tamanho do passo 1. A condição inicial em $x = 0$ é $y = 2$.

Solução. Antes de resolver o problema numericamente, podemos usar o cálculo para determinar a seguinte solução analítica:

$$y = \frac{4}{1,3}(e^{0,8x} - e^{-0,5x}) + 2e^{-0,5x} \tag{E25.5.1}$$

Essa fórmula pode ser usada para gerar os valores da solução verdadeira na Tabela 25.2.

Primeiro, a inclinação em (x_0, y_0) é calculada como

$$y'_0 = 4e^0 - 0,5(2) = 3$$

Esse resultado é bem diferente da inclinação média real no intervalo de 0 a 1,0, que é igual a 4,1946, como calculada a partir da equação diferencial usando a Equação (PT 6.4).

A solução numérica é obtida pelo uso do preditor [Equação (25.15)] para obter uma estimativa de y em 1,0:

$$y^0_1 = 2 + 3(1) = 5$$

TABELA 25.2 Comparação dos valores verdadeiro e aproximado da integral de $y' = 4e^{0,8x} - 0,5y$, com a condição inicial de que $y = 2$ em $x = 0$. Os valores aproximados foram calculados usando o método de Heun com tamanho do passo 1. Dois casos, correspondentes a números diferentes de iterações, são mostrados junto com o erro relativo porcentual absoluto.

		Iterações do Método de Heun			
		1		15	
x	$y_{verdadeiro}$	y_{Heun}	$\lvert\varepsilon_t\rvert$ (%)	y_{Heun}	$\lvert\varepsilon_t\rvert$ (%)
0	2,0000000	2,0000000	0,00	2,0000000	0,00
1	6,1946314	6,7010819	8,18	6,3608655	2,68
2	14,8439219	16,3197819	9,94	15,3022367	3,09
3	33,6771718	37,1992489	10,46	34,7432761	3,17
4	75,3389626	83,3377674	10,62	77,7350962	3,18

Observe que esse é o resultado que seria obtido pelo método de Euler padrão. O valor verdadeiro na Tabela 25.2 mostra que ele corresponde a um erro relativo porcentual de 19,3%.

Agora, para melhorar a estimativa para y_{i+1}, usamos o valor y_1^0 para prever a inclinação no final do intervalo

$$y_1' = f\left(x_1, y_1^0\right) = 4e^{0,8(1)} - 0,5(5) = 6,402164$$

que pode ser combinado com a inclinação inicial para produzir uma inclinação média no intervalo de $x = 0$ a 1

$$y' = \frac{3 + 6,402164}{2} = 4,701082$$

que está mais próxima do valor verdadeiro 4,1946. Esse resultado pode então ser substituído no corretor [Equação (25.16)] para dar uma previsão em $x = 1$

$$y_1 = 2 + 4,701082(1) = 6,701082$$

o que representa um erro relativo porcentual de $-8,18\%$. Logo, o método de Heun sem iteração do corretor reduz o valor absoluto do erro por um fator 2,4 quando comparado ao método de Euler.

Agora essa estimativa pode ser usada para refinar ou corrigir a previsão de y_1 substituindo-se o novo resultado de volta no lado direito da Equação (25.16):

$$y_1 = 2 + \frac{\left[3 + 4e^{0,8(1)} - 0,5(6,701082)\right]}{2}1 = 6,275811$$

o que representa um erro relativo porcentual de 1,31%. Esse resultado, por sua vez, pode ser substituído de volta na Equação (25.16) para posterior correção:

$$y_1 = 2 + \frac{\left[3 + 4e^{0,8(1)} - 0,5(6,275811)\right]}{2}1 = 6,382129$$

o que representa um $|\varepsilon_t|$ de 3,03%. Observe como os erros às vezes crescem conforme a iteração continua. Tais aumentos podem ocorrer especialmente para tamanhos de passo grandes e nos impedem de tirar a conclusão geral de que uma iteração adicional sempre irá melhorar o resultado. Entretanto, para um tamanho de passo suficientemente pequeno, a iteração iria eventualmente convergir para um único valor. No nosso caso, 6,360865, o que representa um erro relativo de 2,68%, obtido depois de 15 iterações. A Tabela 25.2 mostra resultados para o restante dos cálculos usando o método com 1 e 15 iterações por passo.

No exemplo anterior, a derivada é uma função tanto da variável dependente y quanto da variável independente x. Para o caso de polinômios, nos quais a EDO é apenas uma função da variável independente, o passo do preditor [Equação (25.16)] não é necessário e o corretor é aplicado apenas uma vez para cada iteração. Em tais casos, a técnica é expressa concisamente como

$$y_{i+1} = y_i + \frac{f(x_i) + f(x_{i+1})}{2}h \tag{25.18}$$

Observe a semelhança entre o lado direito da Equação (25.18) e a regra do trapézio [Equação (21.3)]. A ligação entre os dois métodos pode ser formalmente demonstrada começando com a equação diferencial ordinária

$$\frac{dy}{dx} = f(x)$$

Essa equação pode ser resolvida por integração:

$$\int_{y_i}^{y_{i+1}} dy = \int_{x_i}^{x_{i+1}} f(x)\,dx \tag{25.19}$$

o que fornece

$$y_{i+1} - y_i = \int_{x_i}^{x_{i+1}} f(x)\,dx \tag{25.20}$$

ou

$$y_{i+1} = y_i + \int_{x_i}^{x_{i+1}} f(x)\,dx \tag{25.21}$$

Agora, lembre-se, do Capítulo 21, de que a regra do trapézio [Equação (21.3)] é definida por

$$\int_{x_i}^{x_{i+1}} f(x)\,dx \cong \frac{f(x_i) + f(x_{i+1})}{2}h \tag{25.22}$$

em que $h = x_{i+1} - x_i$. Substituindo a Equação (25.22) na Equação (25.21), obtemos

$$y_{i+1} = y_i + \frac{f(x_i) + f(x_{i+1})}{2}h \tag{25.23}$$

o que é equivalente à Equação (25.18).

Como a Equação (25.23) é uma expressão direta da regra do trapézio, o erro de truncamento local é dado por [lembre-se da Equação (21.6)]

$$E_t = -\frac{f''(\xi)}{12}h^3 \tag{25.24}$$

em que ξ está entre x_i e x_{i+1}. Logo, o método é de segunda ordem porque a segunda derivada da EDO é zero quando a solução verdadeira é quadrática. Além disso, os erros local e global são $O(h^3)$ e $O(h^2)$, respectivamente. Portanto, a diminuição do tamanho do passo faz que o erro decresça a uma taxa mais rápida do que no caso do método de Euler. A Figura 25.11, que mostra o resultado do uso do método de Heun para resolver o polinômio do Exemplo 25.1, ilustra esse comportamento.

25.2.2 O Método do Ponto Médio (ou do Polígono Melhorado)

A Figura 25.12 ilustra uma outra modificação simples no método de Euler. Chamada de *método do ponto médio* (ou do *polígono melhorado* ou de *Euler modificado*), essa técnica usa o método de Euler para prever um valor de y no ponto médio do intervalo (Figura 25.12*a*):

$$y_{i+1/2} = y_i + f(x_i, y_i)\frac{h}{2} \tag{25.25}$$

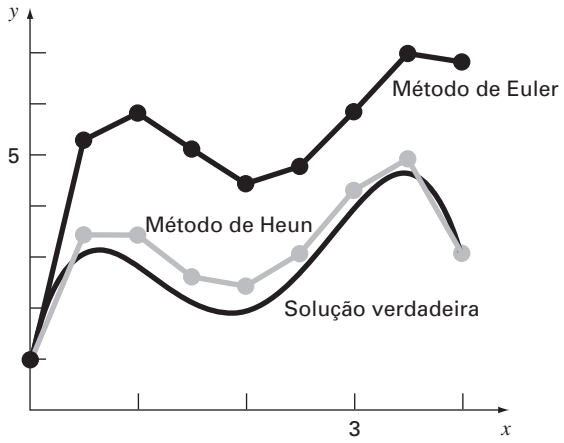


FIGURA 25.11
Comparação da solução verdadeira com soluções numéricas usando os métodos de Euler e de Heun para a integração de $y' = -2x^3 + 12x^2 - 20x + 8{,}5$.

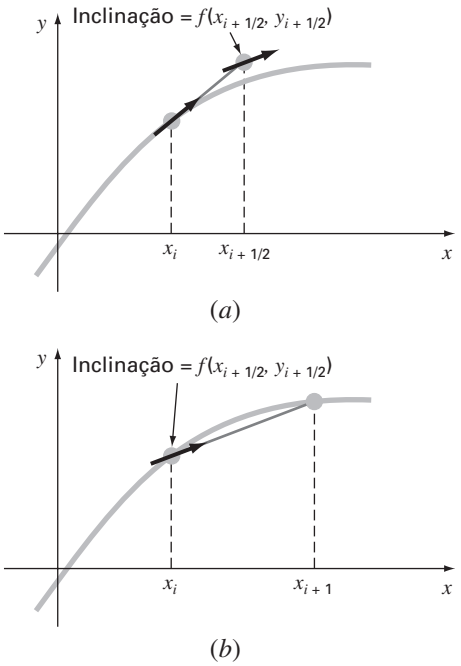


FIGURA 25.12
Descrição gráfica do método do ponto médio. (*a*) Equação (25.25) e (*b*) Equação (25.27).

Então, esse valor previsto é usado para calcular uma inclinação no ponto médio:

$$y'_{i+1/2} = f(x_{i+1/2}, y_{i+1/2}) \tag{25.26}$$

a qual supomos representar uma aproximação válida da inclinação média para o intervalo todo. Essa inclinação é então usada para extrapolar linearmente de x_i a x_{i+1} (Figura 25.12*b*):

$$y_{i+1} = y_i + f(x_{i+1/2}, y_{i+1/2})h \tag{25.27}$$

Observe que, como y_{i+1} não aparece em ambos os lados, o corretor [Equação (25.27)] não pode ser aplicado iterativamente para melhorar a solução.

Como na seção anterior, essa abordagem também pode ser ligada às fórmulas de integração de Newton-Cotes. Lembre-se, da Tabela 21.4, de que a fórmula de integração aberta de Newton-Cotes mais simples, a qual é chamada de método do ponto médio, pode ser representada por

$$\int_a^b f(x)\,dx \cong (b-a)\,f(x_1)$$

em que x_1 é o ponto médio do intervalo (a, b). Usando a nomenclatura do caso presente, ela pode ser expressa como

$$\int_{x_i}^{x_{i+1}} f(x)\,dx \cong hf(x_{i+1/2})$$

A substituição dessa fórmula na Equação (25.21) fornece a Equação (25.27). Logo, do mesmo modo como o método de Heun pode ser chamado de regra do trapézio, o *método do ponto médio* leva seu nome da fórmula de integração subjacente na qual é baseado.

O método do ponto médio é superior ao método de Euler porque utiliza uma estimativa da inclinação no ponto médio do intervalo de previsão. Lembre-se, de nossa discussão de derivação numérica na Seção 4.1.3, de que diferenças divididas finitas centradas são aproximações melhores das derivadas do que as versões tanto progressiva quanto regressiva. No mesmo sentido, uma aproximação centrada tal como a Equação (25.26) tem um erro de truncamento local de $O(h^2)$ em comparação com a aproximação progressiva do método de Euler, o qual tem um erro de $O(h)$. Conseqüentemente, os erros local e global do método do ponto médio são $O(h^3)$ e $O(h^2)$, respectivamente.

25.2.3 Algoritmos Computacionais para os Métodos de Heun e do Ponto Médio

Tanto o método de Heun com um único corretor quanto o método do ponto médio podem ser facilmente programados usando a estrutura geral descrita na Figura 25.7. Como nas Figuras 25.13*a* e *b*, rotinas simples podem ser escritas para tomar o lugar da rotina de Euler na Figura 25.7.

FIGURA 25.13
Pseudocódigo para implementar os métodos (*a*) de Heun, (*b*) do ponto médio e (*c*) de Heun iterativo.

(*a*) Método de Heun Simples sem Corretor

```
SUB Heun (x, y, h, ynew)
  CALL Derivs (x, y, dy1dx)
  ye = y + dy1dx · h
  CALL Derivs(x + h, ye, dy2dx)
  Slope = (dy1dx + dy2dx)/2
  ynew = y + Slope · h
  x = x + h
END SUB
```

(*b*) Método do Ponto Médio

```
SUB Midpoint (x, y, h, ynew)
  CALL Derivs(x, y, dydx)
  ym = y + dydx · h/2
  CALL Derivs (x + h/2, ym, dymdx)
  ynew = y + dymdx · h
  x = x + h
END SUB
```

(*c*) Método de Heun com Corretor

```
SUB HeunIter (x, y, h, ynew)
  es = 0.01
  maxit = 20
  CALL Derivs(x, y, dy1dx)
  ye = y + dy1dx · h
  iter = 0
  DO
    yeold = ye
    CALL Derivs(x + h, ye, dy2dx)
    slope = (dy1dx + dy2dx)/2
    ye = y + slope · h
    iter = iter + 1
    ea = | (ye − yeold)/ye | 100%
    IF (ea ≤ es OR iter > maxit) EXIT
  END DO
  ynew = ye
  x = x + h
END SUB
```

Entretanto, quando a versão iterativa do método de Heun é implementada, as modificações são um pouco mais complicadas (embora elas ainda se localizem em um único módulo). Desenvolvemos um pseudocódigo para esse propósito na Figura 25.13*c*. Esse algoritmo pode ser combinado com a Figura 25.7 para desenvolver um software para o método de Heun iterativo.

25.2.4 Resumo

Consertando o método de Euler, deduzimos duas novas técnicas de segunda ordem. Mesmo que essas versões exijam mais esforço computacional para determinar a inclinação, a conseqüente redução no erro nos permitirá concluir em uma seção subseqüente (Seção 25.3.4) que a melhora na acurácia em geral faz que o esforço valha a pena. Embora existam certos casos nos quais as técnicas de programação fáceis como o método de Euler podem ser usadas com vantagem, os métodos de Heun e do ponto médio são geralmente superiores e devem ser implementados se forem consistentes com os objetivos do problema.

Como observado no início desta seção, o método de Heun (sem iterações), o método do ponto médio e, na realidade, a própria técnica de Euler são versões de uma classe mais ampla de abordagens de passo único, chamadas de métodos de Runge-Kutta. Agora nos voltaremos para uma dedução formal dessas técnicas.

25.3 MÉTODOS DE RUNGE-KUTTA

Os métodos de *Runge-Kutta* (*RK*) alcançam a acurácia de uma abordagem por série de Taylor sem exigir cálculos de derivadas de ordem superior. Há muitas variações, mas todas podem ser postas na forma geral da Equação (25.1):

$$y_{i+1} = y_i + \phi(x_i, y_i, h)h \tag{25.28}$$

em que $\phi(x_i, y_i, h)$ é chamada *função incremento*, a qual pode ser interpretada como representativa da inclinação em um intervalo. A função incremento pode ser escrita na forma geral

$$\phi = a_1k_1 + a_2k_2 + \cdots + a_nk_n \tag{25.29}$$

em que os a's são constantes e os k's são

$$k_1 = f(x_i, y_i) \tag{25.29a}$$

$$k_2 = f(x_i + p_1h, y_i + q_{11}k_1h) \tag{25.29b}$$

$$k_3 = f(x_i + p_2h, y_i + q_{21}k_1h + q_{22}k_2h) \tag{25.29c}$$

$$\vdots$$

$$k_n = f(x_i + p_{n-1}h, y_i + q_{n-1,1}k_1h + q_{n-1,2}k_2h + \cdots + q_{n-1,n-1}k_{n-1}h) \tag{25.29d}$$

em que os p's e os q's são constantes. Observe que os k's são relações de recorrência. Isto é, k_1 aparece na equação para k_2, o qual aparece na equação para k_3, e assim por diante. Como cada k é um cálculo da função, essa recorrência torna os métodos RK eficientes para cálculos computacionais.

Vários tipos de métodos de Runge-Kutta podem ser deduzidos usando-se um número diferente de termos na função incremento, conforme especificado por n. Observe que o método RK de primeira ordem com $n = 1$ é, na realidade, o método de Euler. Uma vez que n seja escolhido, os valores para os a's, os p's e os q's são calculados igualando-se a Equação (25.28) a termos da expansão em série de Taylor (Quadro 25.1). Logo, pelo menos para as versões de ordem mais baixa, o número de termos, n, usualmente representa a ordem da abordagem. Por exemplo, na próxima seção, os métodos RK de segunda ordem usam uma função incremento com dois termos ($n = 2$). Tais métodos de segunda ordem serão exatos se a solução da equação diferencial for quadrática. Além

disso, como os termos com h^3 e com grau mais alto foram desprezados durante a dedução, o erro de truncamento local é $O(h^3)$ e o erro global é $O(h^2)$. Nas seções subseqüentes, serão desenvolvidos os métodos RK de terceira e quarta ordens ($n = 3$ e 4, respectivamente), casos nos quais os erros de truncamento globais são $O(h^3)$ e $O(h^4)$, respectivamente.

25.3.1 Métodos de Runge-Kutta de Segunda Ordem

A versão de segunda ordem da Equação (25.28) é

$$y_{i+1} = y_i + (a_1k_1 + a_2k_2)h \tag{25.30}$$

em que

$$k_1 = f(x_i, y_i) \tag{25.30a}$$

$$k_2 = f(x_i + p_1h, y_i + q_{11}k_1h) \tag{25.30b}$$

Como descrito no Quadro 25.1, os valores a_1, a_2, p_1 e q_{11} são calculados igualando-se a Equação (25.30) à expansão em série de Taylor até os termos de 2º grau. Fazendo isso, deduzimos três equações para calcular as quatro constantes desconhecidas. As três equações são

$$a_1 + a_2 = 1 \tag{25.31}$$

$$a_2p_1 = \frac{1}{2} \tag{25.32}$$

$$a_2q_{11} = \frac{1}{2} \tag{25.33}$$

Quadro 25.1 Dedução dos Métodos de Runge-Kutta de Segunda Ordem

A versão de segunda ordem da Equação (25.28) é

$$y_{i+1} = y_i + (a_1k_1 + a_2k_2)h \tag{Q25.1.1}$$

em que

$$k_1 = f(x_i, y_i) \tag{Q25.1.2}$$

e

$$k_2 = f(x_i + p_1h, y_i + q_{11}k_1h) \tag{Q25.1.3}$$

Para usar a Equação (Q25.1.1), temos de determinar valores para as constantes a_1, a_2, p_1 e q_{11}. Para fazer isso, lembramos que a série de Taylor de segundo grau para y_{i+1} em termos de y_i e $f(x_i, y_i)$ é escrita como [Equação (25.11)]

$$y_{i+1} = y_i + f(x_i, y_i)h + \frac{f'(x_i, y_i)}{2!}h^2 \tag{Q25.1.4}$$

em que $f'(x_i, y_i)$ deve ser determinada pela regra da cadeia (Seção 25.1.3):

$$f'(x_i, y_i) = \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y}\frac{dy}{dx} \tag{Q25.1.5}$$

Substituindo a Equação (Q25.1.5) na Equação (Q25.1.4), obtemos

$$y_{i+1} = y_i + f(x_i, y_i)h + \left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}\right)\frac{h^2}{2!} \tag{Q25.1.6}$$

A estratégia básica por trás dos métodos de Runge-Kutta é o uso de manipulações algébricas para determinar os valores de a_1, a_2, p_1 e q_{11} que tornam as Equações (Q25.1.1) e (Q25.1.6) equivalentes.

Para fazer isso, primeiro usamos uma série de Taylor para expandir a Equação (Q25.1.3). A série de Taylor para uma função de duas variáveis é definida por [lembre-se da Equação (4.26)]

$$g(x + r, y + s) = g(x, y) + r\frac{\partial g}{\partial x} + s\frac{\partial g}{\partial y} + \cdots$$

Aplicando esse método para expandir a Equação (Q25.1.3), obtemos

$$f(x_i + p_1h, y_i + q_{11}k_1h) = f(x_i, y_i) + p_1h\frac{\partial f}{\partial x} + q_{11}k_1h\frac{\partial f}{\partial y} + O(h^2)$$

Esse resultado pode ser substituído junto com a Equação (Q25.1.2) na Equação (Q25.1.1) para fornecer

$$y_{i+1} = y_i + a_1hf(x_i, y_i) + a_2hf(x_i, y_i) + a_2p_1h^2\frac{\partial f}{\partial x} + a_2q_{11}h^2f(x_i, y_i)\frac{\partial f}{\partial y} + O(h^3)$$

ou, reagrupando os termos,

$$y_{i+1} = y_i + [a_1f(x_i, y_i) + a_2f(x_i, y_i)]h + \left[a_2p_1\frac{\partial f}{\partial x} + a_2q_{11}f(x_i, y_i)\frac{\partial f}{\partial y}\right]h^2 + O(h^3) \tag{Q25.1.7}$$

Agora, comparando os termos semelhantes nas Equações (Q25.1.6) e (Q25.1.7), determinamos que, para as duas equações serem equivalentes, deve valer o seguinte:

$$a_1 + a_2 = 1$$

$$a_2p_1 = \frac{1}{2}$$

$$a_2q_{11} = \frac{1}{2}$$

Essas três equações simultâneas contêm as quatro constantes desconhecidas. Como há uma incógnita a mais do que o número de equações, não existe um único conjunto de constantes que satisfaça as equações. Entretanto, escolhendo um valor arbitrário para uma dessas constantes, podemos determinar as outras três. Conseqüentemente, existe uma família de métodos de segunda ordem e não uma única versão.

Como temos três equações com quatro incógnitas, devemos escolher um valor para uma das incógnitas para determinar as outras três. Suponha que especifiquemos um valor para a_2. Então, as Equações (25.31) a (25.33) podem ser resolvidas simultaneamente por

$$a_1 = 1 - a_2 \tag{25.34}$$

$$p_1 = q_{11} = \frac{1}{2a_2} \tag{25.35}$$

Como podemos escolher um número infinito de valores para a_2, existe um número infinito de métodos RK de segunda ordem. Cada versão forneceria exatamente o mesmo resultado se a solução da EDO fosse quadrática, linear ou constante. Entretanto, elas fornecem resultados diferentes quando (como é o caso típico) a solução é mais complicada. Apresentamos três das versões usadas mais freqüentemente e preferidas:

Método de Heun com um Único Corretor ($a_2 = 1/2$). Se for suposto que a_2 é $1/2$, as Equações (25.34) e (25.35) podem ser resolvidas por $a_1 = 1/2$ e $p_1 = q_{11} = 1$. Esses parâmetros, quando substituídos na Equação (25.30), fornecem

$$y_{i+1} = y_i + \left(\frac{1}{2}k_1 + \frac{1}{2}k_2\right)h \tag{25.36}$$

em que

$$k_1 = f(x_i, y_i) \tag{25.36a}$$

$$k_2 = f(x_i + h, y_i + k_1 h) \tag{25.36b}$$

Observe que k_1 é a inclinação no início do intervalo e k_2 é a inclinação no final do intervalo. Conseqüentemente, esse método de segunda ordem de Runge-Kutta é, na realidade, a técnica de Heun sem iteração.

O Método do Ponto Médio ($a_2 = 1$). Se for suposto que a_2 é 1, então $a_1 = 0$, $p_1 = q_{11} = 1/2$, e a Equação (25.30) se torna

$$y_{i+1} = y_i + k_2 h \tag{25.37}$$

em que

$$k_1 = f(x_i, y_i) \tag{25.37a}$$

$$k_2 = f\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1 h\right) \tag{25.37b}$$

Esse é o método do ponto médio.

Método de Ralston ($a_2 = 2/3$). Ralston (1962) e Ralston e Rabinowitz (1978) determinaram que a escolha $a_2 = 2/3$ fornece um limitante mínimo para o erro de truncamento para os algoritmos RK de segunda ordem. Para essa versão, $a_1 = 1/3$ e $p_1 = q_{11} = 3/4$, e obtém-se

$$y_{i+1} = y_i + \left(\frac{1}{3}k_1 + \frac{2}{3}k_2\right)h \tag{25.38}$$

em que

$$k_1 = f(x_i, y_i) \tag{25.38a}$$

$$k_2 = f\left(x_i + \frac{3}{4}h, y_i + \frac{3}{4}k_1 h\right) \tag{25.38b}$$

EXEMPLO 25.6

Comparação de Diversos Esquemas RK de Segunda Ordem

Enunciado do Problema. Use o método do ponto médio [(Equação (25.37)] e o método de Ralston [Equação (25.38)] para integrar numericamente a Equação (PT7.13)

$$f(x, y) = -2x^3 + 12x^2 - 20x + 8{,}5$$

de $x = 0$ a $x = 4$ usando um tamanho de passo de 0,5. A condição inicial em $x = 0$ é $y = 1$. Compare os resultados com os valores obtidos usando um outro algoritmo RK de segunda ordem, isto é, o método de Heun sem iteração do corretor (Tabela 25.3).

Solução. O primeiro passo no método do ponto médio é usar a Equação (25.37*a*) para calcular

$$k_1 = -2(0)^3 + 12(0)^2 - 20(0) + 8{,}5 = 8{,}5$$

Entretanto, como a EDO é uma função apenas de x, esse resultado não terá influência no segundo passo — o uso da Equação (25.37*b*) para calcular

$$k_2 = -2(0{,}25)^3 + 12(0{,}25)^2 - 20(0{,}25) + 8{,}5 = 4{,}21875$$

Observe que essa estimativa da inclinação está muito mais perto do valor médio no intervalo (4,4375) do que a inclinação no início do intervalo (8,5) que teria sido usada na abordagem de Euler. A inclinação no ponto médio pode ser substituída na Equação (25.37) para prever

$$y(0{,}5) = 1 + 4{,}21875(0{,}5) = 3{,}109375 \qquad \varepsilon_t = 3{,}4\%$$

O cálculo é repetido e os resultados estão resumidos na Figura 25.14 e na Tabela 25.3.

Para o método de Ralston, k_1 para o primeiro intervalo também é igual a 8,5 e [Equação (25.38*b*)]

$$k_2 = -2(0{,}375)^3 + 12(0{,}375)^2 - 20(0{,}375) + 8{,}5 = 2{,}58203125$$

A inclinação média é calculada por

$$\phi = \frac{1}{3}(8{,}5) + \frac{2}{3}(2{,}58203125) = 4{,}5546875$$

FIGURA 25.14
Comparação da solução verdadeira com soluções numéricas usando três métodos RK de segunda ordem e o método de Euler.

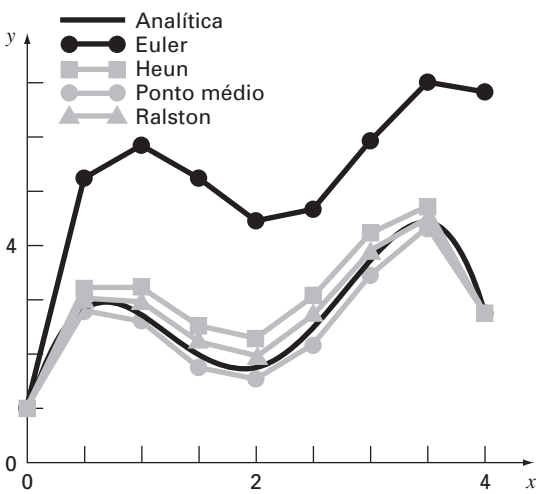


TABELA 25.3 Comparação dos valores verdadeiro e aproximados da integral de $y' = -2x^3 + 12x^2 - 20x + 8{,}5$, com a condição inicial $y = 1$ em $x = 0$. Os valores aproximados foram calculados usando três versões dos métodos RK de segunda ordem, com um tamanho de passo de 0,5.

		Heun		Ponto médio		RK de Segunda Ordem de Ralston	
x	$y_{verdadeiro}$	y	$\lvert\varepsilon_t\rvert$ (%)	y	$\lvert\varepsilon_t\rvert$ (%)	y	$\lvert\varepsilon_t\rvert$ (%)
0,0	1,00000	1,00000	0	1,00000	0	1,00000	0
0,5	3,21875	3,43750	6,8	3,109375	3,4	3,277344	1,8
1,0	3,00000	3,37500	12,5	2,81250	6,3	3,101563	3,4
1,5	2,21875	2,68750	21,1	1,984375	10,6	2,347656	5,8
2,0	2,00000	2,50000	25,0	1,75	12,5	2,140625	7,0
2,5	2,71875	3,18750	17,2	2,484375	8,6	2,855469	5,0
3,0	4,00000	4,37500	9,4	3,81250	4,7	4,117188	2,9
3,5	4,71875	4,93750	4,6	4,609375	2,3	4,800781	1,7
4,0	3,00000	3,00000	0	3	0	3,031250	1,0

a qual pode ser usada para prever

$$y(0{,}5) = 1 + 4{,}5546875(0{,}5) = 3{,}27734375 \qquad \varepsilon_t = -1{,}82\%$$

O cálculo é repetido e os resultados estão resumidos na Figura 25.14 e na Tabela 25.3. Observe como todos os métodos RK de segunda ordem são superiores ao método de Euler.

25.3.2 Métodos de Runge-Kutta de Terceira Ordem

Para $n = 3$, a dedução é parecida com a dos métodos de segunda ordem. O resultado dessa dedução é um conjunto de seis equações com oito incógnitas. Portanto, devem ser especificados valores *a priori* para duas das incógnitas para determinar os parâmetros restantes. Uma versão comum que resulta é

$$y_{i+1} = y_i + \frac{1}{6}(k_1 + 4k_2 + k_3)h \tag{25.39}$$

em que

$$k_1 = f(x_i, y_i) \tag{25.39a}$$

$$k_2 = f\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1 h\right) \tag{25.39b}$$

$$k_3 = f(x_i + h, y_i - k_1 h + 2k_2 h) \tag{25.39c}$$

Observe que, se a derivada for uma função apenas de x, o método de terceira ordem se reduz à regra 1/3 de Simpson. Ralston (1962) e Ralston e Rabinowitz (1978) desenvolveram uma versão alternativa que fornece um limitante mínimo para o erro de truncamento. De qualquer modo, os métodos RK de terceira ordem têm erros local e global de $O(h^4)$ e $O(h^3)$, respectivamente, e fornecem resultados exatos se a solução for uma função cúbica. Tratando-se de polinômios, a Equação (25.39) também será exata quando a equação diferencial for uma função cúbica e a solução for de quarto grau, porque a regra 1/3 de Simpson fornece estimativas exatas da integral para funções cúbicas (lembre-se do Quadro 21.3).

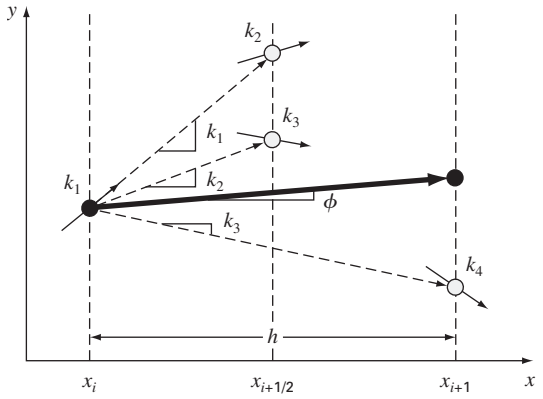


FIGURA 25.15
Descrição gráfica das estimativas de inclinação envolvidas no método RK de quarta ordem.

25.3.3 Métodos de Runge-Kutta de Quarta Ordem

Os métodos RK mais populares são os de quarta ordem. Como no caso das abordagens de segunda ordem, existe um número infinito de versões. A seguinte é a forma mais comumente usada e, assim, a chamamos *método RK de quarta ordem clássico*:

$$y_{i+1} = y_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)h \tag{25.40}$$

em que

$$k_1 = f(x_i, y_i) \tag{25.40a}$$

$$k_2 = f\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1 h\right) \tag{25.40b}$$

$$k_3 = f\left(x_i + \frac{1}{2}h, y_i + \frac{1}{2}k_2 h\right) \tag{25.40c}$$

$$k_4 = f(x_i + h, y_i + k_3 h) \tag{25.40d}$$

Observe que para EDOs que são uma função apenas de x, o método RK de quarta ordem clássico é parecido com a regra 1/3 de Simpson. Além disso, o método RK de quarta ordem é parecido com a abordagem de Heun no fato que são desenvolvidas múltiplas estimativas da inclinação para se chegar a uma inclinação média melhorada no intervalo. Como descrito na Figura 25.15, cada um dos k's representa uma inclinação. Então, a Equação (25.40) representa uma média ponderada delas para chegar a uma inclinação melhorada.

EXEMPLO 25.7 Método RK de Quarta Ordem Clássico

Enunciado do Problema.

(a) Use o método RK de quarta ordem clássico [Equação (25.40)] para integrar
$f(x, y) = -2x^3 + 12x^2 - 20x + 8{,}5$
utilizando um tamanho de passo de $h = 0{,}5$ e uma condição inicial de $y = 1$ em $x = 0$.

(b) Analogamente, integre
$f(x, y) = 4e^{0{,}8x} - 0{,}5y$
usando $h = 0{,}5$ com $y(0) = 2$ de $x = 0$ a 0,5.

Solução.

(a) As Equações (25.40*a*) a (25.40*d*) são usadas para calcular $k_1 = 8{,}5$, $k_2 = 4{,}21875$, $k_3 = 4{,}21875$ e $k_4 = 1{,}25$, que são substituídos na Equação (25.40) para fornecer

$$y(0{,}5) = 1 + \left\{\frac{1}{6}[8{,}5 + 2(4{,}21875) + 2(4{,}21875) + 1{,}25]\right\} 0{,}5$$
$$= 3{,}21875$$

o que é exato. Logo, como a solução verdadeira é de quarto grau [Equação (PT7.16)], o método de quarta ordem dá resultados exatos.

(b) Nesse caso, a inclinação no início do intervalo é calculada como

$$k_1 = f(0{,}2) = 4e^{0{,}8(0)} - 0{,}5(2) = 3$$

Esse valor é usado para calcular um valor de y e uma inclinação no ponto médio,

$$y(0{,}25) = 2 + 3(0{,}25) = 2{,}75$$

$$k_2 = f(0{,}25, 2{,}75) = 4e^{0{,}8(0{,}25)} - 0{,}5(2{,}75) = 3{,}510611$$

Essa inclinação, por sua vez, é usada para calcular um outro valor de y e uma outra inclinação no ponto médio,

$$y(0{,}25) = 2 + 3{,}510611(0{,}25) = 2{,}877653$$

$$k_3 = f(0{,}25, 2{,}877653) = 4e^{0{,}8(0{,}25)} - 0{,}5(2{,}877653) = 3{,}446785$$

A seguir, essa inclinação é usada para calcular um valor de y e uma inclinação no final do intervalo,

$$y(0{,}5) = 2 + 3{,}071785(0{,}5) = 3{,}723392$$

$$k_4 = f(0{,}5, 3{,}723392) = 4e^{0{,}8(0{,}5)} - 0{,}5(3{,}723392) = 4{,}105603$$

Finalmente, as quatro estimativas de inclinação são combinadas para fornecer uma inclinação média. Essa inclinação média é então usada para fazer a previsão final no fim do intervalo,

$$\phi = \frac{1}{6}[3 + 2(3{,}510611) + 2(3{,}446785) + 4{,}105603] = 3{,}503399$$

$$y(0{,}5) = 2 + 3{,}503399(0{,}5) = 3{,}751699$$

o que se compara favoravelmente com a solução verdadeira, de 3,751521.

25.3.4 Métodos de Runge-Kutta de Ordem Superior

Quando resultados mais acurados forem necessários, o *método RK de quinta ordem de Butcher* (1964) é recomendado:

$$y_{i+1} = y_i + \frac{1}{90}(7k_1 + 32k_3 + 12k_4 + 32k_5 + 7k_6)h \tag{25.41}$$

em que

$$k_1 = f(x_i, y_i) \tag{25.41a}$$

$$k_2 = f\left(x_i + \frac{1}{4}h, y_i + \frac{1}{4}k_1 h\right) \tag{25.41b}$$

$$k_3 = f\left(x_i + \frac{1}{4}h, y_i + \frac{1}{8}k_1 h + \frac{1}{8}k_2 h\right) \tag{25.41c}$$

$$k_4 = f\left(x_i + \frac{1}{2}h,\, y_i - \frac{1}{2}k_2 h + k_3 h\right) \tag{25.41d}$$

$$k_5 = f\left(x_i + \frac{3}{4}h,\, y_i + \frac{3}{16}k_1 h + \frac{9}{16}k_4 h\right) \tag{25.41e}$$

$$k_6 = f\left(x_i + h,\, y_i - \frac{3}{7}k_1 h + \frac{2}{7}k_2 h + \frac{12}{7}k_3 h - \frac{12}{7}k_4 h + \frac{8}{7}k_5 h\right) \tag{25.41f}$$

Observe a similaridade entre o método de Butcher e a regra de Boole na Tabela 21.2. Fórmulas RK de ordem superior como o método de Butcher estão disponíveis, mas, em geral, além dos métodos de quarta ordem, o ganho em acurácia é contrabalançado pelo esforço e pela complexidade computacionais adicionais.

EXEMPLO 25.8 Comparação dos Métodos de Runge-Kutta

Enunciado do Problema. Use métodos RK de primeira a quinta ordem para resolver

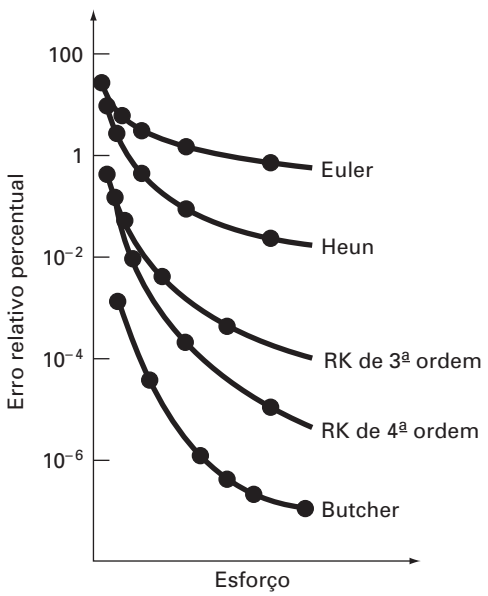
$$f(x, y) = 4e^{0,8x} - 0,5y$$

com $y(0) = 2$ de $x = 0$ a $x = 4$ com diversos tamanhos de passo. Compare a acurácia dos diversos métodos para o resultado em $x = 4$ com base na resposta exata $y(4) = 75,33896$.

Solução. Os cálculos são feitos usando os métodos de Euler, não-iterativo de Heun, RK de terceira ordem [Equação (25.39)], RK de quarta ordem clássico e RK de quinta ordem de Butcher. Os resultados estão apresentados na Figura 25.16, em que marcamos os valores absolutos do erro porcentual relativo em função do esforço computacional. Essa última quantidade é equivalente ao número de cálculos de função necessário para obter o resultado, como em

$$\text{Esforço} = n_f \frac{b - a}{h} \tag{E25.8.1}$$

FIGURA 25.16
Comparação do erro relativo porcentual *versus* o esforço computacional para os métodos RK de primeira a quinta ordem.



em que n_f é o número de cálculos de função envolvidos no cálculo RK particular. Para ordens ≤ 4, n_f é igual à ordem do método. Entretanto, observe que a técnica de quinta ordem de Butcher exige seis cálculos de função [Equações (25.41*a*) a (25.41*f*)]. A quantidade $(b - a)/h$ é o intervalo de integração total dividido pelo tamanho do passo — ou seja, é o número de aplicações da técnica RK necessário para se obter o resultado. Logo, como em geral os cálculos de função são os passos principais no gasto de tempo, a Equação (E25.8.1) fornece uma medida grosseira do tempo de execução necessário para se obter a resposta.

Uma inspeção da Figura 25.16 nos leva a diversas conclusões: primeiro, que os métodos de ordem superior atingem uma acurácia melhor para o mesmo esforço computacional e, segundo, que o ganho em acurácia para o esforço adicional tende a diminuir depois de um ponto. (Observe que as curvas caem rapidamente no início e depois tendem a se nivelar.)

O Exemplo 25.8 e a Figura 25.16 poderiam levar-nos a concluir que as técnicas RK de ordem superior são sempre os métodos preferidos. Contudo, outros fatores, como custo de programação e exigências de acurácia do problema, também devem ser considerados na escolha de uma técnica de solução. Tais prós e contras serão explorados em detalhe nas aplicações em engenharia no Capítulo 28 e no Epílogo da Parte Sete.

25.3.5 Algoritmos Computacionais para os Métodos de Runge-Kutta

Como todos os métodos abordados neste capítulo, as técnicas RK se ajustam bem ao algoritmo geral expresso na Figura 25.7. A Figura 25.17 apresenta um pseudocódigo para determinar a inclinação do método RK de quarta ordem clássico [Equação (25.40)]. Subrotinas para calcular as inclinações para todas as outras versões podem ser facilmente programadas de um modo parecido.

25.4 SISTEMAS DE EQUAÇÕES

Muitos problemas práticos em engenharia e em ciência exigem a solução de um sistema simultâneo de equações diferenciais ordinárias em vez de uma única equação. Tais sistemas podem ser representados de forma geral por

FIGURA 25.17
Pseudocódigo para determinar um único passo do método RK de quarta ordem.

```
SUB RK4 (x, y, h, ynew)
  CALL Derivs(x, y, k1)
  ym = y + k1 · h/2
  CALL Derivs(x + h/2, ym, k2)
  ym = y + k2 · h/2
  CALL Derivs(x + h/2, ym, k3)
  ye = y + k3 · h
  CALL Derivs(x + h, ye, k4)
  slope = (k1 + 2(k2 + k3) + k4)/6
  ynew = y + slope · h
  x = x + h
END SUB
```

$$\frac{dy_1}{dx} = f_1(x, y_1, y_2, \ldots, y_n)$$
$$\frac{dy_2}{dx} = f_2(x, y_1, y_2, \ldots, y_n)$$
$$\vdots$$
$$\frac{dy_n}{dx} = f_n(x, y_1, y_2, \ldots, y_n) \tag{25.42}$$

A solução de tal sistema exige que sejam conhecidas n condições iniciais no valor inicial de x.

25.4.1 O Método de Euler

Todos os métodos discutidos neste capítulo para uma única equação podem ser estendidos ao sistema que acabamos de mostrar. As aplicações em engenharia podem envolver milhares de equações simultâneas. Em cada caso, o procedimento para resolver um sistema de equações envolve simplesmente a aplicação da técnica de passo único em todas as equações para cada passo, antes de prosseguir para o próximo passo. Isso é mais bem ilustrado pelo próximo exemplo para o método de Euler, que é simples.

EXEMPLO 25.9 Resolvendo Sistemas de EDOs Usando o Método de Euler

Enunciado do Problema. Resolva o seguinte conjunto de equações diferenciais usando o método de Euler, supondo que, em $x = 0$, $y_1 = 4$ e $y_2 = 6$. Integre até $x = 2$ com um tamanho de passo de 0,5.

$$\frac{dy_1}{dx} = -0{,}5y_1 \qquad \frac{dy_2}{dx} = 4 - 0{,}3y_2 - 0{,}1y_1$$

Solução. O método de Euler é implementado para cada variável como na Equação (25.2):

$$y_1(0{,}5) = 4 + [-0{,}5(4)]0{,}5 = 3$$

$$y_2(0{,}5) = 6 + [4 - 0{,}3(6) - 0{,}1(4)]0{,}5 = 6{,}9$$

Observe que $y_1(0) = 4$ é usado na segunda equação em vez de $y_1(0{,}5) = 3$ calculado na primeira equação. Procedendo dessa maneira, obtemos

x	y_1	y_2
0	4	6
0,5	3	6,9
1,0	2,25	7,715
1,5	1,6875	8,44525
2,0	1,265625	9,094087

25.4.2 Métodos de Runge-Kutta

Observe que qualquer um dos métodos RK de ordem superior deste capítulo pode ser aplicado para sistemas de equações — mas é preciso tomar cuidado na determinação das inclinações. A Figura 25.15 ajuda na visualização da maneira adequada de fazer isso no método de quarta ordem. Ou seja, primeiro desenvolvemos inclinações para todas as variáveis no valor inicial. Essas inclinações (um conjunto de k_1's) são então usadas para

fazer previsões da variável independente no ponto médio do intervalo. Os valores no ponto médio são, por sua vez, usados para calcular um conjunto de inclinações nos pontos médios (os k_2's). Essas novas inclinações são então levadas de volta para o ponto inicial para fazer um outro conjunto de previsões no ponto médio que leva a novas previsões de inclinações nos pontos médios (os k_3's). Estes então são usados para fazer previsões no ponto final do intervalo, que são utilizadas para desenvolver inclinações no final do intervalo (os k_4's). Finalmente, os k's são combinados em um conjunto de funções incremento [como na Equação (25.40)] e trazidos de volta para o ponto inicial para fazer a previsão final. O exemplo a seguir ilustra a abordagem.

EXEMPLO 25.10 Resolvendo Sistemas de EDOs Usando o Método RK de Quarta Ordem

Enunciado do Problema. Use os métodos RK de quarta ordem para resolver as EDOs do Exemplo 25.9.

Solução. Primeiro, precisamos determinar todas as inclinações no início do intervalo:

$$k_{1,1} = f_1(0, 4, 6) = -0{,}5(4) = -2$$

$$k_{1,2} = f_2(0, 4, 6) = 4 - 0{,}3(6) - 0{,}1(4) = 1{,}8$$

em que $k_{i,j}$ é o i-ésimo valor de k para a j-ésima variável dependente. A seguir, devemos calcular os primeiros valores de y_1 e y_2 no ponto médio:

$$y_1 + k_{1,1}\frac{h}{2} = 4 + (-2)\frac{0{,}5}{2} = 3{,}5$$

$$y_2 + k_{1,2}\frac{h}{2} = 6 + (1{,}8)\frac{0{,}5}{2} = 6{,}45$$

o que pode ser usado para calcular o primeiro conjunto de inclinações no ponto médio,

$$k_{2,1} = f_1(0{,}25, 3{,}5, 6{,}45) = -1{,}75$$

$$k_{2,2} = f_2(0{,}25, 3{,}5, 6{,}45) = 1{,}715$$

Esses são usados para determinar o segundo conjunto de previsões para o ponto médio,

$$y_1 + k_{2,1}\frac{h}{2} = 4 + (-1{,}75)\frac{0{,}5}{2} = 3{,}5625$$

$$y_2 + k_{2,2}\frac{h}{2} = 6 + (1{,}715)\frac{0{,}5}{2} = 6{,}42875$$

que podem ser utilizados para calcular o segundo conjunto de inclinações no ponto médio,

$$k_{3,1} = f_1(0{,}25, 3{,}5625, 6{,}42875) = -1{,}78125$$

$$k_{3,2} = f_2(0{,}25, 3{,}5625, 6{,}42875) = 1{,}715125$$

Esses são usados para determinar as previsões no final do intervalo

$$y_1 + k_{3,1}h = 4 + (-1{,}78125)(0{,}5) = 3{,}109375$$

$$y_2 + k_{3,2}h = 6 + (1{,}715125)(0{,}5) = 6{,}857563$$

que podem ser utilizados para calcular as inclinações na extremidade final,

$$k_{4,1} = f_1(0{,}5, 3{,}109375, 6{,}857563) = -1{,}554688$$

$$k_{4,2} = f_2(0{,}5, 3{,}109375, 6{,}857563) = 1{,}631794$$

Os valores de k podem então ser usados para calcular [Equação (25.40)]:

$$y_1(0,5) = 4 + \frac{1}{6}[-2 + 2(-1,75 - 1,78125) - 1,554688]0,5 = 3,115234$$

$$y_2(0,5) = 6 + \frac{1}{6}[1,8 + 2(1,715 + 1,715125) + 1,631794]0,5 = 6,857670$$

Procedendo de forma análoga nos próximos passos, obtemos

x	y_1	y_2
0	4	6
0,5	3,115234	6,857670
1,0	2,426171	7,632106
1,5	1,889523	8,326886
2,0	1,471577	8,946865

25.4.3 Algoritmos Computacionais para a Resolução de Sistemas de EDOs

O código computacional para resolver uma única EDO pelo método de Euler (Figura 25.7) pode ser facilmente estendido para sistemas de equações. As modificações incluem:

1. Entrar o número de equações, n.
2. Entrar os valores iniciais de cada uma das n variáveis dependentes.
3. Modificar o algoritmo de modo que ele calcule inclinações para cada uma das variáveis dependentes.
4. Incluir equações adicionais para calcular os valores da derivada para cada uma das EDOs.
5. Incluir laços para calcular um novo valor para cada variável dependente.

Tal algoritmo está delineado na Figura 25.18 para o método RK de quarta ordem. Observe quão parecido ele é na estrutura e organização com a Figura 25.7. A maior parte das diferenças se relaciona ao fato de que

1. Existem n equações.
2. Há detalhes adicionais do método RK de quarta ordem.

EXEMPLO 25.11 Resolvendo um Sistema de EDOs com o Computador

Enunciado do Problema. Um programa de computador para implementar o método RK de quarta ordem para sistemas pode ser desenvolvido facilmente com base na Figura 25.18. Tal software torna conveniente a comparação de diferentes modelos para um sistema físico. Por exemplo, um modelo linear para o pêndulo oscilante é dado por [lembre-se da Equação (PT7.11)]

$$\frac{dy_1}{dx} = y_2 \qquad \frac{dy_2}{dx} = -16,1y_1$$

em que y_1 e y_2 são o deslocamento angular e a velocidade, respectivamente. Um modelo não-linear para o mesmo sistema é [lembre-se da Equação (PT7.9)]

$$\frac{dy_3}{dx} = y_4 \qquad \frac{dy_4}{dx} = -16,1 \operatorname{sen}(y_3)$$

em que y_3 e y_4 são o deslocamento angular e a velocidade para o caso não-linear. Resolva esses sistemas para dois casos: (a) um pequeno deslocamento inicial ($y_1 = y_3 = 0,1$ radianos; $y_2 = y_4 = 0$) e (b) um grande deslocamento ($y_1 = y_3 = \pi/4 = 0,785398$ radianos; $y_2 = y_4 = 0$).

(a) Programa Principal

```
Atribua valores para
n = número de equações
yi = valores iniciais das n
     variáveis dependentes
xi = valor inicial da variável
     independente
xf = valor final da variável
     independente
dx = tamanho do passo de cálculo
xout = intervalo de saída

x = xi
m = 0
xp_m = x
DOFOR i = 1, n
  yp_i,m = yi_i
END DO
DO
  xend = x + xout
  IF (xend > xf) THEN xend = xf
  h = dx
  CALL Integrator (x, y, n, h, xend)
  m = m + 1
  xp_m = x
  DOFOR i = 1, n
    yp_i,m = y_i
  END DO
  IF (x ≥ xf) EXIT
LOOP
DISPLAY RESULTS
END
```

(b) Rotina para Fazer um Passo de Saída

```
SUB Integrator (x, y, n, h, xend)
  DO
    IF (xend − x < h) THEN h = xend − x
    CALL RK4 (x, y, n, h)
    IF (x ≥ xend) EXIT
  END DO
END SUB
```

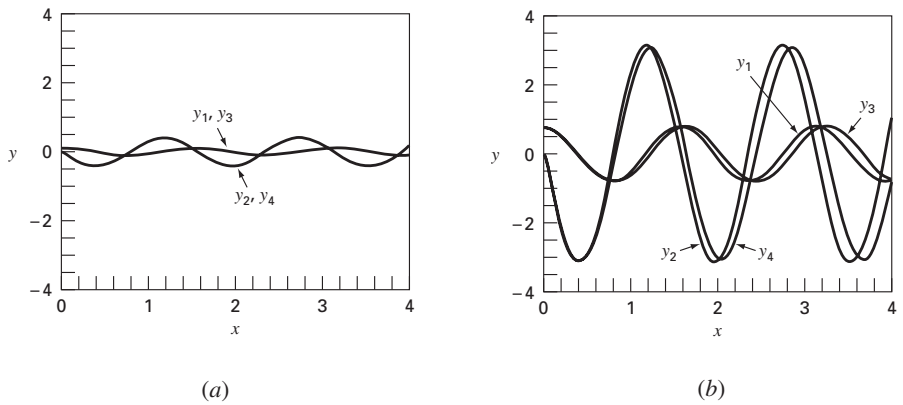
(c) Método RK de Quarta Ordem para um Sistema de EDOs

```
SUB RK4 (x, y, n, h)
  CALL Derivs (x, y, k1)
  DOFOR i = 1, n
    ym_i = y_i + k1_i * h / 2
  END DO
  CALL Derivs (x + h / 2, ym, k2)
  DOFOR i = 1, n
    ym_i = y_i + k2_i * h / 2
  END DO
  CALL Derivs (x + h / 2, ym, k3)
  DOFOR i = 1, n
    ye_i = y_i + k3_i * h
  END DO
  CALL Derivs (x + h, ye, k4)
  DOFOR i = 1, n
    slope_i = (k1_i + 2*(k2_i+k3_i)+k4_i)/6
    y_i = y_i + slope_i * h
  END DO
  x = x + h
END SUB
```

(d) Rotina para Determinar Derivadas

```
SUB Derivs (x, y, dy)
  dy_1 = ...
  dy_2 = ...
END SUB
```

FIGURA 25.18
Pseudocódigo para o método RK de quarta ordem para sistemas.



(*a*) (*b*)

FIGURA 25.19
Soluções obtidas com um programa computacional para o método RK de quarta ordem. Os gráficos representam soluções tanto para o pêndulo linear quanto para o não-linear com deslocamentos (*a*) pequenos e (*b*) grandes.

Solução.

(a) Os resultados calculados para os modelos linear e não-linear são quase idênticos (Figura 25.19*a*). Isso é como esperado, pois, quando o deslocamento inicial é pequeno, sen $(\theta) \cong \theta$.

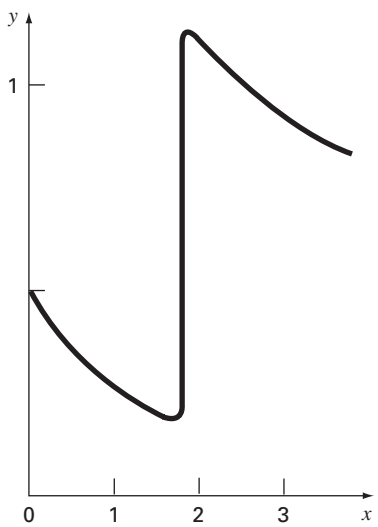
(b) Quando o deslocamento inicial é $\pi/4 = 0{,}785398$, as soluções são muito diferentes e a diferença aumenta conforme o tempo se torna cada vez maior (Figura 25.19*b*). Isso é esperado porque a hipótese de que sen $(\theta) = \theta$ é ruim quando theta é grande.

25.5 MÉTODOS DE RUNGE-KUTTA ADAPTATIVOS

Até este ponto, apresentamos métodos para resolver EDOs que empregam um tamanho de passo constante. Para um número significativo de problemas, isso pode representar uma limitação séria. Por exemplo, suponha que estejamos integrando uma EDO com uma solução do tipo descrito na Figura 25.20. Para a maior parte do intervalo, a solução varia gradualmente. Tal comportamento sugere que um tamanho de passo razoavelmente grande poderia ser utilizado para obter resultados adequados. Entretanto, para uma região localizada de $x = 1{,}75$ a $x = 2{,}25$, a solução passa por uma variação abrupta. A conseqüência prática de lidar com tais funções é que seria necessário um tamanho de passo muito pequeno para capturar adequadamente o comportamento impulsivo. Se fosse utilizado um algoritmo com tamanho de passo constante, o pequeno tamanho de passo necessário para a região de variação abrupta teria de ser aplicado em todo o cálculo. Como conseqüência, um tamanho de passo muito menor do que o necessário — e, portanto, muito mais cálculos — seria desperdiçado na região de variação gradual.

Os algoritmos que ajustam automaticamente o tamanho do passo podem evitar tais exageros e, portanto, apresentam grande vantagem. Como eles se "adaptam" à trajetória da solução, dizemos que têm *controle adaptativo do tamanho do passo*. A implementação dessas abordagens exige que uma estimativa do erro de truncamento local seja obtida em cada passo. Então, essa estimativa de erro pode servir de base ou para aumentar ou para diminuir o tamanho do passo.

FIGURA 25.20
Um exemplo de uma solução de uma EDO que exibe uma variação abrupta. Ajuste automático no tamanho do passo tem grandes vantagens nesses casos.



Antes de continuar, devemos mencionar que, além de resolver EDOs, os métodos descritos neste capítulo também podem ser usados para calcular integrais definidas. Como mencionado anteriormente na introdução da Parte Seis, o cálculo da integral

$$I = \int_a^b f(x)\,dx$$

é equivalente à resolução da equação diferencial

$$\frac{dy}{dx} = f(x)$$

para determinar $y(b)$ dada a condição inicial $y(a) = 0$. Logo, as técnicas a seguir podem ser usadas para calcular eficientemente integrais definidas envolvendo funções que são lisas em geral, mas que exibem regiões de variação abrupta.

Há duas abordagens principais para incorporar controle adaptativo do tamanho do passo em métodos de passo único. Na primeira, a estimativa do erro é feita como a diferença entre duas previsões usando um método RK de mesma ordem, mas com tamanho de passo diferente. Na segunda, a estimativa do erro de truncamento local é feita como a diferença entre duas previsões usando métodos RK de ordens diferentes.

25.5.1 Método RK Adaptativo ou de Divisão do Passo na Metade

A *divisão do passo na metade* (também chamado *RK adaptativo*) envolve fazer cada passo duas vezes, uma como um passo completo, e independentemente, como dois meio passos. A diferença nos dois resultados representa uma estimativa do erro de truncamento local. Se y_1 designar a previsão por um único passo e y_2 designar a previsão usando dois meio passos, o erro Δ pode ser representado por

$$\Delta = y_2 - y_1 \tag{25.43}$$

Além de fornecer um critério para o controle do tamanho do passo, a Equação (25.43) também pode ser usada para corrigir a previsão y_2. Para a versão RK de quarta ordem, a correção é

$$y_2 \leftarrow y_2 + \frac{\Delta}{15} \tag{25.44}$$

Essa estimativa é acurada até a quinta ordem.

EXEMPLO 25.12 Método RK de Quarta Ordem Adaptativo

Enunciado do Problema. Use o método RK de quarta ordem adaptativo para integrar $y' = 4e^{0,8x} - 0,5y$ de $x = 0$ a 2 usando $h = 2$ e uma condição inicial $y(0) = 2$. Esta é a mesma equação diferencial que foi resolvida anteriormente no Exemplo 25.5. Lembre-se de que a solução verdadeira é $y(2) = 14,84392$.

Solução. A previsão única com um passo h é calculada como

$$y(2) = 2 + \frac{1}{6}[3 + 2(6,40216 + 4,70108) + 14,11105]2 = 15,10584$$

As duas previsões de meio passo são

$$y(1) = 2 + \frac{1}{6}[3 + 2(4,21730 + 3,91297) + 5,945681]1 = 6,20104$$

e

$$y(2) = 6,20104 + \frac{1}{6}[5,80164 + 2(8,72954 + 7,99756) + 12,71283]1 = 14,86249$$

Portanto, o erro aproximado é

$$E_a = \frac{14,86249 - 15,10584}{15} = -0,01622$$

que se compara favoravelmente com o erro verdadeiro de

$$E_t = 14{,}84392 - 14{,}86249 = -0{,}01857$$

A estimativa de erro também pode ser usada para corrigir a previsão

$$y(2) = 14{,}86249 - 0{,}01622 = 14{,}84627$$

a qual tem um $E_t = -0{,}00235$.

25.5.2 Runge-Kutta Fehlberg

Além de dividir o passo na metade como estratégia para ajustar o tamanho do passo, uma abordagem alternativa para obter uma estimativa do erro envolve o cálculo de duas previsões RK de ordens diferentes. Os resultados podem então ser subtraídos para se obter uma estimativa do erro de truncamento local. Uma desvantagem dessa abordagem é que ela aumenta muito o gasto computacional. Por exemplo, previsões de quarta e quinta ordens equivalem a um cálculo de dez valores de função no total por passo. O método de *Runge-Kutta Fehlberg* ou *RK embutido* contorna esse problema de forma inteligente pelo uso de um método RK de quinta ordem que utiliza os cálculos de função do método RK de quarta ordem que o acompanha. Assim, a abordagem fornece uma estimativa de erro com base em apenas seis cálculos de função!

No caso presente, usamos a seguinte estimativa de quarta ordem

$$y_{i+1} = y_i + \left(\frac{37}{378}k_1 + \frac{250}{621}k_3 + \frac{125}{594}k_4 + \frac{512}{1.771}k_6\right)h \tag{25.45}$$

junto com a fórmula de quinta ordem:

$$y_{i+1} = y_i + \left(\frac{2.825}{27.648}k_1 + \frac{18.575}{48.384}k_3 + \frac{13.525}{55.296}k_4 + \frac{277}{14.336}k_5 + \frac{1}{4}k_6\right)h \tag{25.46}$$

em que

$$k_1 = f(x_i, y_i)$$

$$k_2 = f\left(x_i + \frac{1}{5}h, y_i + \frac{1}{5}k_1h\right)$$

$$k_3 = f\left(x_i + \frac{3}{10}h, y_i + \frac{3}{40}k_1h + \frac{9}{40}k_2h\right)$$

$$k_4 = f\left(x_i + \frac{3}{5}h, y_i + \frac{3}{10}k_1h - \frac{9}{10}k_2h + \frac{6}{5}k_3h\right)$$

$$k_5 = f\left(x_i + h, y_i - \frac{11}{54}k_1h + \frac{5}{2}k_2h - \frac{70}{27}k_3h + \frac{35}{27}k_4h\right)$$

$$k_6 = f\left(x_i + \frac{7}{8}h, y_i + \frac{1.631}{55.296}k_1h + \frac{175}{512}k_2h + \frac{575}{13.824}k_3h + \frac{44.275}{110.592}k_4h + \frac{253}{4.096}k_5h\right)$$

Assim, a EDO pode ser resolvida com a Equação (25.46) e a estimativa de erro obtida, como a diferença das estimativas de quinta e quarta ordens. Deve ser observado que os coeficientes particulares usados foram desenvolvidos por Cash e Karp (1990). Portanto, esse método é algumas vezes chamado de RK de *Cash-Karp*.

EXEMPLO 25.13 Método de Runge-Kutta Fehlberg

Enunciado do Problema. Use a versão Cash-Karp da abordagem de Runge-Kutta Fehlberg para fazer os mesmos cálculos que no Exemplo 25.12 de $x = 0$ a 2 usando $h = 2$.

Solução. O cálculo dos k's pode ser resumido na seguinte tabela:

	x	y	$f(x, y)$
k_1	0	2	3
k_2	0,4	3,2	3,908511
k_3	0,6	4,20883	4,359883
k_4	1,2	7,228398	6,832587
k_5	2	15,42765	12,09831
k_6	1,75	12,17686	10,13237

Eles podem ser usados para calcular a previsão de quarta ordem

$$y_1 = 2 + \left(\frac{37}{378}3 + \frac{250}{621}4{,}359883 + \frac{125}{594}6{,}832587 + \frac{512}{1.771}10{,}13237\right)2 = 14{,}83192$$

junto com a fórmula de quinta ordem:

$$y_1 = 2 + \left(\frac{2.825}{27.648}3 + \frac{18.575}{48.384}4{,}359883 + \frac{13.525}{55.296}6{,}832587 + \frac{277}{14.336}12{,}09831 + \frac{1}{4}10{,}13237\right)2 = 14{,}83677$$

A estimativa de erro é obtida subtraindo-se essas duas equações para obter

$$E_a = 14{,}83677 - 14{,}83192 = 0{,}004842$$

25.5.3 Controle do Tamanho do Passo

Agora que desenvolvemos maneiras de obter estimativas do erro de truncamento local, elas podem ser usadas para ajustar o tamanho do passo. Em geral, a estratégia é aumentar o tamanho do passo se o erro for muito pequeno e diminuir se o erro for muito grande. Press et al. (1992) sugeriram o seguinte critério para conseguir isso:

$$h_{\text{novo}} = h_{\text{atual}} \left| \frac{\Delta_{\text{novo}}}{\Delta_{\text{atual}}} \right|^{\alpha} \tag{25.47}$$

em que h_{atual} e h_{novo} são os tamanhos de passo atual e novo, respectivamente, Δ_{atual} é a acurácia atual calculada, Δ_{novo} é a acurácia desejada e α é uma potência constante que é igual a 0,2 quando o tamanho do passo é aumentado (isto é, quando $\Delta_{\text{atual}} \leq \Delta_{\text{novo}}$) e a 0,25 quando o tamanho do passo é diminuído ($\Delta_{\text{atual}} > \Delta_{\text{novo}}$).

O parâmetro chave na Equação (25.47) é obviamente Δ_{novo}, pois é o seu veículo para especificar a acurácia desejada. Uma maneira de fazer isso seria relacionar Δ_{novo} a um nível de erro relativo. Embora isso funcione bem quando ocorrem apenas valores positivos, pode causar problemas para soluções que passem pelo zero. Por exemplo, você poderia estar simulando uma função oscilante que passe repetidamente pelo zero, mas que seja limitada por valores absolutos máximos. Para tal caso, você poderia querer incluir esses valores máximos na acurácia desejada.

Uma forma mais geral de lidar com tais casos é determinar Δ_{novo} como

$$\Delta_{\text{novo}} = \varepsilon y_{\text{escala}}$$

em que ε é um nível de tolerância geral. Sua escolha de y_{escala} determinará, então, como é a escala do erro. Por exemplo, se $y_{\text{escala}} = y$, a precisão será formulada em termos de erros relativos fracionários. Se você estiver tratando um caso no qual queira erros constantes em relação a um limitante máximo prescrito, iguale y_{escala} a esse limitante. Um truque sugerido por Press et al. (1992) para obter erros relativos constantes, exceto muito próximo a cruzamentos com o zero, é

$$y_{\text{escala}} = |y| + \left| h\frac{dy}{dx} \right|$$

Essa é a versão que usaremos em nosso algoritmo.

25.5.4 Algoritmo Computacional

As Figuras 25.21 e 25.22 esboçam pseudocódigos para implementar a versão Cash-Karp do algoritmo de Runge-Kutta Fehlberg. Esse algoritmo usou como modelo, uma implementação mais detalhada por Press et al. (1992) para sistemas de EDOs.

A Figura 25.21 implementa um único passo da rotina de Cash-Karp [isto é, as equações (25.45) e (25.46)]. A Figura 25.22 delineia um programa principal geral junto com uma sub-rotina que de fato adapta o tamanho do passo.

FIGURA 25.21
Pseudocódigo para um único passo do método RK de Cash-Karp.

```
SUBROUTINE RKkc (y,dy,x,h,yout,yerr)
PARAMETER (a2=0.2,a3=0.3,a4=0.6,a5=1.,a6=0.875,
  b21=0.2,b31=3./40.,b32=9./40.,b41=0.3,b42=−0.9,
  b43=1.2,b51=−11./54.,b52=2.5,b53=−70./27.,
  b54=35./27.,b61=1631./55296.,b62=175./512.,
  b63=575./13824.,b64=44275./110592.,b65=253./4096.,
  c1=37./378.,c3=250./621.,c4=125./594.,
  c6=512./1771.,dc1=c1−2825./27648.,
  dc3=c3−18575./48384.,dc4=c4−13525./55296.,
  dc5=−277./14336.,dc6=c6−0.25)
ytemp=y+b21*h*dy
CALL Derivs (x+a2*h,ytemp,k2)
ytemp=y+h*(b31*dy+b32*k2)
CALL Derivs(x+a3*h,ytemp,k3)
ytemp=y+h*(b41*dy+b42*k2+b43*k3)
CALL Derivs(x+a4*h,ytemp,k4)
ytemp=y+h*(b51*dy+b52*k2+b53*k3+b54*k4)
CALL Derivs(x+a5*h,ytemp,k5)
ytemp=y+h*(b61*dy+b62*k2+b63*k3+b64*k4+b65*k5)
CALL Derivs(x+a6*h,ytemp,k6)
yout=y+h*(c1*dy+c3*k3+c4*k4+c6*k6)
yerr=h*(dc1*dy+dc3*k3+dc4*k4+dc5*k5+dc6*k6)
END RKkc
```

FIGURA 25.22
Pseudocódigo para um (a) programa principal e (b) rotina adaptativa do passo para resolver uma única EDO.

(*a*) Programa Principal

```
INPUT xi, xf, yi
maxstep=100
hi=.5; tiny = 1. × 10^−30
eps=0.00005
print *, xi,yi
x=xi
y=yi
h=hi
istep=0
DO
  IF (istep > maxstep AND x ≤ xf) EXIT
  istep=istep+1
  CALL Derivs(x,y,dy)
  yscal=ABS(y)+ABS(h*dy)+tiny
  IF (x+h>xf) THEN h=xf−x
  CALL Adapt (x,y,dy,h,yscal,eps,hnxt)
  PRINT x,y
  h=hnxt
END DO
END
```

(*b*) Rotina Adaptativa do Passo

```
SUB Adapt (x,y,dy,htry,yscal,eps,hnxt)
PARAMETER (safety=0.9,econ=1.89e−4)
h=htry
DO
  CALL RKkc(y,dy,x,h,ytemp,yerr)
  emax=abs(yerr/yscal/eps)
  IF emax ≤ 1 EXIT
  htemp=safety*h*emax^−0.25
  h=max(abs(htemp),0.25*abs(h))
  xnew=x+h
  IF xnew = x THEN pause
END DO
IF emax > econ THEN
  hnxt=safety*emax^−.2*h
ELSE
  hnxt=4.*h
END IF
x=x+h
y=ytemp
END Adapt
```

EXEMPLO 25.14 Aplicação Computacional de um Esquema RK de Quarta Ordem Adaptativo

Enunciado do Problema. O método RK adaptativo é adequado para a seguinte equação diferencial ordinária

$$\frac{dy}{dx} + 0{,}6y = 10e^{-(x-2)^2/[2(0{,}075)^2]} \tag{E25.14.1}$$

Observe que, para a condição inicial $y(0) = 0{,}5$, a solução da equação homogênea associada é

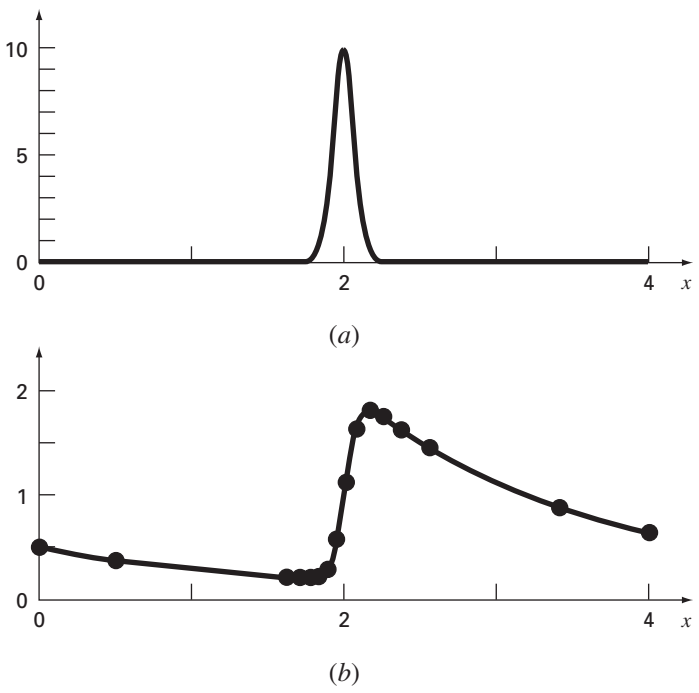
$$y = 0{,}5e^{-0{,}6x} \tag{E25.14.2}$$

que é uma curva lisa que gradualmente se aproxima de zero conforme x aumenta. Em contraste, a solução da equação não-homogênea passa por uma transição abrupta na vizinhança de $x = 2$ por causa da natureza do termo forçante (Figura 25.23*a*). Use um esquema RK de quarta ordem padrão para resolver a Equação (E25.14.1) de $x = 0$ a 4. A seguir, use o esquema adaptativo descrito nesta seção para fazer o mesmo cálculo.

Solução. Primeiro, é utilizado o esquema de quarta ordem clássico para calcular a curva sólida na Figura 25.23*b*. Para esse cálculo, foi usado um tamanho de passo de 0,1 de modo que foram feitas $4/(0{,}1) = 40$ aplicações da técnica. Então, os cálculos foram repetidos com um tamanho de passo de 0,05, para um total de 80 aplicações. A maior discrepância entre os dois resultados ocorre na região entre 1,8 e 2,0. O valor absoluto da discrepância é de cerca de 0,1% a 0,2%.

A seguir, o algoritmo nas Figuras 25.21 e 25.22 é desenvolvido em um programa computacional e usado para resolver o mesmo problema. Foram escolhidos um tamanho de passo inicial de 0,5 e um $\varepsilon = 0{,}00005$. Os resultados foram sobrepostos na Figura 25.23*b*. Observe quão grandes são os passos tomados na região de variação gradual. Então, na vizinhança de $x = 2$, os passos são diminuídos para acomodar a natureza abrupta da função forçante.

FIGURA 25.23
(*a*) Função forçante em forma de sino que induz uma variação abrupta na solução de uma EDO [Equação (E25.14.1)]. (*b*) A solução. Os pontos indicam a previsão de uma rotina adaptativa para o tamanho do passo.



A utilidade de um esquema de integração adaptativo obviamente depende da natureza da função que está sendo modelada. Ele é particularmente vantajoso para aquelas soluções com longas extensões lisas e regiões curtas de variação abrupta. Além disso, tem utilidade nas situações nas quais o tamanho do passo correto não é conhecido *a priori*. Nesses casos, uma rotina adaptativa irá "tatear" seu caminho pela solução enquanto mantém os resultados dentro da tolerância desejada. Logo, vai "andar na ponta dos pés" nas regiões de variação abrupta e caminhar rapidamente quando a variação se tornar mais gradual.

PROBLEMAS

25.1 Resolva o seguinte problema de valor inicial analiticamente no intervalo de $x = 0$ a 2:

$$\frac{dy}{dx} = yx^2 - 1{,}1y$$

em que $y(0) = 1$. Faça o gráfico da solução.

25.2 Use o método de Euler com $h = 0{,}5$ e 0,25 para resolver o Problema 25.1. Trace seus resultados no mesmo gráfico para comparar visualmente a acurácia para os dois tamanhos de passo.

25.3 Use o método de Heun com $h = 0{,}5$ para resolver o Problema 25.1. Faça iterações no corretor até $\varepsilon_s = 1\%$.

25.4 Use o método do ponto médio com $h = 0{,}5$ e 0,25 para resolver o Problema 25.1.

25.5 Use o método RK de quarta ordem clássico com $h = 0{,}5$ para resolver o Problema 25.1.

25.6 Repita os Problemas 25.1 a 25.5, mas para o seguinte problema de valor inicial no intervalo de $x = 0$ a 1:

$$\frac{dy}{dx} = (1 + 2x)\sqrt{y} \qquad y(0) = 1$$

25.7 Use os métodos de (**a**) Euler e (**b**) Heun (sem iteração) para resolver

$$\frac{d^2y}{dt^2} - 0{,}5t + y = 0$$

em que $y(0) = 2$ e $y'(0) = 0$. Resolva de $x = 0$ a 4 usando $h = 0{,}1$. Compare os métodos fazendo os gráficos da solução.

25.8 Resolva o seguinte problema com o método RK de quarta ordem:

$$\frac{d^2y}{dx^2} + 0{,}6\frac{dy}{dx} + 8y = 0$$

em que $y(0) = 4$ e $y'(0) = 0$. Resolva de $x = 0$ a 5 com $h = 0{,}5$. Represente os resultados graficamente.

25.9 Resolva de $t = 0$ a 3 com $h = 0{,}1$ usando o método de (**a**) Heun (sem iteração) e (**b**) RK de segunda ordem de Ralston:

$$\frac{dy}{dt} = y\ \operatorname{sen}^3(t) \qquad y(0) = 1$$

25.10 Resolva o seguinte problema numericamente de $t = 0$ a 3.

$$\frac{dy}{dt} = -y + t^2 \qquad y(0) = 1$$

Use o método RK de terceira ordem com tamanho de passo de 0,5.

25.11 Use o método (**a**) de Euler e (**b**) RK de quarta ordem para resolver

$$\frac{dy}{dx} = -2y + 4e^{-x}$$

$$\frac{dz}{dx} = -\frac{yz^2}{3}$$

no intervalo de $x = 0$ a 1 usando um tamanho de passo de 0,2, com $y(0) = 2$ e $z(0) = 4$.

25.12 Calcule o primeiro passo do Exemplo 25.14 usando o método RK de quarta ordem adaptativo com $h = 0{,}5$. Verifique se o ajuste no tamanho do passo está bom.

25.13 Se $\varepsilon = 0{,}001$, determine se é necessário ajuste no tamanho do passo no Exemplo 25.12.

25.14 Use a abordagem de RK-Fehlberg para fazer os mesmos cálculos que no Exemplo 25.12 de $x = 0$ a 1 com $h = 1$.

25.15 Escreva um programa computacional baseado na Figura 25.7. Entre outras coisas, faça comentários em todo o programa para identificar o que cada seção deve fazer.

25.16 Teste o programa que você desenvolveu no Problema 25.15 para duplicar os cálculos dos Exemplos 25.1 e 25.4.

25.17 Desenvolva um programa amigável ao usuário para o método de Heun com um corretor iterativo. Teste seu programa duplicando os resultados na Tabela 25.2.

25.18 Desenvolva um programa amigável ao usuário para o método RK de quarta ordem clássico. Teste seu programa duplicando o Exemplo 25.7.

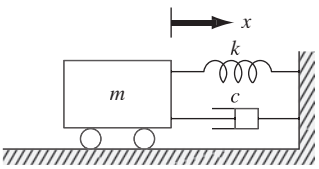
25.19 Desenvolva um programa amigável ao usuário para sistemas de equações usando o método RK de quarta ordem. Use esse programa para duplicar os cálculos no Exemplo 25.10.

25.20 O movimento de um sistema massa-mola amortecido (Figura P25.20) é descrito pela seguinte equação diferencial ordinária:

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = 0$$

em que x é o deslocamento a partir da posição de equilíbrio (m), t é o tempo (s), $m = 20$ kg é a massa e c é o coeficiente de amortecimento (N · s/m). O coeficiente de amortecimento c toma os três valores de 5 (subamortecido), 40 (criticamente amortecido) e 200 (superamortecido). A constante da mola é $k = 20$ N/m. A velocidade inicial é zero e o deslocamento inicial é $x = 1$ m. Resolva essa equação usando um método numérico no período de tempo $0 \le t \le 15$ s. Faça um gráfico do deslocamento em função do tempo para cada um dos três valores de coeficiente de amortecimento na mesma curva.

Figura P25.20



25.21 Se água for drenada de um tanque cilíndrico vertical abrindo-se uma válvula na base, ela escoará rapidamente quando o tanque estiver cheio e mais lentamente conforme ele continuar a ser drenado. Como pode ser mostrado, a taxa pela qual o nível de água abaixa é:

$$\frac{dy}{dt} = -k\sqrt{y}$$

em que k é uma constante dependente da forma do orifício, da área da seção transversal do tanque e do orifício de drenagem. A profundidade da água y é medida em metros e o tempo t, em minutos. Se $k = 0{,}06$, determine quanto tempo levará para que o tanque fique vazio se o nível de fluido for inicialmente 3 m. Resolva aplicando a equação de Euler e escrevendo um programa computacional ou usando o Excel. Use um passo de 0,5 minuto.

25.22 A seguir está uma equação diferencial de segunda ordem, com condição inicial:

$$\frac{d^2x}{dt^2} + (5x)\frac{dx}{dt} + (x+7)\ \text{sen}(\omega t) = 0$$

em que

$$\frac{dx}{dt}(0) = 1{,}5 \quad \text{e} \quad x(0) = 6$$

Observe que $\omega = 1$. Decomponha a equação em duas equações diferenciais de primeira ordem. Depois da decomposição, resolva o sistema de $t = 0$ a 15 e faça o gráfico dos resultados.

25.23 Supondo que o arrasto seja proporcional ao quadrado da velocidade, podemos modelar a velocidade de um objeto em queda livre como o pára-quedista com a seguinte equação diferencial:

$$\frac{dv}{dt} = g - \frac{c_d}{m}v^2$$

em que v é a velocidade (m/s), t é o tempo (s), g é a aceleração da gravidade (9,81 m/s^2), c_d é um coeficiente de arrasto de segunda ordem (kg/m) e m é a massa (kg). Resolva para determinar a velocidade e a distância percorrida na queda por um objeto de 90 kg com um coeficiente de arrasto de 0,225 kg/m. Se a altura inicial for 1 km, determine quando ele atinge o chão. Obtenha sua solução com (**a**) o método de Euler e (**b**) o método RK de quarta ordem.

25.24 Um tanque esférico tem um orifício circular no fundo pelo qual o líquido escoa para fora (Figura P25.24). A vazão pelo buraco pode ser estimada por

$$Q_{\text{out}} = CA\sqrt{2gh}$$

em que Q_{out} é o escoamento para fora (m^3/s), C é um coeficiente obtido empiricamente, A é a área do orifício (m^2), g é a constante gravitacional (= 9,81 m/s^2) e h é a profundidade do líquido no tanque. Use um dos métodos numéricos descritos neste capítulo para determinar quanto tempo levará para que a água escoe para fora de um tanque de 3 m de diâmetro com uma altura inicial de 2,75 m. Observe que o orifício tem um diâmetro de 3 cm e $C = 0{,}55$.

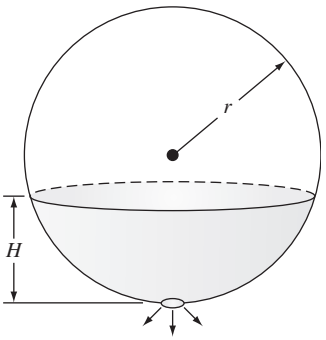


Figura P25.24
Um tanque esférico.

25.25 O modelo logístico é usado para simular a população por

$$\frac{dp}{dt} = k_{gm}\,(1 - p/p_{\max})p$$

em que p é a população, k_{gm} é a taxa de crescimento máxima sob condições ilimitadas e $p_{\text{máx}}$ é a capacidade de suporte. Simule a população mundial de 1950 a 2000 usando um dos métodos numéricos descritos neste capítulo. Use as seguintes condições iniciais e valores de parâmetros para sua simulação: p_0 (em 1950) = 2,555 milhões de pessoas, $k_{gm} = 0{,}026$/ano, e $p_{\text{máx}} = 12{,}000$ milhões de pessoas. Faça a função gerar saídas correspondentes às datas para os seguintes dados populacionais medidos. Desenvolva um gráfico de sua simulação junto com os dados.

t	1950	1960	1970	1980	1990	2000
p	2,555	3,040	3,708	4,454	5,276	6,079

25.26 Suponha que um projétil seja lançado da superfície da Terra. Suponha que a única força agindo no objeto seja a força gravitacional para baixo. Nessas condições, o balanço de forças pode ser usado para deduzir que

$$\frac{dv}{dt} = -g(0)\frac{R^2}{(R+x)^2}$$

em que v é a velocidade para cima (m/s), t é o tempo (s), x é a altitude (m) medida para cima a partir da superfície da Terra, $g(0)$ é a aceleração gravitacional na superfície da Terra ($\cong 9{,}81$ m/s^2) e R é o raio da Terra ($\cong 6{,}37 \times 10^6$ m). Percebendo que $dx/dt = v$, use o método de Euler para determinar a altura máxima que seria atingida se $v(t = 0) = 1.400$ m/s.

25.27 A seguinte função exibe tanto regiões achatadas quanto íngremes em uma região relativamente curta de x:

$$f(x) = \frac{1}{(x-0{,}3)^2 + 0{,}01} + \frac{1}{(x-0{,}9)^2 + 0{,}04} - 6$$

Determine o valor da integral definida dessa função entre $x = 0$ e 1 usando um método RK adaptativo.

CAPÍTULO 26

Rigidez e Métodos de Passo Múltiplo

Este capítulo abrange duas áreas. Primeiro, descrevemos *EDOs rígidas*. Elas são tanto EDOs individuais quanto sistemas de EDOs que têm ambas as componentes rápida e lenta em sua solução. Introduzimos a idéia de uma técnica de *solução implícita* como um recurso comumente usado para tratar esse problema. A seguir, discutiremos *métodos de passo múltiplo*. Esses algoritmos retêm informações de passos anteriores para capturar mais eficientemente a trajetória da solução. Também fornecem estimativas do erro de truncamento que podem ser usadas para implementar um controle adaptativo do tamanho do passo.

26.1 RIGIDEZ

A rigidez é um problema especial que pode aparecer na solução de equações diferenciais ordinárias. Um *sistema rígido* é aquele que envolve componentes variando rapidamente junto com componentes variando lentamente. Em muitos casos, as componentes que variam rapidamente são transientes efêmeros que desaparecem rapidamente, depois do que a solução passa a ser dominada pelas componentes que variam lentamente. Embora os fenômenos transientes existam apenas para uma pequena parte do intervalo de integração, podem determinar o passo no tempo para toda a solução.

Tanto as EDOs individuais quanto os sistemas de EDO podem ser rígidos. Um exemplo de uma única EDO rígida é

$$\frac{dy}{dt} = -1.000y + 3.000 - 2.000e^{-t} \tag{26.1}$$

Se $y(0) = 0$, a solução analítica pode ser escrita como

$$y = 3 - 0{,}998e^{-1000t} - 2{,}002e^{-t} \tag{26.2}$$

Como na Figura 26.1, a solução é inicialmente dominada pelo termo exponencial rápido (e^{-1000t}). Depois de um curto período ($t < 0{,}005$), esse transiente desaparece e a solução passa a ser dominada pela exponencial lenta (e^{-t}).

Podemos ganhar uma percepção do tamanho do passo necessário para a estabilidade de tal solução examinando a parte homogênea da Equação (26.1),

$$\frac{dy}{dt} = -ay \tag{26.3}$$

Se $y(0) = y_0$, o cálculo pode ser usado para determinar a solução como

$$y = y_0 e^{-at}$$

Logo, a solução começa em y_0 e se aproxima assintoticamente de zero.

O método de Euler pode ser usado para resolver o mesmo problema numericamente:

$$y_{i+1} = y_i + \frac{dy_i}{dt}h$$

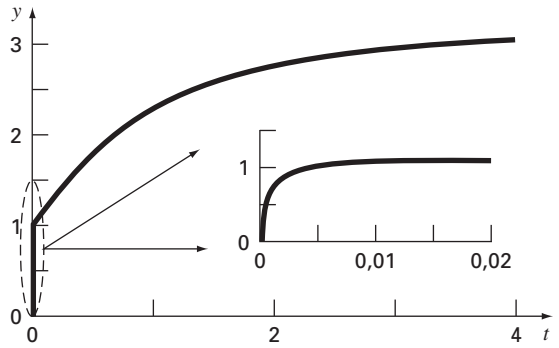


FIGURA 26.1
Gráfico de uma solução rígida de uma única EDO. Embora a solução pareça começar em 1, na realidade existe um transiente rápido de $y = 0$ a 1, que ocorre em menos do que 0,005 unidade de tempo. Esse transiente é perceptível apenas quando a resposta é vista na inserção em uma escala de tempo mais fina.

Substituindo a Equação (26.3) obtemos

$$y_{i+1} = y_i - ay_i h$$

ou

$$y_{i+1} = y_i(1 - ah) \tag{26.4}$$

A estabilidade dessa fórmula depende claramente do tamanho do passo h. Isto é, $|1 - ah|$ deve ser menor do que 1. Logo, se $h > 2/a$, $|y_i| \to \infty$ quando $i \to \infty$.

Para a parte transiente rápida da Equação (26.2), esse critério pode ser usado para mostrar que o tamanho do passo para manter a estabilidade deve ser $< 2/1.000 = 0{,}002$. Além disso, deve ser observado que, embora esse critério mantenha a estabilidade (ou seja, uma solução limitada), seria necessário um tamanho de passo ainda menor para obter uma solução acurada. Logo, embora o transiente ocorra apenas para uma pequena fração do intervalo de integração, ele controla o tamanho máximo permissível do passo.

Superficialmente, você poderia supor que as rotinas adaptativas do tamanho do passo descritas no final do último capítulo poderiam oferecer uma solução para esse dilema. Poderia pensar que elas usariam pequenos passos durante os transientes rápidos e passos grandes caso contrário. Entretanto, esse não é o caso, porque as exigências de estabilidade ainda fariam que fosse necessário usar tamanhos de passo muito pequenos ao longo de toda a solução.

Em vez de usar abordagens explícitas, os métodos implícitos oferecem um recurso alternativo. Tais representações são chamadas *implícitas* porque a incógnita aparece em ambos os lados da equação. Uma forma implícita do método de Euler pode ser desenvolvida pelo cálculo da derivada em um instante futuro,

$$y_{i+1} = y_i + \frac{dy_{i+1}}{dt}h$$

Isso é chamado de *método de Euler regressivo ou implícito*. A substituição da Equação (26.3) fornece

$$y_{i+1} = y_i - ay_{i+1}h$$

que pode ser resolvida por

$$y_{i+1} = \frac{y_i}{1 + ah} \tag{26.5}$$

Nesse caso, independentemente do tamanho do passo, $|y_i| \to 0$ quando $i \to \infty$. Assim, a abordagem é chamada *incondicionalmente estável*.

EXEMPLO 26.1

Euler Explícito e Implícito

Enunciado do Problema. Use os métodos de Euler implícito e explícito para resolver

$$\frac{dy}{dt} = -1.000y + 3.000 - 2.000e^{-t}$$

em que $y(0) = 0$. (**a**) Use Euler explícito com tamanhos de passo de 0,0005 e 0,0015 para determinar y entre $t = 0$ e 0,006. (**b**) Use Euler implícito com tamanho de passo de 0,05 para determinar y entre 0 e 0,4.

Solução.

(a) Para esse problema, o método de Euler explícito é

$$y_{i+1} = y_i + (-1000y_i + 3000 - 2000e^{-t_i})h$$

O resultado para $h = 0{,}0005$ está mostrado na Figura 26.2*a* junto com a solução analítica. Embora ela apresente algum erro de truncamento, o resultado captura a forma geral da solução analítica. Em contraste, quando o tamanho do passo é aumentado para um valor um pouco abaixo do limite de estabilidade ($h = 0{,}0015$), a solução apresenta oscilações. Usar $h > 0{,}002$ resultaria em uma solução totalmente instável, isto é, que iria para infinito conforme a solução prosseguisse.

(b) O método de Euler implícito é

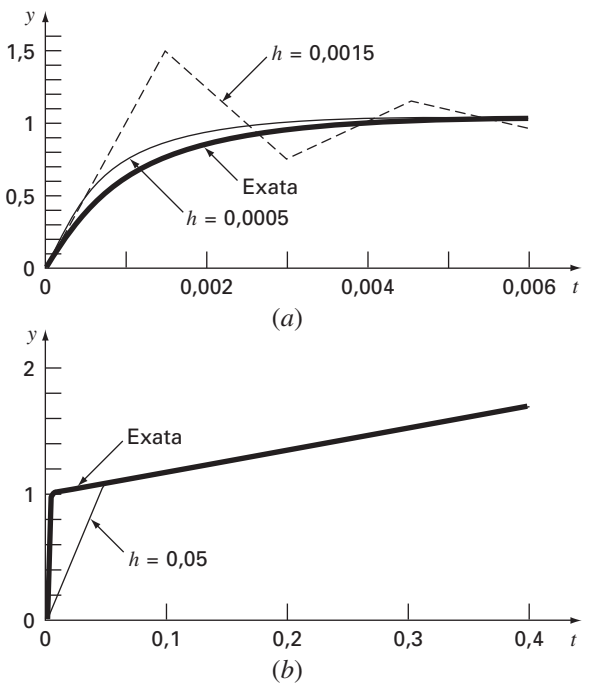
$$y_{i+1} = y_i + (-1000y_{i+1} + 3000 - 2000e^{-t_{i+1}})h$$

Agora, como a EDO é linear, podemos reescrever essa equação de modo que y_{i+1} fique isolado no lado esquerdo,

$$y_{i+1} = \frac{y_i + 3000h - 2000he^{-t_{i+1}}}{1 + 1000h}$$

O resultado para $h = 0{,}05$ está mostrado na Figura 26.2*b* junto com a solução analítica. Observe que, embora tenhamos usado um tamanho de passo muito maior do que aquele que induziu estabilidade no Euler explícito, a solução numérica segue de perto o resultado analítico.

FIGURA 26.2
Solução de uma EDO rígida com os métodos de Euler (*a*) explícito e (*b*) implícito.



Os sistemas de EDOs também podem ser rígidos. Um exemplo é

$$\frac{dy_1}{dt} = -5y_1 + 3y_2 \tag{26.6a}$$

$$\frac{dy_2}{dt} = 100y_1 - 301y_2 \tag{26.6b}$$

Para condições iniciais $y_1(0) = 52{,}29$ e $y_2(0) = 83{,}82$, a solução exata é

$$y_1 = 52{,}96e^{-3{,}9899t} - 0{,}67e^{-302{,}0101t} \tag{26.7a}$$

$$y_2 = 17{,}83e^{-3{,}9899t} + 65{,}99e^{-302{,}0101t} \tag{26.7b}$$

Observe que os expoentes são negativos e diferem em cerca de duas ordens de grandeza. Como no caso de uma única equação, são os expoentes grandes que respondem rapidamente e são o cerne da rigidez do problema.

Um método de Euler implícito para sistemas pode ser formulado para o caso presente como

$$y_{1,i+1} = y_{1,i} + (-5y_{1,i+1} + 3y_{2,i+1})h \tag{26.8a}$$

$$y_{2,i+1} = y_{2,i} + (100y_{1,i+1} - 301y_{2,i+1})h \tag{26.8b}$$

Agrupando os termos, obtemos

$$(1 + 5h)y_{1,i+1} - 3hy_{2,i+1} = y_{1,i} \tag{26.9a}$$

$$-100hy_{1,i+1} + (1 + 301h)y_{2,i+1} = y_{2,i} \tag{26.9b}$$

Logo, podemos ver que o problema consiste em resolver um conjunto de equações simultâneas para cada passo no tempo.

Para EDOs não-lineares, a solução se torna ainda mais difícil, já que envolve a resolução de um sistema de equações simultâneas não-lineares (lembre-se da Seção 6.5). Assim, embora se ganhe estabilidade com as abordagens implícitas, é pago um preço na forma de um aumento da complexidade da solução.

O método de Euler implícito é incondicionalmente estável e acurado apenas em primeira ordem. Também é possível desenvolver de forma parecida um esquema de integração usando a regra do trapézio implícita, acurada até segunda ordem, para sistemas rígidos. Em geral é desejável obter métodos de ordem superior. As fórmulas de Adams-Moulton descritas depois neste capítulo também podem ser usadas para obter métodos implícitos de ordem superior. Entretanto, os limites de estabilidade de tais abordagens são muito severos quando aplicados a sistemas rígidos. Gear (1971) desenvolveu uma série especial de esquemas implícitos que têm limites de estabilidade muito maiores baseados em fórmulas de diferença regressiva. Esforços consideráveis têm sido feitos para desenvolver softwares que implementem eficientemente os métodos de Gear. Como resultado, esse provavelmente é o método mais amplamente usado para resolver sistemas rígidos. Além disso, Rosenbrock e outros (ver Press et al., 1992) propuseram algoritmos de Runge-Kutta implícitos em que k termos apareciam implicitamente. Esses métodos têm boas características de estabilidade e são bastante adequados para a solução de sistemas de equações diferenciais rígidos.

26.2 MÉTODOS DE PASSO MÚLTIPLO

Os métodos de passo único descritos nas seções anteriores utilizam informações em um único ponto x_i para prever um valor da variável dependente y_{i+1} em um ponto futuro x_{i+1} (Figura 26.3*a*). Abordagens alternativas chamadas *métodos de passo múltiplo* (Figura 26.3*b*) são baseadas na percepção de que, uma vez que o cálculo tenha começado, informação valiosa de pontos anteriores está a nossa disposição. A curvatura das linhas ligando esses valores anteriores fornece informação com relação à trajetória da solução. Os métodos de passo múltiplo explorados neste capítulo usam essa informação para resolver EDOs. Antes de descrever as versões de ordem superior, apresentaremos um método simples de segunda ordem que serve para ilustrar as características gerais das abordagens de passo múltiplo.

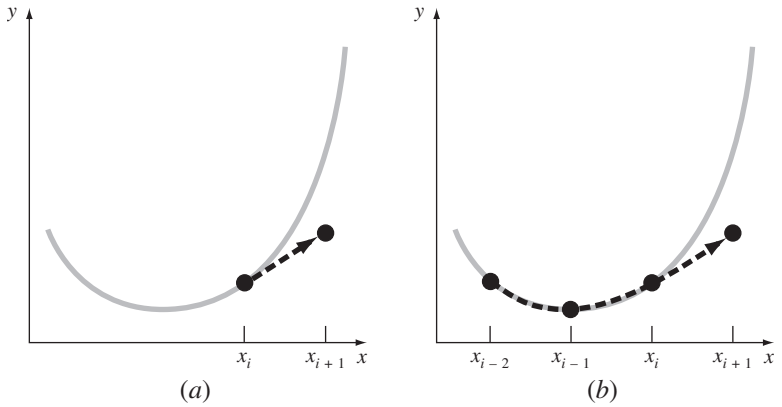


FIGURA 26.3
Descrição gráfica da diferença fundamental entre os métodos (*a*) de passo único e (*b*) de passo múltiplo para resolver EDOs.

26.2.1 O Método de Heun de Início Não-automático

Lembre-se de que a abordagem de Heun usa o *método de Euler* como um *preditor* [Equação (25.15)]:

$$y_{i+1}^0 = y_i + f(x_i, y_i)h \tag{26.10}$$

e a *regra do trapézio* como um *corretor* [Equação (25.16)]:

$$y_{i+1} = y_i + \frac{f(x_i, y_i) + f\left(x_{i+1}, y_{i+1}^0\right)}{2}h \tag{26.11}$$

Logo, o preditor e o corretor têm erros de truncamento locais de $O(h^2)$ e $O(h^3)$, respectivamente. Isso sugere que o preditor é o elo fraco no método, já que tem o erro maior. Essa fraqueza é significativa, pois a eficiência do passo iterativo de correção depende da acurácia da previsão inicial. Conseqüentemente, uma forma de melhorar o método de Heun é desenvolver um preditor que tenha um erro local de $O(h^3)$. Pode-se conseguir isso usando o método de Euler e a inclinação em y_i e informação extra de um ponto anterior y_{i-1}, como em

$$y_{i+1}^0 = y_{i-1} + f(x_i, y_i)2h \tag{26.12}$$

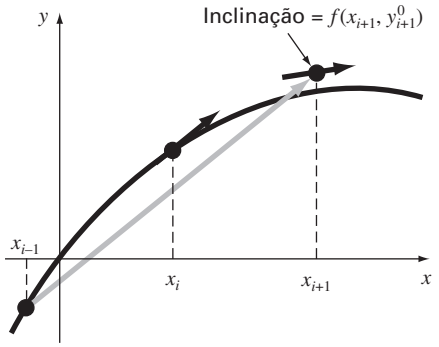
Observe que a Equação (26.12) alcança $O(h^3)$ usando um tamanho de passo maior, $2h$. Além disso, observe que a Equação (26.12) não é de início automático, pois envolve um valor anterior da variável dependente y_{i-1}. Tal valor não estaria disponível em um problema de valor inicial típico. Por causa desse fato, as Equações (26.11) e (26.12) são chamadas de *método de Heun de início não-automático*.

Como descrito na Figura 26.4, a estimativa da derivada na Equação (26.12) está agora localizada no ponto médio em vez de no início do intervalo no qual a previsão é feita. Como será mostrado a seguir, a centralização melhora o erro do preditor para $O(h^3)$. Entretanto, antes de prosseguir para uma dedução formal do Heun de início não-automático, vamos resumir o método e expressá-lo usando uma nomenclatura ligeiramente modificada:

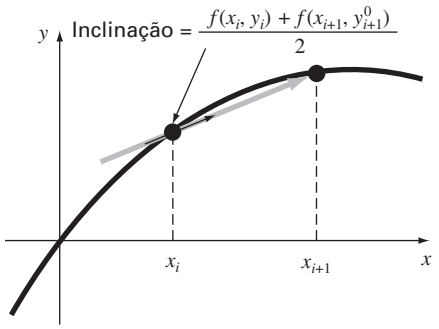
Preditor:
$$y_{i+1}^0 = y_{i-1}^m + f\left(x_i, y_i^m\right)2h \tag{25.13}$$

Corretor:
$$y_{i+1}^j = y_i^m + \frac{f\left(x_i, y_i^m\right) + f\left(x_{i+1}, y_{i+1}^{j-1}\right)}{2}h \tag{25.14}$$

$$(\text{para } j = 1, 2, \ldots, m)$$



(*a*)



(*b*)

FIGURA 26.4
Uma descrição gráfica do método de Heun de início não-automático. (*a*) O método do ponto médio, que é usado como um preditor. (*b*) A regra do trapézio que é usada como um corretor.

em que os sobrescritos foram adicionados para denotar que o corretor é aplicado iterativamente de $j = 1$ a m para obter soluções refinadas. Observe que y_i^m e y_{i-1}^m são o resultado final das iterações do corretor nos passos anteriores no tempo. Param-se as iterações em qualquer passo no tempo com base no seguinte critério de parada

$$|\varepsilon_a| = \left| \frac{y_{i+1}^j - y_{i+1}^{j-1}}{y_{i+1}^j} \right| 100\% \tag{26.15}$$

Quando ε_a for menor do que uma tolerância de erro pré-especificada, param-se as iterações. Nesse ponto, $j = m$. O uso das Equações (26.13) a (26.15) para resolver uma EDO está ilustrado no seguinte exemplo.

EXEMPLO 26.2 Método de Heun com Início Não-automático

Enunciado do Problema. Use o método de Heun de início não-automático para fazer os mesmos cálculos feitos anteriormente no Exemplo 25.5 com o método de Heun. Isto é, integre $y' = 4e^{0,8x} - 0,5y$ de $x = 0$ a $x = 4$ usando um tamanho de passo de 1,0. Como no Exemplo 25.5, a condição inicial em $x = 0$ é $y = 2$. Entretanto, como agora estamos lidando com um método de passo múltiplo, precisamos da informação adicional de que y é igual a $-0,3929953$ em $x = -1$.

Solução. O preditor [Equação (26.13)] é usado para extrapolar linearmente de $x = -1$ a $x = 1$.

$$y_1^0 = -0,3929953 + \left[4e^{0,8(0)} - 0,5(2)\right] 2 = 5,607005$$

O corretor [Equação (26.14)] é então usado para calcular o valor:

$$y_1^1 = 2 + \frac{4e^{0,8(0)} - 0,5(2) + 4e^{0,8(1)} - 0,5(5,607005)}{2}1 = 6,549331$$

o qual representa um erro relativo porcentual de $-5,73\%$ (valor verdadeiro = 6,194631). Esse erro é um pouco menor que o valor de 8,18% que ocorreu no método de Heun de início automático.

Agora, a Equação (26.14) pode ser aplicada iterativamente para melhorar a solução:

$$y_1^2 = 2 + \frac{3 + 4e^{0,8(1)} - 0,5(6,549331)}{2}1 = 6,313749$$

o que representa um ε_t de $-1,92\%$. Uma estimativa aproximada do erro também pode ser determinada usando a Equação (26.15):

$$|\varepsilon_a| = \left|\frac{6,313749 - 6,549331}{6,313749}\right| 100\% = 3,7\%$$

A Equação (26.14) pode ser aplicada iterativamente até que ε_a cai abaixo de um valor pré-especificado ε_s. Como no caso do método de Heun (lembre-se do Exemplo 25.5), as iterações convergem para um valor de 6,360865 ($\varepsilon_t = -2,68\%$). No entanto, como o valor inicial do preditor é mais acurado, o método de passo múltiplo converge a uma taxa um pouco mais rápida.

Para o segundo passo, o preditor é

$$y_2^0 = 2 + \left[4e^{0,8(1)} - 0,5(6,360865)\right]2 = 13,44346 \qquad \varepsilon_t = 9,43\%$$

o qual é superior à previsão de 12,08260 ($\varepsilon_t = 18\%$) que foi calculada com o método de Heun original. O primeiro corretor fornece 15,76693 ($\varepsilon_t = 6,8\%$) e as iterações subseqüentes convergem para o mesmo resultado obtido com o método de Heun de início automático: 15,30224 ($\varepsilon_t = -3,1\%$). Como no passo anterior, a taxa de convergência do corretor é um pouco melhor por causa da previsão inicial melhor.

Dedução e Análise de Erro das Fórmulas Preditor-Corretor. Acabamos de usar conceitos gráficos para deduzir o método de Heun de início não-automático. Agora mostraremos como as mesmas equações podem ser deduzidas matematicamente. Essa dedução pode ser particularmente interessante porque liga idéias de ajustes de curvas, integração numérica e EDOs. O exercício também é útil porque fornece um procedimento simples para deduzir métodos de passo múltiplo de ordem superior e estimativas para seus erros.

A dedução é baseada na solução da EDO geral

$$\frac{dy}{dx} = f(x, y)$$

Essa equação pode ser resolvida multiplicando-se ambos os lados por dx e integrando entre os extremos em i e $i + 1$:

$$\int_{y_i}^{y_{i+1}} dy = \int_{x_i}^{x_{i+1}} f(x, y)\, dx$$

O lado esquerdo pode ser integrado com o uso do teorema fundamental [lembre-se da Equação (25.21)]:

$$y_{i+1} = y_i + \int_{x_i}^{x_{i+1}} f(x, y)\, dx \tag{26.16}$$

A Equação (26.16) representa uma solução da EDO se a integral puder ser calculada, ou seja, ela fornece um meio de calcular um novo valor da variável dependente y_{i+1} com base no valor anterior y_i e na equação diferencial.

As fórmulas de integração numéricas como as que desenvolvemos no Capítulo 21 fornecem uma maneira de fazer esse cálculo. Por exemplo, a regra do trapézio [Equação (21.13)] pode ser usada para calcular a integral, como em

$$\int_{x_i}^{x_{i+1}} f(x, y)\, dx = \frac{f(x_i, y_i) + f(x_{i+1}, y_{i+1})}{2} h \tag{26.17}$$

em que $h = x_{i+1} - x_i$ é o tamanho do passo. Substituindo a Equação (26.17) na Equação (26.16), obtemos

$$y_{i+1} = y_i + \frac{f(x_i, y_i) + f(x_{i+1}, y_{i+1})}{2} h$$

que é a equação do corretor para o método de Heun. Como essa equação é baseada na regra do trapézio, o erro de truncamento pode ser obtido diretamente da Tabela 21.2,

$$E_c = -\frac{1}{12} h^3 y^{(3)}(\xi_c) = -\frac{1}{12} h^3 f''(\xi_c) \tag{26.18}$$

em que o subscrito c indica que esse é o erro do corretor.

Uma abordagem parecida pode ser usada para deduzir o preditor. Nesse caso, os limites de integração são de $i - 1$ a $i + 1$:

$$\int_{y_{i-1}}^{y_{i+1}} dy = \int_{x_{i-1}}^{x_{i+1}} f(x, y)\, dx$$

o que pode ser integrado e reorganizado para fornecer

$$y_{i+1} = y_{i-1} + \int_{x_{i-1}}^{x_{i+1}} f(x, y)\, dx \tag{26.19}$$

Agora, em vez de usar uma fórmula fechada da Tabela 21.2, a primeira fórmula de integração aberta de Newton-Cotes (veja a Tabela 21.4) pode ser utilizada para calcular a integral, como em

$$\int_{x_{i-1}}^{x_{i+1}} f(x, y)\, dx = 2h f(x_i, y_i) \tag{26.20}$$

o que é chamado *método do ponto médio*. A substituição da Equação (26.20) na Equação (26.19) fornece

$$y_{i+1} = y_{i-1} + 2h f(x_i, y_i)$$

que é o preditor para o método de Heun de início não-automático. Como no caso do corretor, o erro de truncamento local pode ser tomado diretamente da Tabela 21.4:

$$E_p = \frac{1}{3} h^3 y^{(3)}(\xi_p) = \frac{1}{3} h^3 f''(\xi_p) \tag{26.21}$$

em que o subscrito p indica que esse é o erro do preditor.

Assim, o preditor e o corretor para o método de Heun de início não-automático têm erros de mesma ordem. Além de melhorar a acurácia do preditor, esse fato tem as vantagens adicionais relacionadas com a análise de erro, como elaborado na próxima seção.

Estimativas de Erro. Se o preditor e o corretor de um método de passo múltiplo forem ambos da mesma ordem, pode ser obtida uma estimativa para o erro de truncamento local durante o transcorrer de um cálculo. Essa é uma tremenda vantagem, pois ela estabelece um critério para ajustes do tamanho do passo.

A estimativa para o erro de truncamento local é dada pela Equação (26.21). Essa estimativa de erro pode ser combinada com a estimativa de y_{i+1} do passo do preditor para fornecer [lembre-se da nossa definição básica na Equação (3.1)]

$$\text{Valor verdadeiro} = y_{i+1}^0 + \frac{1}{3} h^3 y^{(3)}(\xi_p) \tag{26.22}$$

Usando uma abordagem parecida, a estimativa de erro para o corretor [Equação (26.18)] pode ser combinada com o resultado do corretor y_{i+1} para fornecer

$$\text{Valor verdadeiro} = y_{i+1}^m - \frac{1}{12}h^3 y^{(3)}(\xi_c) \tag{26.23}$$

A Equação (26.22) pode ser subtraída da Equação (26.23) para fornecer

$$0 = y_{i+1}^m - y_{i+1}^0 - \frac{5}{12}h^3 y^{(3)}(\xi) \tag{26.24}$$

onde ξ agora está entre x_{i-1} e x_{i+1}. Agora, dividindo a Equação (26.24) por 5 e reorganizando o resultado, obtemos

$$\frac{y_{i+1}^0 - y_{i+1}^m}{5} = -\frac{1}{12}h^3 y^{(3)}(\xi) \tag{26.25}$$

Observe que os lados direitos das Equações (26.18) e (26.25) são idênticos, com exceção do argumento na terceira derivada. Se a terceira derivada não variar muito no intervalo em questão, podemos supor que os lados direitos são iguais e, portanto, os lados esquerdos também deveriam ser iguais, como em

$$E_c = -\frac{y_{i+1}^0 - y_{i+1}^m}{5} \tag{26.26}$$

Assim, chegamos a uma relação que pode ser usada para obter uma estimativa do erro de truncamento em cada passo com base em duas quantidades — o preditor (y_{i+1}^0) e o corretor (y_{i+1}^m) — que são subprodutos rotineiros dos cálculos.

EXEMPLO 26.3 Estimativa do Erro de Truncamento em cada Passo

Enunciado do Problema. Use a Equação (26.26) para obter uma estimativa do erro de truncamento em cada passo do Exemplo 26.2. Observe que os valores verdadeiros em $x = 1$ e 2 são 6,194631 e 14,84392, respectivamente.

Solução. Em $x_{i+1} = 1$, o preditor dá 5,607005 e o corretor fornece 6,360865. Esses valores podem ser substituídos na Equação (26.26) para dar

$$E_c = -\frac{6{,}360865 - 5{,}607005}{5} = -0{,}1507722$$

o qual se compara bem com o erro exato,

$$E_t = 6{,}194631 - 6{,}360865 = -0{,}1662341$$

Em $x_{i+1} = 2$, o preditor dá 13,44346 e o corretor fornece 15,30224, que podem ser usados para calcular

$$E_c = -\frac{15{,}30224 - 13{,}44346}{5} = -0{,}3717550$$

que também se compara favoravelmente com o erro exato, $E_t = 14{,}84392 - 15{,}30224 = -0{,}4583148$.

A facilidade com a qual a estimativa de erro pode ser obtida usando-se a Equação (26.26) fornece uma base racional para ajustes no tamanho do passo durante o decorrer de um cálculo. Por exemplo, se a Equação (26.26) indicar que o erro é maior do que o nível aceitável, o tamanho do passo poderia ser diminuído.

Modificadores. Antes de discutir os algoritmos computacionais, devemos observar duas outras formas pelas quais o método de Heun de início não-automático pode ser tornado mais acurado e eficiente. Primeiro, você deveria perceber que, além de fornecer um critério para o ajuste do tamanho do passo, a Equação (26.26) representa uma estimativa numérica da discrepância entre o valor corrigido final em cada passo y_{i+1} e o valor verdadeiro. Assim, ela pode ser adicionada diretamente a y_{i+1} para refinar a estimativa ainda mais:

$$y_{i+1}^m \leftarrow y_{i+1}^m - \frac{y_{i+1}^m - y_{i+1}^0}{5} \tag{26.27}$$

A Equação (26.27) é chamada de *modificador do corretor*. (O símbolo ← é lido "é substituído por".) O lado esquerdo é o valor modificado de y_{i+1}^m.

Uma segunda melhoria, que se relaciona mais com a eficiência do programa, é um *modificador do preditor*, cujo objetivo é ajustar o resultado do preditor de modo que ele esteja mais próximo do valor final para o qual o corretor converge. Isso é vantajoso porque, como observado no início da seção, o número de iterações do corretor depende muito da acurácia da previsão inicial. Conseqüentemente, se o preditor for modificado adequadamente, poderemos reduzir o número de iterações necessárias para convergir para o valor final do corretor.

Tal modificador pode ser deduzido simplesmente supondo que a terceira derivada é relativamente constante de passo a passo. Portanto, usando o resultado do passo anterior em i, a Equação (26.25) pode ser escrita como

$$h^3 y^{(3)}(\xi) = -\frac{12}{5}\left(y_i^0 - y_i^m\right) \tag{26.28}$$

a qual, supondo que $y^{(3)}(\xi) \cong y^{(3)}(\xi_p)$, pode ser substituída na Equação (26.21) para fornecer

$$E_p = \frac{4}{5}\left(y_i^m - y_i^0\right) \tag{26.29}$$

a qual pode então ser usada para modificar o resultado do preditor:

$$y_{i+1}^0 \leftarrow y_{i+1}^0 + \frac{4}{5}\left(y_i^m - y_i^0\right) \tag{26.30}$$

EXEMPLO 26.4 Efeito dos Modificadores nos Resultados do Preditor-Corretor

Enunciado do Problema. Recalcule o Exemplo 26.3 usando ambos os modificadores.

Solução. Como no Exemplo 26.3, o resultado inicial do preditor é 5,607005. Como o modificador do preditor [Equação (26.30)] usa valores de uma iteração anterior, ele não pode ser usado para melhorar o resultado inicial. Entretanto, a Equação (26.27) pode ser usada para modificar o valor corrigido de 6,360865 ($\varepsilon_t = -2{,}684\%$), como em

$$y_1^m = 6{,}360865 - \frac{6{,}360865 - 5{,}607005}{5} = 6{,}210093$$

o que representa um $\varepsilon_t = -0{,}25\%$. Assim, o erro é reduzido em mais de uma ordem de grandeza.

Para a próxima iteração, o preditor [Equação (26.13)] é usado para calcular

$$y_2^0 = 2 + \left[4e^{0{,}8(0)} - 0{,}5(6{,}210093)\right]2 = 13{,}59423 \qquad \varepsilon_t = 8{,}42\%$$

o que é cerca da metade do erro do preditor para a segunda iteração do Exemplo 26.3, que era de $\varepsilon_t = 18{,}6\%$. Essa melhora ocorre porque estamos usando uma estimativa superior de y (6,210093 em vez de 6,360865) no preditor. Em outras palavras, os erros propagado e global são reduzidos pela inclusão do modificador do corretor.

Agora, como temos informação da iteração anterior, a Equação (26.30) pode ser usada para modificar o preditor, como em

$$y_2^0 = 13{,}59423 + \frac{4}{5}(6{,}360865 - 5{,}607005) = 14{,}19732 \qquad \varepsilon_t = -4{,}36\%$$

o que, novamente, reduz o erro à metade.

Essa modificação não tem efeito na saída final do passo subseqüente do corretor. Independentemente de serem usados preditores modificados ou não, o corretor, no final das contas, convergirá para a mesma resposta. Mas, como a taxa ou eficiência da convergência depende da acurácia da previsão inicial, a modificação pode reduzir o número de iterações necessárias para a convergência.

Implementando o corretor, obtemos um resultado de 15,21178 ($\varepsilon_t = -2{,}48\%$), o que representa uma melhora em relação ao Exemplo 26.3, por causa da redução no erro global. Finalmente, o resultado pode ser modificado usando-se a Equação (26.27):

$$y_2^m = 15{,}21178 - \frac{15{,}21178 - 13{,}59423}{5} = 14{,}88827 \qquad \varepsilon_t = -0{,}30\%$$

Novamente, o erro foi reduzido de uma ordem de grandeza.

Como no exemplo anterior, a adição de modificadores aumenta tanto a eficiência quanto a acurácia dos métodos de passo múltiplo. Em particular, o modificador do corretor aumenta efetivamente a ordem da técnica. Assim, o método de Heun de início não-automático com modificadores é de ordem três em vez de ordem dois, como era o caso para a versão sem modificadores. Contudo, deve ser observado que existem situações nas quais o modificador do corretor irá afetar a estabilidade do processo de iteração do corretor. Como conseqüência, o modificador não é incluído no algoritmo de Heun de início não-automático delineado na Figura 26.5. Apesar disso, o modificador do corretor ainda pode ter utilidade no controle do tamanho do passo, como discutido a seguir.

FIGURA 26.5
A seqüência de fórmulas usadas para implementar o método de Heun de início não-automático. Observe que as estimativas de erro para o corretor podem ser usadas para modificar o corretor. Entretanto, como isso pode afetar a estabilidade do corretor, o modificador não é incluído nesse algoritmo. A estimativa de erro para o corretor é incluída por causa de sua utilidade para o ajuste do tamanho do passo.

Preditor:

$y_{i+1}^0 = y_{i-1}^m + f(x_i, y_i^m)2h$

(Armazene esse resultado como $y_{i+1,u}^0 = y_{i+1}^0$, em que o subscrito u indica que a variável não foi modificada.)

Modificador do Preditor:

$y_{i+1}^0 \leftarrow y_{i+1,u}^0 + \frac{4}{5}(y_{i,u}^m - y_{i,u}^0)$

Corretor:

$y_{i+1}^j = y_i^m + \frac{f(x_i, y_i^m) + f(x_{i+1}, y_{i+1}^{j-1})}{2}h$ (para $j = 1$ a um máximo m de iterações)

Verificação do Erro:

$|\varepsilon_a| = \left|\frac{y_{i+1}^j - y_{i+1}^{j-1}}{y_{i+1}^j}\right| 100\%$

(Se $|\varepsilon_a| >$ critério de erro, faça $j = j+1$ e repita o corretor; se $\varepsilon_a \leq$ critério de erro, armazene o resultado como $y_{i+1,u}^m = y_{i+1}^m$.)

Estimativa de Erro do Corretor:

$E_c = -\frac{1}{5}(y_{i+1,u}^m - y_{i+1,u}^0)$

(Se os cálculos forem continuar, faça $i = i + 1$ e volte ao preditor.)

26.2.2 Controle do Tamanho do Passo e Programas Computacionais

Tamanho de Passo Constante. É relativamente simples desenvolver uma versão com tamanho de passo constante do método de Heun com início não-automático. Uma das únicas complicações é que é necessário um método de passo único para gerar um ponto extra para iniciar os cálculos.

Além disso, como é usado um tamanho de passo constante, deve ser escolhido um valor para h antes dos cálculos. Em geral, a experiência indica que um tamanho de passo ótimo deveria ser suficientemente pequeno para garantir a convergência com duas iterações do corretor (Hull e Creemer, 1963). Além disso, ele deveria ser pequeno o bastante para fornecer um erro de truncamento suficientemente pequeno. Ao mesmo tempo, o tamanho do passo deveria ser tão grande quanto possível para minimizar o custo do tempo de execução e os erros de arredondamento. Como com os outros métodos para EDOs, a única maneira prática de avaliar o módulo do erro global é comparar os resultados para o mesmo problema, mas com o tamanho do passo dividido por 2.

Tamanho de Passo Variável. São usados tipicamente dois critérios para decidir se uma mudança no tamanho do passo é necessária. Primeiro, se a Equação (26.26) for maior do que algum critério de erro pré-especificado, o tamanho do passo é diminuído. Segundo, o tamanho do passo é escolhido de modo que o critério de convergência do corretor seja satisfeito em duas iterações. Esse critério tem por objetivo levar em conta os prós e contras entre a taxa de convergência e o número total de passos nos cálculos. Para valores pequenos de h, a convergência será mais rápida, porém, serão necessários mais passos. Para h maiores, a convergência é mais lenta, mas existirão menos passos. A experiência (Hull e Creemer, 1963) sugere que os passos totais serão minimizados se h for escolhido de modo que o corretor convirja dentro de duas iterações. Portanto, se forem necessárias mais de duas iterações, o tamanho do passo é diminuído, e se forem necessárias menos de duas iterações, o tamanho do passo é aumentado.

Embora essa estratégia especifique quando as modificações do passo devem ser feitas, ela não indica como eles devem ser modificados. Essa é uma questão crítica, porque os métodos de passo múltiplo exigem, por definição, diversos pontos para calcular um novo ponto. Uma vez que o tamanho do passo seja mudado, deve ser determinado um novo conjunto de pontos. Uma abordagem é recomeçar e usar o método de passo único para gerar um novo conjunto de pontos iniciais.

Uma estratégia mais eficiente que faz uso da informação existente é aumentar e diminuir o tamanho do passo dobrando-o ou dividindo-o por 2. Como descrito na Figura 26.6*b*, se um número suficiente de valores anteriores tiver sido gerado, o aumento do tamanho do passo dobrando-o é uma tarefa relativamente simples (Figura 26.6*c*). Tudo que é necessário é não perder de vista os subscritos, de modo que os valores antigos de x e y se tornem valores novos apropriados. Dividir o tamanho do passo por 2 é um pouco mais complicado, porque alguns dos novos valores não estarão disponíveis (Figura 26.6*a*). Entretanto, polinômios interpoladores desenvolvidos no Capítulo 18 podem ser usados para determinar esses valores intermediários.

De qualquer modo, a decisão de incorporar controle do tamanho do passo representa o balanço de prós e contras entre o investimento inicial na complexidade do programa e o retorno a longo prazo, por causa do aumento da eficiência. Obviamente, o tamanho e a importância do próprio problema terão uma grande influência nesse balanço. Felizmente, diversos pacotes e bibliotecas de software têm rotinas de passo múltiplo que você pode usar para obter soluções sem ter de programá-las a partir do zero. Mencionaremos algumas delas quando revisarmos os pacotes e bibliotecas no final do Capítulo 27.

26.2.3 Fórmulas de Integração

O método de Heun de início não-automático é característico da maioria dos métodos de passo múltiplo. Ele usa uma fórmula de integração aberta (o método do ponto médio) para fazer uma estimativa inicial. Esse passo do preditor exige um ponto dado anterior. A seguir, uma fórmula de integração fechada (regra do trapézio) é aplicada iterativamente para melhorar a solução.

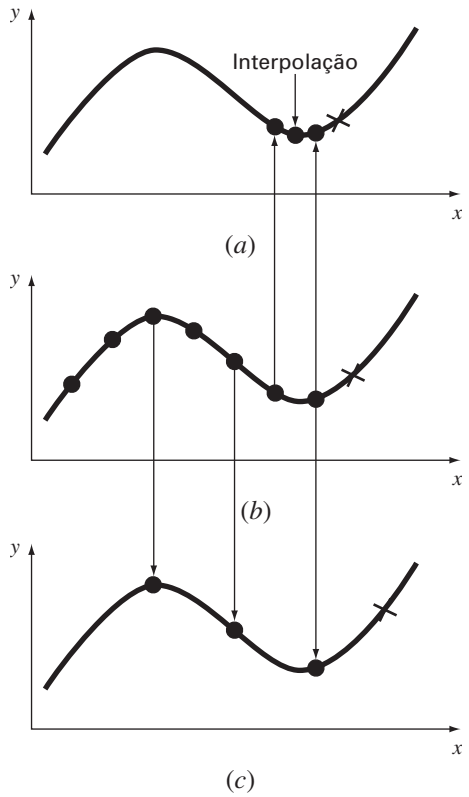


FIGURA 26.6
Um gráfico indicando como a estratégia de dividir por 2 e dobrar permite o uso de (*b*) valores calculados previamente para um método de passo múltiplo de terceira ordem. (*a*) Dividir por 2; (*c*) dobrar.

Deveria ser óbvio que uma estratégia para melhorar métodos de passo múltiplo seria usar fórmulas de integração de ordem superior como preditores e corretores. Por exemplo, as fórmulas de Newton-Cotes de ordem superior desenvolvidas no Capítulo 21 poderiam ser usadas para esse propósito.

Antes de descrever esses métodos de ordem superior, vamos revisar as fórmulas de integração mais comuns nas quais eles se baseiam. Como mencionado, as primeiras delas são as fórmulas de Newton-Cotes. Mas existe uma segunda classe chamada de fórmulas de Adams, que também revisaremos e que, muitas vezes, são preferidas. Como descrito na Figura 26.7, a diferença fundamental entre as fórmulas de Newton-Cotes e de Adams se relaciona com a maneira pela qual a integral é aplicada para se obter a solução. Como descrito na Figura 26.7*a*, as fórmulas de Newton-Cotes fazem uma estimativa da integral em um intervalo gerando diversos pontos. Essa integral então é usada para projetar do início do intervalo para o final. Em contraste, as fórmulas de Adams (Figura 26.7*b*) usam um conjunto de pontos de um intervalo para obter uma estimativa da integral somente para o último segmento do intervalo. Essa integral então é usada para projetar através do último segmento.

Fórmulas de Newton-Cotes. Algumas das fórmulas mais comuns para resolver equações diferenciais ordinárias baseiam-se em ajustar um polinômio interpolador de grau n a $n + 1$ valores conhecidos de y e então usar essa equação para calcular a integral. Como discutido no Capítulo 21, as fórmulas de integração de Newton-Cotes são baseadas nessa abordagem. Essas fórmulas são de dois tipos: abertas e fechadas.

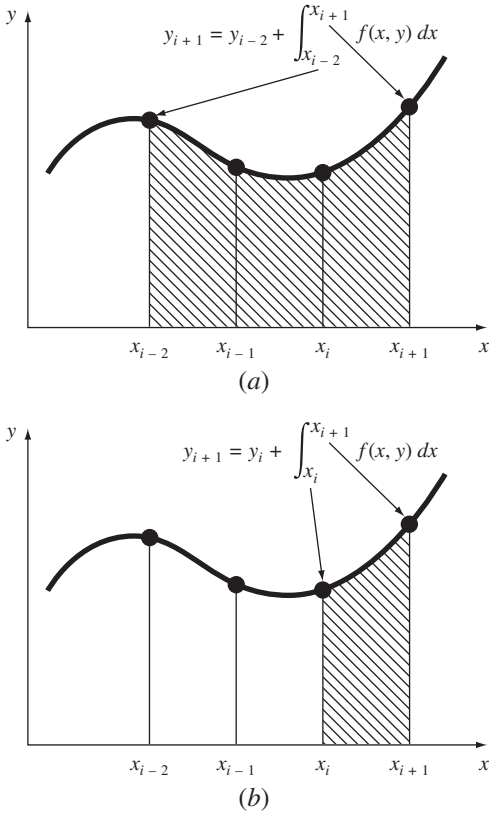


FIGURA 26.7
Ilustração da diferença fundamental entre as fórmulas de integração de Newton-Cotes e de Adams. (*a*) As fórmulas de Newton-Cotes usam uma série de pontos para obter uma estimativa da integral em diversos segmentos. A estimativa é então usada para projetar através de todo o intervalo. (*b*) As fórmulas de Adams usam uma série de pontos para obter uma estimativa da integral em um único segmento. A estimativa é então usada para projetar através do segmento.

***Fórmulas Abertas*.** Para n pontos dados igualmente espaçados, as fórmulas abertas podem ser expressas na forma de uma solução de uma EDO, como feito anteriormente para a Equação (26.19). A equação geral para esse propósito é

$$y_{i+1} = y_{i-n} + \int_{x_{i-n}}^{x_{i+1}} f_n(x)\,dx \tag{26.31}$$

onde $f_n(x)$ é um polinômio interpolador de grau n. O cálculo da integral usa a fórmula de integração aberta de Newton-Cotes de ordem n (Tabela 21.4). Por exemplo, se $n = 1$,

$$y_{i+1} = y_{i-1} + 2hf_i \tag{26.32}$$

onde f_i é uma abreviação de $f(x_i, y_i)$ — ou seja, a equação diferencial calculada em x_i e y_i. A Equação (26.32) é conhecida como o método do ponto médio e foi usada anteriormente como preditor no método de Heun com início não-automático. Para $n = 2$,

$$y_{i+1} = y_{i-2} + \frac{3h}{2}(f_i + f_{i-1})$$

e para $n = 3$,

$$y_{i+1} = y_{i-3} + \frac{4h}{3}(2f_i - f_{i-1} + 2f_{i-2}) \tag{26.33}$$

A Equação (26.33) é descrita graficamente na Figura 26.8*a*.

Fórmulas Fechadas. A fórmula fechada, em geral, pode ser expressa como

$$y_{i+1} = y_{i-n+1} + \int_{x_{i-n+1}}^{x_{i+1}} f_n(x)\,dx \tag{26.34}$$

onde a integral é aproximada por uma fórmula de integração fechada de Newton-Cotes de ordem n (Tabela 21.2). Por exemplo, para $n = 1$,

$$y_{i+1} = y_i + \frac{h}{2}(f_i + f_{i+1})$$

que é equivalente à regra do trapézio. Para $n = 2$,

$$y_{i+1} = y_{i-1} + \frac{h}{3}(f_{i-1} + 4f_i + f_{i+1}) \tag{26.35}$$

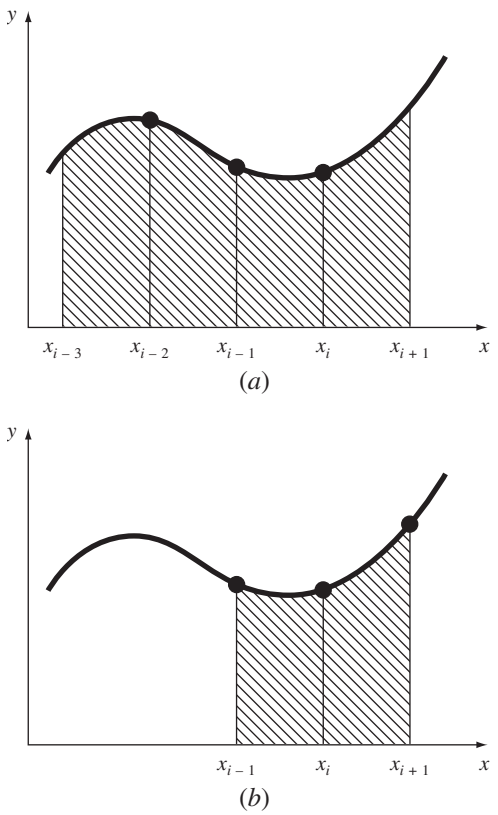
que é equivalente à regra 1/3 de Simpson. A Equação (26.35) é descrita na Figura 26.8*b*.

Fórmulas de Adams. O outro tipo de integração que pode ser usada para resolver EDOs são as fórmulas de Adams. Muitos algoritmos computacionais populares para solução de EDOs são baseados nesses métodos.

Fórmulas Abertas (Adams-Bashforth). As fórmulas de Adams podem ser deduzidas de diversas maneiras. Uma técnica é escrever uma expansão em série de Taylor progressiva em torno de x_i:

$$y_{i+1} = y_i + f_i h + \frac{f_i'}{2}h^2 + \frac{f_i''}{6}h^3 + \cdots$$

FIGURA 26.8
Descrição gráfica de fórmulas de integração de Newton-Cotes abertas e fechadas. (*a*) A terceira fórmula aberta [Equação (26.33)] e (b) regra 1/3 de Simpson [Equação (26.35)].



a qual também pode ser escrita como

$$y_{i+1} = y_i + h\left(f_i + \frac{h}{2}f_i' + \frac{h^2}{3!}f_i'' + \cdots\right) \tag{26.36}$$

Lembre-se (Seção 4.1.3) de que uma diferença regressiva pode aproximar a derivada:

$$f_i' = \frac{f_i - f_{i-1}}{h} + \frac{f_i''}{2}h + O(h^2)$$

a qual pode ser substituída na Equação (26.36),

$$y_{i+1} = y_i + h\left\{f_i + \frac{h}{2}\left[\frac{f_i - f_{i-1}}{h} + \frac{f_i''}{2}h + O(h^2)\right] + \frac{h^2}{6}f_i'' + \cdots\right\}$$

ou, agrupando os termos,

$$y_{i+1} = y_i + h\left(\frac{3}{2}f_i - \frac{1}{2}f_{i-1}\right) + \frac{5}{12}h^3 f_i'' + O(h^4) \tag{26.37}$$

Essa fórmula é chamada de *fórmula de Adams aberta de segunda ordem*. As fórmulas de Adams abertas também são conhecidas como *fórmulas de Adams-Bashforth*. Conseqüentemente, a Equação (26.37) algumas vezes é chamada de *segunda fórmula de Adams-Bashforth.*

Fórmulas de Adams-Bashforth de ordem superior podem ser deduzidas substituindo-se aproximações por diferenças de ordem superior na Equação (26.36). A fórmula de Adams aberta de ordem n pode ser representada, em geral, por

$$y_{i+1} = y_i + h\sum_{k=0}^{n-1}\beta_k f_{i-k} + O(h^{n+1}) \tag{26.38}$$

Os coeficientes β_k são compilados na Figura 26.1. A versão de quarta ordem é descrita na Figura 26.9*a*. Observe que a versão de primeira ordem é o método de Euler.

TABELA 26.1 Coeficientes e erro de truncamento para os preditores de Adams-Bashforth.

Ordem	β_0	β_1	β_2	β_3	β_4	β_5	Erro de Truncamento Local
1	1						$\frac{1}{2}h^2 f'(\xi)$
2	3/2	−1/2					$\frac{5}{12}h^3 f''(\xi)$
3	23/12	−16/12	5/12				$\frac{9}{24}h^4 f^{(3)}(\xi)$
4	55/24	−59/24	37/24	−9/24			$\frac{251}{720}h^5 f^{(4)}(\xi)$
5	1901/720	−2774/720	2616/720	−1274/720	251/720		$\frac{475}{1440}h^6 f^{(5)}(\xi)$
6	4277/720	−7923/720	9982/720	−7298/720	2877/720	−475/720	$\frac{19,087}{60,480}h^7 f^{(6)}(\xi)$

Fórmulas Fechadas (Adams-Moulton). Uma série de Taylor regressiva em torno de x_{i+1} pode ser escrita como

$$y_i = y_{i+1} - f_{i+1}h + \frac{f'_{i+1}}{2}h^2 - \frac{f''_{i+1}}{3!}h^3 + \cdots$$

Isolando y_{i+1}, obtemos

$$y_{i+1} = y_i + h\left(f_{i+1} - \frac{h}{2}f'_{i+1} + \frac{h^2}{6}f''_{i+1} + \cdots\right) \tag{26.39}$$

Uma diferença pode ser usada para aproximar a primeira derivada:

$$f'_{i+1} = \frac{f_{i+1} - f_i}{h} + \frac{f''_{i+1}}{2}h + O(h^2)$$

que pode ser substituída na Equação (26.39) e agrupando os termos obtemos

$$y_{i+1} = y_i + h\left(\frac{1}{2}f_{i+1} + \frac{1}{2}f_i\right) - \frac{1}{12}h^3 f''_{i+1} - O(h^4)$$

Essa fórmula é conhecida como *fórmula de Adams fechada de segunda ordem* ou *segunda fórmula de Adams-Moulton*. Além disso, observe que ela é a regra do trapézio.

A fórmula de Adams fechada de ordem n pode ser escrita em geral como

$$y_{i+1} = y_i + h\sum_{k=0}^{n-1} \beta_k f_{i+1-k} + O(h^{n+1})$$

Os coeficientes β_k são listados na Tabela 26.2. O método de quarta ordem é descrito na Figura 26.9*b*.

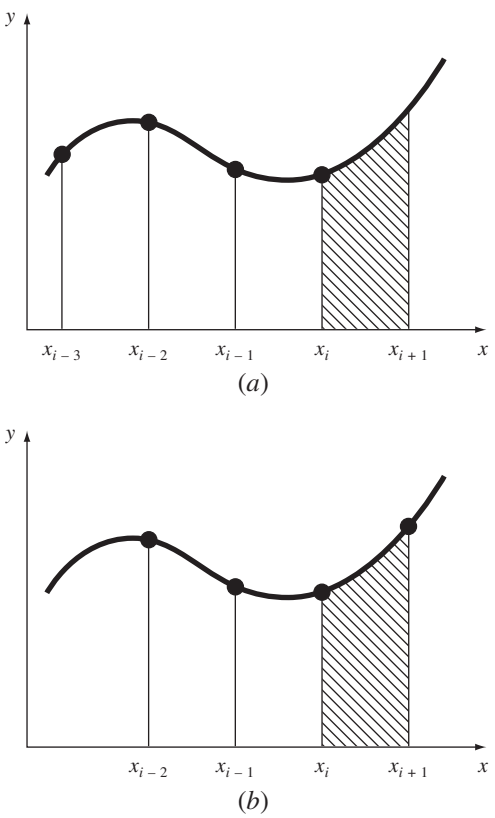


FIGURA 26.9
Descrição gráfica das fórmulas de integração de Adams abertas e fechadas. (*a*) A quarta fórmula aberta de Adams-Bashforth e (*b*) a quarta fórmula fechada de Adams-Moulton.

Quadro 26.1 Dedução das Relações Gerais para os Modificadores

A relação entre o valor verdadeiro, a aproximação e o erro de um preditor pode ser representada, em geral, por

$$\text{Valor verdadeiro} = y_{i+1}^0 + \frac{\eta_p}{\delta_p} h^{n+1} y^{(n+1)}(\xi_p) \qquad \text{(Q26.1.1)}$$

em que η_p e δ_p são o numerador e o denominador, respectivamente, da constante do erro de truncamento ou para um preditor de Newton-Cotes aberto (Tabela 21.4) ou um de Adams-Bashforth (Tabela 26.1), e n é a ordem.

Uma relação parecida pode ser desenvolvida para o corretor:

$$\text{Valor verdadeiro} = y_{i+1}^m - \frac{\eta_c}{\delta_c} h^{n+1} y^{(n+1)}(\xi_c) \qquad \text{(Q26.1.2)}$$

em que η_c e δ_c são o numerador e o denominador, respectivamente, da constante do erro de truncamento ou para um corretor de Newton-Cotes fechado (Tabela 21.2) ou um de Adams-Moulton (Tabela 26.2). Como foi feito na dedução da Equação (26.24), a Equação (Q26.1.1) pode ser subtraída da Equação (Q26.1.2) para fornecer

$$0 = y_{i+1}^m - y_{i+1}^0 - \frac{\eta_c + \eta_p \delta_c / \delta_p}{\delta_c} h^{n+1} y^{(n+1)}(\xi) \qquad \text{(Q26.1.3)}$$

Agora, dividindo a equação por $\eta_c + \eta_p \delta_c / \delta_p$, multiplicando o último termo por δ_p / δ_p e reorganizando, obtemos uma estimativa do erro de truncamento local do corretor:

$$E_c \cong -\frac{\eta_c \delta_p}{\eta_c \delta_p + \eta_p \delta_c} \left(y_{i+1}^m - y_{i+1}^0 \right) \qquad \text{(Q26.1.4)}$$

Para o modificador do preditor, a Equação (Q26.1.3) pode ser reescrita no passo anterior como

$$h^n y^{(n+1)}(\xi) = -\frac{\delta_c \delta_p}{\eta_c \delta_p + \eta_p \delta_c} \left(y_i^0 - y_i^m \right)$$

a qual pode ser substituída no termo de erro da Equação (Q26.1.1) para fornecer

$$E_p = \frac{\eta_p \delta_c}{\eta_c \delta_p + \eta_p \delta_c} \left(y_i^m - y_i^0 \right) \qquad \text{(Q26.1.5)}$$

As Equações (Q26.1.4) e (Q26.1.5) são versões gerais de modificadores que podem ser usados para melhorar os algoritmos de passo múltiplo. Por exemplo, o método de Milne tem $\eta_p = 14$, $\delta_p = 45$, $\eta_c = 1$, $\delta_c = 90$. Substituindo esses valores nas Equações (Q26.1.4) e (Q26.1.5), obtemos as Equações (26.43) e (26.42), respectivamente. Modificadores parecidos podem ser desenvolvidos para outros pares de fórmulas abertas e fechadas que tenham erros de truncamento locais de mesma ordem.

TABELA 26.2 Coeficientes e erros de truncamento para os corretores de Adams-Moulton.

Ordem	β_0	β_1	β_2	β_3	β_4	β_5	Erro de Truncamento Local
2	1/2	1/2					$-\frac{1}{12}h^3 f''(\xi)$
3	5/12	8/12	−1/12				$-\frac{1}{24}h^4 f^{(3)}(\xi)$
4	9/24	19/24	−5/24	1/24			$-\frac{19}{720}h^5 f^{(4)}(\xi)$
5	251/720	646/720	−264/720	106/720	−19/720		$-\frac{27}{1440}h^6 f^{(5)}(\xi)$
6	475/1440	1427/1440	−798/1440	482/1440	−173/1440	27/1440	$-\frac{863}{60{,}480}h^7 f^{(6)}(\xi)$

26.2.4 Métodos de Passo Múltiplo de Ordem Superior

Agora que desenvolvemos formalmente fórmulas de integração de Newton-Cotes e de Adams, podemos usá-las para deduzir métodos de passo múltiplo de ordem superior. Como foi o caso no método de Euler de início não-automático, as fórmulas de integração são aplicadas em conjunto como métodos preditor-corretor. Além disso, se as fórmulas abertas e fechadas tiverem erro de truncamento local da mesma ordem, modificadores do tipo listado na Figura 26.5 podem ser incorporados para melhorar a acurácia e permitir controle do tamanho do passo. O Quadro 26.1 fornece equações gerais para esses modificadores. Na próxima seção, apresentamos duas das abordagens mais comuns de passo múltiplo de ordem superior: o método de Milne e o método de Adams de quarta ordem.

Método de Milne. O método de Milne é o método de passo múltiplo mais comum baseado em fórmulas de integração de Newton-Cotes. Ele usa a fórmula aberta de Newton-Cotes de três pontos com um preditor:

$$y_{i+1}^0 = y_{i-3}^m + \frac{4h}{3}\left(2f_i^m - f_{i-1}^m + 2f_{i-2}^m\right) \tag{26.40}$$

e a fórmula fechada de Newton-Cotes de três pontos (regra 1/3 de Simpson) como um corretor:

$$y_{i+1}^j = y_{i-1}^m + \frac{h}{3}\left(f_{i-1}^m + 4f_i^m + f_{i+1}^{j-1}\right) \tag{26.41}$$

onde j é um índice representando o número de iterações do modificador. Os modificadores do preditor e do corretor para o método de Milne podem ser deduzidos das fórmulas no Quadro 26.1 e dos coeficientes de erro nas Tabelas 21.2 e 21.4:

$$E_p = \frac{28}{29}\left(y_i^m - y_i^0\right) \tag{26.42}$$

$$E_c \cong -\frac{1}{29}\left(y_{i+1}^m - y_{i+1}^0\right) \tag{26.43}$$

EXEMPLO 26.5 Método de Milne

Enunciado do Problema. Use o método de Milne para integrar $y' = 4e^{0,8x} - 0,5y$ de $x = 0$ a $x = 4$ utilizando um tamanho de passo de 1. A condição inicial em $x = 0$ é $y = 2$. Como estamos lidando com um método de passo múltiplo, são necessários pontos anteriores. Em uma aplicação real, um método de passo único tal como RK de quarta ordem seria usado para calcular os pontos necessários. No exemplo presente, vamos usar a solução analítica [lembre-se da Equação (E25.5.1) do Exemplo 25.5] para calcular valores exatos em $x_{i-3} = -3$, $x_{i-2} = -2$, e $x_{i-1} = -1$ de $y_{i-3} = -4,547302$, $y_{i-2} = -2,306160$, e $y_{i-1} = -0,3929953$, respectivamente.

Solução. O preditor [Equação (26.40)] é usado para calcular um valor em $x = 1$:

$$y_1^0 = -4,54730 + \frac{4(1)}{3}[2(3) - 1,99381 + 2(1,96067)] = 6,02272 \qquad \varepsilon_t = 2,8\%$$

O corretor [Equação (26.41) é então usado para calcular

$$y_1^1 = -0,3929953 + \frac{1}{3}[1,99381 + 4(3) + 5,890802] = 6,235210 \qquad \varepsilon_t = -0,66\%$$

Esse resultado pode ser substituído de volta na Equação (26.41) para corrigir iterativamente a estimativa. Esse processo converge para um valor final corrigido de 6,204855 ($\varepsilon_t = -0,17\%$).

Esse valor é mais acurado do que a estimativa comparável de 6,360865 ($\varepsilon_t = -2,68\%$) obtida anteriormente com o método de Heun de início não-automático (Exemplos 26.2 a 26.4). Os resultados dos passos restantes são $y(2) = 14,86031$ ($\varepsilon_t = -0,11\%$), $y(3) = 33,72426$ ($\varepsilon_t = -0,14\%$), e $y(4) = 75,43295$ ($\varepsilon_t = -0,12\%$).

Como no exemplo anterior, o método de Milne em geral fornece resultados de alta acurácia. Entretanto, há certos casos nos quais ele tem um desempenho fraco (ver Ralston e Rabinowitz, 1978). Antes de analisar esses casos, vamos descrever outra abordagem de passo múltiplo de ordem superior — o método de Adams de quarta ordem.

Método de Adams de Quarta Ordem. Um método de passo múltiplo popular baseado nas fórmulas de integração de Adams usa a fórmula de Adams-Bashforth de quarta ordem (Tabela 26.1) como o preditor:

$$y_{i+1}^0 = y_i^m + h\left(\frac{55}{24}f_i^m - \frac{59}{24}f_{i-1}^m + \frac{37}{24}f_{i-2}^m - \frac{9}{24}f_{i-3}^m\right) \tag{26.44}$$

e a fórmula de Adams-Moulton de quarta ordem (Tabela 26.2) como o corretor:

$$y_{i+1}^j = y_i^m + h\left(\frac{9}{24}f_{i+1}^{j-1} + \frac{19}{24}f_i^m - \frac{5}{24}f_{i-1}^m + \frac{1}{24}f_{i-2}^m\right) \tag{26.45}$$

Os modificadores para o preditor e para o corretor no método de Adams de quarta ordem podem ser desenvolvidos das fórmulas no Quadro 26.1 e dos coeficientes dos erros nas Tabelas 26.1 e 26.2 como

$$E_p = \frac{251}{270}(y_i^m - y_i^0) \tag{26.46}$$

$$E_c = -\frac{19}{270}(y_{i+1}^m - y_{i+1}^0) \tag{26.47}$$

EXEMPLO 26.6 Método de Adams de Quarta Ordem

Enunciado do Problema. Use o método de Adams de quarta ordem para resolver o mesmo problema que do Exemplo 26.5.

Solução. O preditor [Equação (26.44)] é usado para calcular um valor em $x = 1$.

$$y_1^0 = 2 + 1\left(\frac{55}{24}3 - \frac{59}{24}1{,}993814 + \frac{37}{24}1{,}960667 - \frac{9}{24}2{,}6365228\right) = 6{,}007539$$

$$\varepsilon_t = 3{,}1\%$$

o qual é comparável ao resultado usando o método de Milne, mas um pouco menos acurado. O corretor [Equação (26.45)] é então usado para calcular

$$y_1^1 = 2 + 1\left(\frac{9}{24}5{,}898394 + \frac{19}{24}3 - \frac{5}{24}1{,}993814 + \frac{1}{24}1{,}960666\right) = 6{,}253214$$

$$\varepsilon_t = -0{,}96\%$$

que novamente é comparável ao resultado usando o método de Milne, mas menos acurado. Esse resultado pode ser substituído de volta na Equação (26.45) para corrigir iterativamente a estimativa. O processo converge para um valor final corrigido de 6,214424 ($\varepsilon_t = 0{,}32\%$), que é um resultado acurado, mas novamente um pouco inferior ao obtido com o método de Milne.

Estabilidade dos Métodos de Passo Múltiplo. A acurácia superior do método de Milne exibida nos Exemplos 26.5 e 26.6 poderia ter sido antecipada com base nos termos de erro para os preditores [Equações (26.42) e (26.46)] e os corretores [Equações (26.43) e (26.47)]. Os coeficientes para o método de Milne, 14/45 e 1/90, são menores do que aqueles para o método de Adams de quarta ordem, 251/720 e 19/720. Além disso, o método de Milne emprega menos cálculos de funções para atingir essas acurácias superiores. À primeira vista, tais resultados poderiam levar à conclusão de que o método de Milne é superior e, portanto, preferível ao método de Adams de quarta ordem. Embora essa conclusão seja válida em muitos casos, existem situações nas quais o desempenho do método de Milne é inaceitável. Tal comportamento é exibido no seguinte exemplo.

EXEMPLO 26.7 Estabilidade dos Métodos de Milne e de Adams de Quarta Ordem

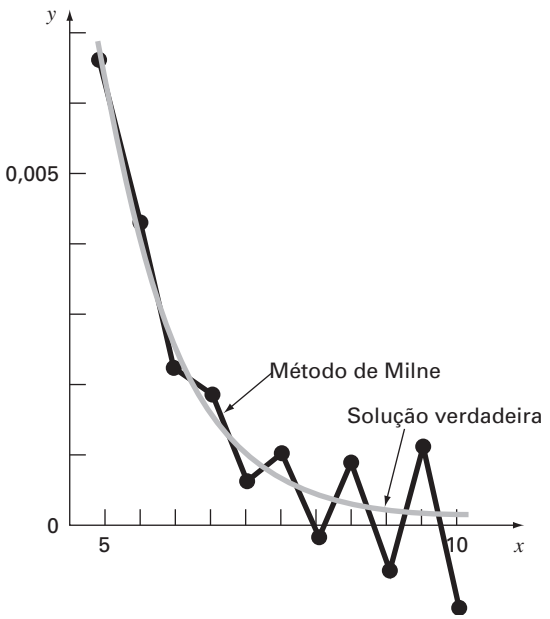
Enunciado do Problema. Use os métodos de Milne e de Adams de quarta ordem para resolver

$$\frac{dy}{dx} = -y$$

com a condição inicial de que $y = 1$ em $x = 0$. Resolva essa equação de $x = 0$ a $x = 10$, usando um tamanho de passo de $h = 0{,}5$. Observe que a solução analítica é $y = e^{-x}$.

Solução. Os resultados, como resumidos na Figura 26.10, indicam problemas com o método de Milne. Pouco depois do início dos cálculos, os erros começam a crescer, e seu sinal, oscilar. Em $x = 10$, o erro relativo cresceu para 2.831% e o sinal do próprio valor previsto começa a oscilar.

FIGURA 26.10
Descrição gráfica da instabilidade do método de Milne.



Em contraste, os resultados do método de Adams seriam muito mais aceitáveis. Embora o erro também cresça, ele faria isso a uma taxa menor. Além disso, as discrepâncias não exibiriam as grandes oscilações de sinal exibidas pelo método de Milne.

O comportamento inaceitável manifestado no exemplo anterior pelo método de Milne é conhecido como instabilidade. Embora ela não ocorra sempre, sua possibilidade leva à conclusão de que a abordagem de Milne deveria ser evitada. Logo, o método de Adams de quarta ordem é normalmente preferido.

A instabilidade no método de Milne decorre do corretor. Conseqüentemente, foram feitas tentativas de corrigir esse defeito, desenvolvendo-se corretores estáveis. Uma alternativa que é usada comumente e que utiliza essa abordagem é o *método de Hamming*, que usa o preditor de Milne e um corretor estável:

$$y_{i+1}^{j} = \frac{9y_i^m - y_{i-2}^m + 3h\left(y_{i+1}^{j-1} + 2f_i^m - f_{i-1}^m\right)}{8}$$

que tem o erro de truncamento local:

$$E_c = \frac{1}{40}h^5 y^{(4)}(\xi_c)$$

O método de Hamming também inclui modificadores da forma

$$E_p = \frac{9}{121}\left(y_i^m - y_i^0\right)$$

$$E_c = -\frac{112}{121}\left(y_{i+1}^m - y_{i+1}^0\right)$$

O leitor pode obter informações adicionais sobre esse e outros métodos de passo múltiplo em outras referências (Hamming, 1973; Lapidus e Seinfield, 1971).

PROBLEMAS

26.1 É dada

$$\frac{dy}{dx} = -200{,}000y + 200{,}000e^{-x} - e^{-x}$$

(a) Faça uma estimativa do tamanho do passo necessário para manter a estabilidade usando o método de Euler explícito.
(b) Se $y(0) = 0$, use o método de Euler implícito para obter uma solução de $x = 0$ a 2 usando um tamanho de passo de 0,1.

26.2 É dada

$$\frac{dy}{dt} = 30(\cos t - y) + 3\text{sen}\, t$$

Se $y(0) = 1$, use Euler implícito para obter uma solução de $t = 0$ a 4 usando um tamanho de passo de 0,4.

26.3 É dado o sistema

$$\frac{dx_1}{dt} = 1999x_1 + 2999x_2$$

$$\frac{dx_2}{dt} = -2000x_1 - 3000x_2$$

Se $x_1(0) = x_2(0) = 1$, obtenha uma solução de $t = 0$ a 0,2 usando um tamanho de passo de 0,05 com os métodos de Euler (**a**) explícito e (**b**) implícito.

26.4 Resolva o seguinte problema de valor inicial no intervalo de $t = 2$ a $t = 3$:

$$\frac{dy}{dt} = -0{,}4y + e^{-2t}$$

Use o método de Heun de início não-automático com um tamanho de passo de 0,5 e condições iniciais de $y(1{,}5) = 5{,}800007$ e $y(2{,}0) = 4{,}762673$. Itere o corretor até $\varepsilon_s = 0{,}1\%$. Calcule os erros relativos porcentuais verdadeiros ε_t para seus resultados com base na solução analítica.

26.5 Repita o Problema 26.4, mas use o método de Adams de quarta ordem. [Observação: $y(0{,}5) = 8{,}46909$ e $y(1{,}0) = 7{,}037566$.] Itere o corretor até $\varepsilon_s = 0{,}01\%$.

26.6 Resolva o seguinte problema de valor inicial de $t = 4$ a 5:

$$\frac{dy}{dt} = -\frac{2y}{t}$$

Use um tamanho de passo de 0,5 e valores iniciais de $y(2{,}5) = 0{,}48$, $y(3) = 0{,}333333$, $y(3{,}5) = 0{,}244898$ e $y(4) = 0{,}1875$. Obtenha sua solução usando as seguintes técnicas: (**a**) o método de Heun de início não-automático ($\varepsilon_s = 1\%$) e (**b**) o método de Adams de quarta ordem ($\varepsilon_s = 0{,}01\%$). [Observação: as respostas exatas obtidas analiticamente são $y(4{,}5) = 0{,}148148$ e $y(5) = 0{,}12$.] Calcule os erros relativos porcentuais verdadeiros ε_t para seus resultados.

26.7 Resolva o seguinte problema de valor inicial de $x = 0$ a $x = 0{,}75$:

$$\frac{dy}{dx} = yx^2 - y$$

Use o método de Heun de início não-automático com um tamanho de passo de 0,25. Se $y(0) = 1$, empregue um método RK de quarta ordem com um tamanho de passo de 0,25 para prever o valor inicial em $y(0{,}25)$.

26.8 Resolva o seguinte problema de valor inicial de $t = 1{,}5$ a $t = 2{,}5$

$$\frac{dy}{dt} = \frac{-2y}{1+t}$$

Use o método de Adams de quarta ordem. Utilize um tamanho de passo de 0,5 e o método de RK de quarta ordem para prever os valores iniciais se $y(0) = 2$.

26.9 Desenvolva um programa para o método de Euler implícito para uma única EDO linear. Teste-o duplicando o Problema 26.1*b*.

26.10 Desenvolva um programa para o método de Euler implícito para um par de EDOs lineares. Teste-o resolvendo a Equação (26.6).

26.11 Desenvolva um programa amigável ao usuário para o método de Heun de início não-automático com um modificador para o preditor. Empregue um método RK de quarta ordem para calcular os valores iniciais. Teste o programa duplicando o Exemplo 26.4.

26.12 Use o programa desenvolvido no Problema 26.11 para resolver o Problema 26.7.

26.13 Considere uma haste fina de comprimento l se movendo no plano xy como mostrado na Figura P26.13. A haste está fixada por um pino em uma extremidade e existe uma massa na outra. Observe que $g = 9{,}81$ m/s^2 e $l = 0{,}5$ m. Esse sistema pode ser resolvido usando

$$\ddot{\theta} - \frac{g}{l}\theta = 0$$

Seja $\theta(0) = 0$ e $\dot{\theta}(0) = 0{,}25$ rad/s. Resolva usando qualquer método estudado neste capítulo. Faça um gráfico do ângulo em função do tempo e da velocidade angular em função do tempo. (*Sugestão*: decomponha a EDO de segunda ordem.)

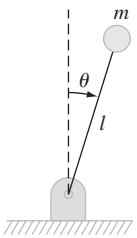


Figura P26.13

26.14 É dada a EDO de primeira ordem

$$\frac{dx}{dt} = -700x - 1000e^{-t}$$
$$x(t = 0) = 4$$

Resolva essa equação diferencial rígida usando um método numérico no período de tempo $0 \leq t \leq 5$. Resolva também analiticamente e trace a solução analítica e a numérica para o transiente rápido e a fase de transição lenta da escala de tempo.

26.15 A seguinte EDO de segunda ordem é considerada rígida

$$\frac{d^2y}{dx^2} = -1001\frac{dy}{dx} - 1000y$$

Resolva essa equação diferencial (**a**) analiticamente e (**b**) numericamente para $x = 0$ a 5. Para (**b**), use uma abordagem implícita com $h = 0{,}5$. Observe que as condições iniciais são $y(0) = 1$ e $y'(0) = 0$. Mostre ambos os resultados graficamente.

26.16 Resolva a seguinte equação diferencial de $t = 0$ a 1

$$\frac{dy}{dt} = -10y$$

com a condição inicial $y(0) = 1$. Use as seguintes técnicas para obter sua solução: (**a**) analiticamente, (**b**) método de Euler explícito e (**c**) método de Euler implícito. Para (**b**) e (**c**), use $h = 0{,}1$ e 0,2. Faça um gráfico de seus resultados.

CAPÍTULO 27

Problemas de Contorno e de Autovalores

Lembre-se de nossa discussão no início da Parte Sete em que uma equação diferencial ordinária é acompanhada por condições auxiliares. Essas condições são usadas para calcular as constantes de integração que aparecem durante a solução da equação. Para uma equação de ordem n, são necessárias n condições. Se todas as condições forem especificadas no mesmo valor da variável independente, então, estamos lidando com um *problema de valor inicial* (Figura 27.1*a*). Até esse ponto, o material da Parte Sete foi dedicado a esse tipo de problema.

Em contraste, existe outra aplicação para a qual as condições não são conhecidas em um único ponto, mas, em vez disso, são conhecidas em valores diferentes da variável independente. Como esses valores normalmente são especificados nas extremidades ou fronteiras do sistema, eles, em geral, são conhecidos como *problemas de contorno* (Figura 27.1*b*). Uma variedade de aplicações significativas na engenharia cai nessa classe. Neste capítulo, discutimos duas abordagens gerais para obter sua solução: o método *shooting* e a abordagem por diferença finita. Adicionalmente, apresentamos técnicas para abordar um tipo especial de problema de contorno: a determinação de autovalores. É claro que os autovalores também têm muitas aplicações além daquelas envolvidas nos problemas de contorno.

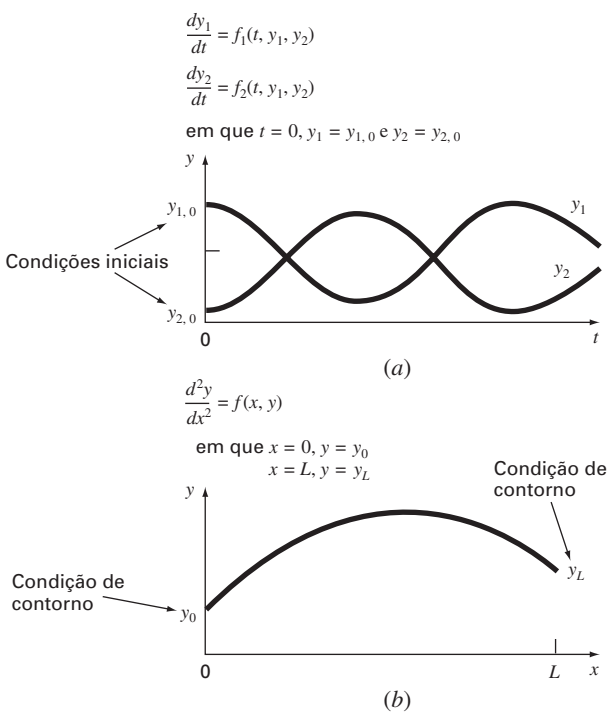


FIGURA 27.1
Problemas de valor inicial *versus* problemas de contorno. (*a*) Um problema de valor inicial no qual todas as condições são especificadas no mesmo valor da variável independente. (*b*) Um problema de contorno em que as condições são especificadas em valores diferentes da variável independente.

27.1 MÉTODOS GERAIS PARA PROBLEMAS DE CONTORNO

A conservação do calor pode ser usada para desenvolver um balanço de calor para uma haste longa e fina (Figura 27.2). Se a haste não estiver isolada ao longo de seu comprimento e se o sistema estiver em um estado estacionário, a equação resultante será

$$\frac{d^2T}{dx^2} + h'(T_a - T) = 0 \tag{27.1}$$

onde h' é o coeficiente de transferência de calor (m^{-2}) que parametriza as taxas de dissipação de calor para o ar em torno e T_a é a temperatura do ar em torno (ºC).

Para obter uma solução da Equação (27.1), devem existir condições de contorno adequadas. Um caso simples é quando as temperaturas nas extremidades da barra são mantidas em valores fixos. Estas podem ser expressas matematicamente por

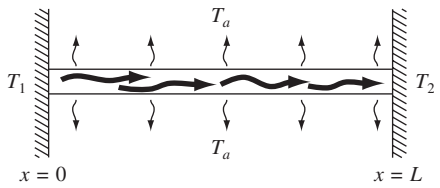
$$T(0) = T_1$$
$$T(L) = T_2$$

Com essas condições, a Equação (27.1) pode ser resolvida analiticamente usando-se o cálculo. Para uma haste de 10 m com $T_a = 20$, $T_1 = 40$, $T_2 = 200$ e $h' = 0{,}01$, a solução é

$$T = 73{,}4523e^{0{,}1x} - 53{,}4523e^{-0{,}1x} + 20 \tag{27.2}$$

Nas seções seguintes, o mesmo problema será resolvido usando-se abordagens numéricas.

FIGURA 27.2
Uma haste uniforme não isolada posicionada entre dois corpos de temperatura constante, mas diferente. Nesse caso, $T_1 > T_2$ e $T_2 > T_a$.



27.1.1 O Método *Shooting*

O *método shooting* (de atirar, em inglês) é baseado na conversão de um problema de contorno em um problema de valor inicial equivalente. Uma abordagem de tentativa e erro é, então, implementada para resolver a versão de valor inicial. Essa abordagem pode ser ilustrada por um exemplo.

EXEMPLO 27.1 O Método *Shooting*

Enunciado do Problema. Use o método *shooting* para resolver a Equação (27.1) para uma haste de 10 m com $h' = 0{,}01$ m^{-2}, $T_a = 20$ e condições de contorno

$$T(0) = 40 \qquad T(10) = 200$$

Solução. Utilizando a mesma abordagem que foi usada para transformar a Equação (PT7.2) nas Equações (PT7.3) a (PT7.6), a equação de segunda ordem pode ser expressa como duas EDOs de primeira ordem:

$$\frac{dT}{dx} = z \tag{E27.1.1}$$

$$\frac{dz}{dx} = h'(T - T_a) \tag{E27.1.2}$$

Para resolver essas equações, precisamos de um valor inicial para z. Para o método *shooting*, tentamos um valor — digamos, $z(0) = 10$. A solução é então obtida por integração das Equações (E27.1.1) e (E27.1.2) simultaneamente. Por exemplo, usando um método RK de quarta ordem com um tamanho de passo 2, obtemos um valor na extremidade do intervalo de $T(10) = 168{,}3797$ (Figura 27.3*a*), o que difere da condição de contorno de $T(10) = 200$. Portanto, fazemos outra tentativa, $z(0) = 20$, e fazemos os cálculos novamente. Desta vez, é obtido o resultado $T(10) = 285{,}8980$ (Figura 27.3*b*).

Agora, como a EDO original é linear, os valores

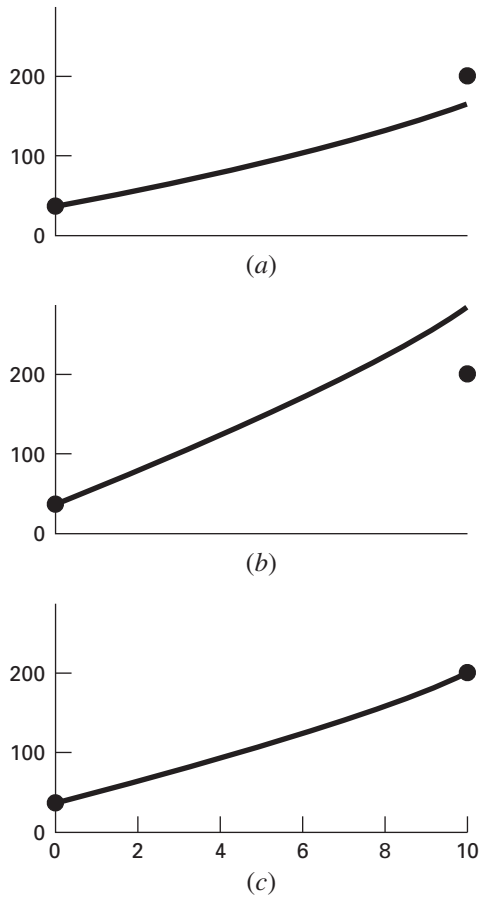
$$z(0) = 10 \qquad T(10) = 168{,}3797$$

e

$$z(0) = 20 \qquad T(10) = 285{,}8980$$

estão relacionados de forma linear. Como tal, eles podem ser usados para calcular o valor de $z(0)$ que fornece $T(10) = 200$. Uma fórmula de interpolação linear [lembre-se da Equação (18.2)] pode ser empregada para esse propósito:

FIGURA 27.3
O método *shooting*: (*a*) o primeiro "tiro", (*b*) o segundo "tiro" e (*c*) a "boa pontaria" final exata.



$$z(0) = 10 + \frac{20 - 10}{285{,}8980 - 168{,}3797}(200 - 168{,}3797) = 12{,}6907$$

Esse valor pode então ser usado para determinar a solução correta, como descrito na Figura 27.3*c*.

Problemas Não-lineares de Dois Pontos. Para problemas de contorno não-lineares, a interpolação ou extrapolação linear pelos dois pontos das soluções não resultará necessariamente em uma estimativa acurada da condição de contorno exigida para alcançar uma solução exata. Uma alternativa é fazer três aplicações do método *shooting* e usar um polinômio interpolador quadrático para obter uma estimativa da condição de contorno adequada. Entretanto, não é provável que tal abordagem forneça a resposta exata e iterações adicionais seriam necessárias para obter a solução.

Outra abordagem para um problema não-linear envolve reformulá-lo como um problema de raízes. Lembre-se de que a forma geral de um problema de raízes é encontrar o valor de x que torne a função $f(x) = 0$. Agora, vamos usar o Exemplo 27.1 para entender como o método *shooting* pode ser reformulado dessa forma.

Primeiro, perceba que a solução do par de equações diferenciais também é uma "função" no sentido de que tentamos uma condição na extremidade esquerda da barra, z_0, e a integração fornece uma previsão da temperatura na extremidade direita, T_{10}. Logo, podemos pensar na integração como

$$T_{10} = f(z_0)$$

Isto é, ela representa um processo pelo qual uma tentativa de z_0 fornece uma previsão de T_{10}. Visto dessa forma, podemos perceber que queremos o valor de z_0 que forneça um valor específico de T_{10}. Se, como no exemplo, quisermos $T_{10} = 200$, o problema pode ser enunciado como

$$200 = f(z_0)$$

Trazendo a meta de 200 para o lado direito da equação, geramos uma nova função, $g(z_0)$, que representa a diferença entre o que temos, $f(x_0)$, e o que queremos, 200.

$$g(z_0) = f(z_0) - 200$$

Se fizermos que essa nova função se anule, teremos obtido a solução. O próximo exemplo ilustra essa abordagem.

EXEMPLO 27.2 O Método *Shooting* para Problemas Não-lineares

Enunciado do Problema. Embora tenha servido aos nossos propósitos para resolver um problema de contorno simples, nosso modelo para a barra na Equação (27.1) não foi muito realista. Para começar, essa barra perderia calor por mecanismos tais como radiação, que são não-lineares.

Suponha que a seguinte EDO não-linear seja usada para simular a temperatura da barra aquecida:

$$\frac{d^2T}{dx^2} + h''(T_a - T)^4 = 0$$

onde $h'' = 5 \times 10^{-8}$. Agora, embora ainda não seja uma representação muito boa da transferência de calor, essa equação é simples o suficiente para nos permitir ilustrar como o método *shooting* pode ser usado para resolver um problema de contorno não-linear de dois pontos. As condições remanescentes do problema são como especificado no Exemplo 27.1.

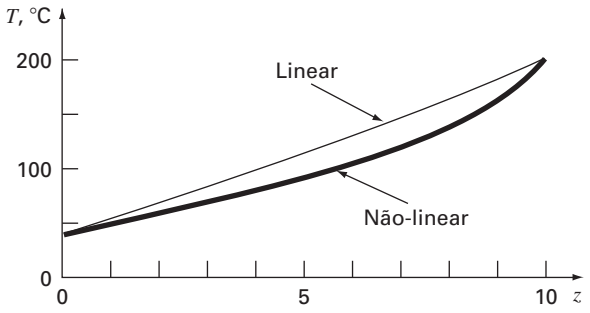


FIGURA 27.4
O resultado de usar o método *shooting* para resolver um problema não-linear.

Solução. A equação de segunda ordem pode ser expressa como duas EDOs:

$$\frac{dT}{dx} = z$$

$$\frac{dz}{dx} = h''(T - T_a)^4$$

Agora, essas equações podem ser integradas usando qualquer um dos métodos descritos nos Capítulos 25 e 26. Usamos uma versão com tamanho do passo constante da abordagem RK de quarta ordem descrita no Capítulo 25. Implementamos essa abordagem como uma função macro do Excel escrita em Visual BASIC. A função integrou a equação com base em uma tentativa inicial para $z(0)$ e retornou a temperatura em $x = 10$. A diferença entre esse valor e o objetivo de 200 foi então colocada em uma célula de planilha. O Excel Solver foi então chamado para ajustar o valor de $z(0)$ até que a diferença ficasse próxima a zero.

O resultado é mostrado na Figura 27.4 juntamente com o caso linear. Como poderia ser esperado, o caso não-linear é mais curvado que o modelo linear. Isso é devido ao termo de potência quatro na relação de transferência de calor.

O método *shooting* pode tornar-se trabalhoso para equações de ordem superior nas quais a necessidade de supor duas ou mais condições torna essa abordagem um pouco mais difícil. Por essas razões, estão disponíveis métodos alternativos, como descrito a seguir.

27.1.2 Métodos de Diferenças Finitas

As alternativas mais comuns ao método *shooting* são as *abordagens por diferenças finitas*. Nessas técnicas, as diferenças divididas finitas substituem as derivadas na equação original. Assim, uma equação diferencial linear é transformada em um conjunto de equações algébricas simultâneas que podem ser resolvidas usando os métodos da Parte Três.

Para o caso da Figura 27.2, a aproximação por diferenças divididas finitas para a segunda derivada é (lembre-se da Figura 23.3)

$$\frac{d^2T}{dx^2} = \frac{T_{i+1} - 2T_i + T_{i-1}}{\Delta x^2}$$

Essa aproximação pode ser substituída na Equação (27.1) para fornecer

$$\frac{T_{i+1} - 2T_i + T_{i-1}}{\Delta x^2} - h'(T_i - T_a) = 0$$

Agrupando os termos obtemos

$$-T_{i-1} + (2 + h'\Delta x^2)T_i - T_{i+1} = h'\Delta x^2 T_a \tag{27.3}$$

Essa equação se aplica para cada nó interior da haste. O primeiro e o último nós interiores, T_{i-1} e T_{i+1}, respectivamente, são especificados pelas condições de contorno. Portanto, o conjunto resultante de equações algébricas lineares será tridiagonal. Como tal, pode ser resolvido com os algoritmos eficientes que estão disponíveis para tais sistemas (Seção 11.1).

EXEMPLO 27.3 Aproximação por Diferença Finita de Problemas de Contorno

Enunciado do Problema. Use a abordagem por diferenças finitas para resolver o mesmo problema que no Exemplo 27.1.

Solução. Usando os parâmetros do Exemplo 27.1, podemos escrever a Equação (27.3) para a haste da Figura 27.2. Usando quatro nós interiores com um segmento de comprimento $\Delta x = 2$ m resulta nas seguintes equações:

$$\begin{bmatrix} 2{,}04 & -1 & 0 & 0 \\ -1 & 2{,}04 & -1 & 0 \\ 0 & -1 & 2{,}04 & -1 \\ 0 & 0 & -1 & 2{,}04 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} 40{,}8 \\ 0{,}8 \\ 0{,}8 \\ 200{,}8 \end{Bmatrix}$$

as quais pode ser resolvida por

$$\{T\}^T = \lfloor 65{,}9698 \quad 93{,}7785 \quad 124{,}5382 \quad 159{,}4795 \rfloor$$

A Tabela 27.1 fornece uma comparação entre a solução analítica [Equação (27.2)] e as soluções numéricas obtidas nos Exemplos 27.1 a 27.3. Observe que há alguma discrepância entre as abordagens. Para ambos os métodos numéricos, esses erros podem ser atenuados diminuindo-se seus respectivos tamanhos de passo. Embora ambas as técnicas tenham bom desempenho no caso presente, a abordagem por diferenças finitas é preferida por causa da facilidade com que pode ser estendida a casos mais complexos.

Além dos métodos *shooting* e de diferenças finitas, existem outras técnicas disponíveis para resolver problemas de contorno. Algumas delas serão descritas na Parte Oito. Estas incluem a solução estacionária (Capítulo 29) e transiente (Capítulo 30) de problemas de contorno bidimensionais usando diferenças finitas e soluções estacionárias de problema unidimensionais com abordagem por elementos finitos (Capítulo 31).

TABELA 27.1 Comparação da solução analítica exata com os métodos *shooting* e de diferenças finitas.

x	Verdadeiro	Método *Shooting*	Diferença Finita
0	40	40	40
2	65,9518	65,9520	65,9698
4	93,7478	93,7481	93,7785
6	124,5036	124,5039	124,5382
8	159,4534	159,4538	159,4795
10	200	200	200

27.2 PROBLEMAS DE AUTOVALORES

Problemas de *autovalores* ou de valores característicos são uma classe especial de problemas de contorno que são comuns em contextos de problemas em engenharia envolvendo vibrações, elasticidade e outros sistemas oscilantes. Além disso, eles são usados em uma ampla variedade de contextos de engenharia, além dos problemas de contorno. Antes de descrever os métodos numéricos para resolver esses problemas, vamos apresentar alguma informação geral básica. Isso inclui a discussão dos significados matemático e na engenharia dos autovalores.

27.2.1 Fundamentos Matemáticos

A Parte Três trata de métodos para resolver conjuntos de equações algébricas lineares da forma geral

$$[A]\{X\} = \{B\}$$

Tais sistemas são chamados *não homogêneos* por causa da presença do vetor $\{B\}$ no lado direito da igualdade. Se as equações que fazem parte de tal sistema forem linearmente independentes (isto é, tiverem um determinante não-nulo), elas terão uma única solução. Em outras palavras, existe um único conjunto de valores de x que balanceará as equações.

Em contraste, um sistema algébrico linear homogêneo tem a forma geral

$$[A]\{X\} = 0$$

Embora soluções não-triviais (isto é, soluções distintas de todos os x's iguais a zero) de tais sistemas sejam possíveis, elas, geralmente, não são únicas. Em vez disso, as equações simultâneas estabelecem relações entre os x's que podem ser satisfeitas por diversas combinações de valores.

Os problemas de autovalores associados com a engenharia são tipicamente da forma geral

$$\begin{aligned}
(a_{11} - \lambda)x_1 + \quad & a_{12}x_2 + \cdots + \quad a_{1n}x_n = 0 \\
a_{21}x_1 + \; & (a_{22} - \lambda)x_2 + \cdots + \quad a_{2n}x_n = 0 \\
\cdot \quad & \cdot \quad \cdot \quad \cdot \\
\cdot \quad & \cdot \quad \cdot \quad \cdot \\
\cdot \quad & \cdot \quad \cdot \quad \cdot \\
a_{n1}x_1 + \quad & a_{n2}x_2 + \cdots + (a_{nn} - \lambda)x_n = 0
\end{aligned}$$

onde λ é um parâmetro desconhecido chamado *autovalor* ou *valor característico*. Uma solução $\{X\}$ para tal sistema é conhecida como um *autovetor*. O conjunto de equações anteriores pode ser expresso de forma concisa como

$$\big[[A] - \lambda[I]\big]\{X\} = 0 \tag{27.4}$$

A solução da Equação (27.4) depende da determinação de λ. Uma maneira de conseguir isso é baseada no fato de que o determinante da matriz $[[A] - \lambda[I]]$ deve ser igual a zero para que soluções não-triviais sejam possíveis. Expandindo o determinante, obtemos um polinômio em λ. As raízes desse polinômio são as soluções para os autovalores. Um exemplo dessa abordagem será fornecido na próxima seção.

27.2.2 Fundamentos Físicos

O sistema massa-mola na Figura 27.5*a* é um contexto simples que ilustra como os autovalores ocorrem em um problema físico. Ele também vai ajudar a ilustrar alguns dos conceitos matemáticos introduzidos na seção anterior.

Para simplificar a análise, suponha que cada massa não esteja sobre a ação de forças externas ou de amortecimento. Além disso, suponha que cada mola tenha o mesmo comprimento natural l e a mesma constante da mola k. Finalmente, suponha que o deslocamento de cada mola seja medido em relação ao seu próprio sistema de coordenada local com a origem na posição de equilíbrio da mola (Figura 27.5*a*). Sob essas hipóteses, a

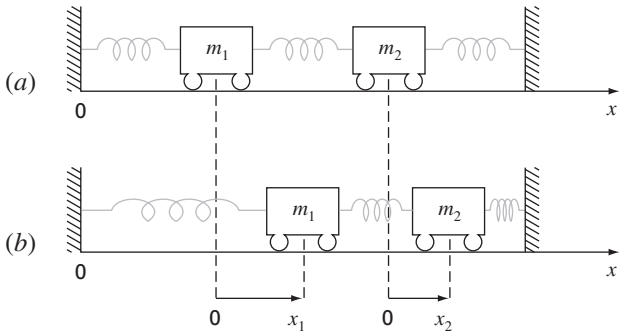


FIGURA 27.5
A colocação das massas fora do equilíbrio cria forças nas molas, que, depois de soltas, levam a oscilações das massas. As posições das massas podem ser dadas em relação a coordenadas locais com origens em suas respectivas posições de equilíbrio.

segunda lei de Newton pode ser usada para desenvolver um balanço de força para cada massa (lembre-se da Seção 12.4),

$$m_1\frac{d^2x_1}{dt^2} = -kx_1 + k(x_2 - x_1)$$

e

$$m_2\frac{d^2x_2}{dt^2} = -k(x_2 - x_1) - kx_2$$

onde x_i é o deslocamento da massa i de sua posição de equilíbrio (Figura 27.5*b*). Essas equações podem ser expressas como

$$m_1\frac{d^2x_1}{dt^2} - k(-2x_1 + x_2) = 0 \tag{27.5a}$$

$$m_2\frac{d^2x_2}{dt^2} - k(x_1 - 2x_2) = 0 \tag{27.5b}$$

Da teoria da vibração, é conhecido que algumas das soluções da Equação (27.5) tem a forma

$$x_i = A_i \operatorname{sen}(\omega t) \tag{27.6}$$

onde A_i é a amplitude da vibração da massa i e ω é a freqüência de vibração que é igual a

$$\omega = \frac{2\pi}{T_p} \tag{27.7}$$

na qual T_p é o período. Da Equação (27.6), segue que

$$x_i'' = -A_i\omega^2 \operatorname{sen}(\omega t) \tag{27.8}$$

As Equações (27.6) e (27.8) podem ser substituídas na Equação (27.5), a qual, depois de agruparmos os termos, pode ser expressa como

$$\left(\frac{2k}{m_1} - \omega^2\right)A_1 - \frac{k}{m_1}A_2 = 0 \tag{27.9a}$$

$$-\frac{k}{m_2}A_1 + \left(\frac{2k}{m_2} - \omega^2\right)A_2 = 0 \tag{27.9b}$$

A comparação da Equação (27.9) com a Equação (27.4) indica que, nesse ponto, a solução foi reduzida a um problema de autovalor.

EXEMPLO 27.4 Autovalores e Autovetores para um Sistema Massa-Mola

Enunciado do Problema. Calcule os autovalores e os autovetores da Equação (27.9) para o caso em que $m_1 = m_2 = 40$ kg e $k = 200$ N/m.

Solução. Substituindo-se os valores dos parâmetros na Equação (27.9), obtemos

$$(10 - \omega^2)A_1 - 5A_2 = 0$$

$$-5A_1 + (10 - \omega^2)A_2 = 0$$

O determinante desse sistema é [lembre-se da Equação (9.3)]

$$(\omega^2)^2 - 20\omega^2 + 75 = 0$$

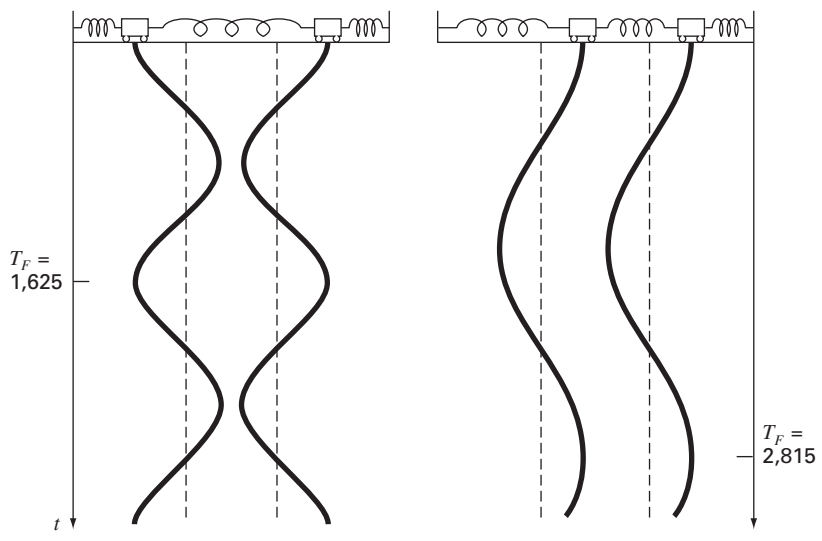
o qual pode ser resolvido pela fórmula quadrática como $\omega^2 = 15$ e $5\ s^{-2}$. Portanto, as freqüências para as vibrações das massas são $\omega = 3{,}873\ s^{-1}$ e $2{,}236\ s^{-1}$, respectivamente. Esses valores podem ser usados para determinar os períodos para as vibrações com a Equação (27.7). Para o primeiro modo, $T_p = 1{,}62$ s, e, para o segundo modo, $T_p = 2{,}81$ s.

Como afirmado na Seção 27.2.1, não pode ser obtido um único conjunto de valores para as incógnitas. Entretanto, suas razões podem ser especificadas pela substituição dos autovalores de volta nas equações. Por exemplo, para o primeiro modo ($\omega^2 = 15\ s^{-2}$), $A_1 = -A_2$. Para o segundo modo ($\omega^2 = 5\ s^{-2}$), $A_1 = A_2$.

Esse exemplo fornece informação valiosa com relação ao comportamento do sistema na Figura 27.5. Além de seu período, sabemos que, se o sistema estiver vibrando no primeiro modo, a amplitude da segunda massa será igual, mas de sinal oposto à amplitude da primeira. Como na Figura 27.6*a*, as massas vibram se afastando e depois se aproximando, indefinidamente.

No segundo modo, as duas massas têm amplitudes iguais em todos os instantes. Logo, como na Figura 27.6*b*, elas vibram para a frente e para trás juntas. Deve ser observado que a configuração das amplitudes fornece uma diretriz de como escolher os seus valores iniciais para alcançar movimentos puros em qualquer um dos dois modos. Qualquer outra configuração levará à superposição dos modos (lembre-se do Capítulo 19).

FIGURA 27.6
Os modos principais de vibração de duas massas iguais ligadas por três molas idênticas, entre paredes fixas.



(*a*) Primeiro modo (*b*) Segundo modo

27.2.3 Um Problema de Contorno

Agora que você já conheceu os autovalores, voltaremo-nos para o tipo de problema que é o assunto do capítulo presente: problemas de contorno para equações diferenciais ordinárias. A Figura 27.7 mostra um sistema físico que pode servir como contexto para o exame desse tipo de problema.

A curvatura de uma coluna delgada sujeita a uma carga axial P pode ser modelada por

$$\frac{d^2y}{dx^2} = \frac{M}{EI} \tag{27.10}$$

onde d^2y/dx^2 especifica a curvatura, M é o momento fletor, E é o modo de elasticidade e I é o momento de inércia da secção transversal em torno de seu eixo neutro. Considerando o corpo livre na Figura 27.7*b*, é claro que o momento fletor em x é $M = -Py$. Substituindo esse valor na Equação (27.10), temos

$$\frac{d^2y}{dx^2} + p^2y = 0 \tag{27.11}$$

onde

$$p^2 = \frac{P}{EI} \tag{27.12}$$

Para o sistema na Figura 27.7, sujeito às condições de contorno

$$y(0) = 0 \tag{27.13a}$$

$$y(L) = 0 \tag{27.13b}$$

a solução geral da Equação (27.11) é

$$y = A\ \text{sen}(px) + B\cos(px) \tag{27.14}$$

onde A e B são constantes arbitrárias que devem ser calculadas por meio das condições de contorno. De acordo com a primeira condição [Equação (27.13*a*)],

$$0 = A\ \text{sen}(0) + B\cos(0)$$

Portanto, concluímos que $B = 0$.

De acordo com a segunda condição [Equação (27.13*b*)],

$$0 = A\ \text{sen}\ (pL) + B\cos\ (pL)$$

Mas, como $B = 0$, A sen $(pL) = 0$. Como $A = 0$ representaria uma solução trivial, concluímos que sen $(pL) = 0$. Para essa igualdade ser válida,

$$pL = n\pi \qquad \text{para } n = 1, 2, 3, \ldots \tag{27.15}$$

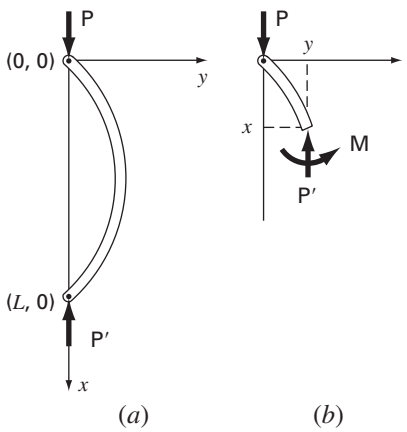


FIGURA 27.7
(*a*) Uma haste delgada. (*b*) Um diagrama de corpo livre da haste.

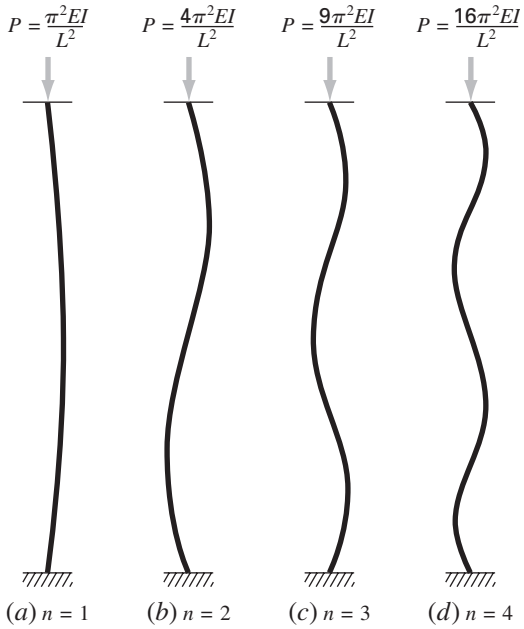


FIGURA 27.8
Os primeiros quatro autovalores para a haste delgada da Figura 27.7.

Logo, existe um número infinito de valores que satisfazem as condições de contorno. A Equação (27.15) pode ser resolvida por

$$p = \frac{n\pi}{L} \qquad \text{para } n = 1, 2, 3, \ldots \tag{27.16}$$

que são os autovalores para a coluna.

A Figura 27.8, que mostra a solução para os primeiros quatro autovalores, pode fornecer uma percepção do significado físico dos resultados. Cada autovalor corresponde a uma maneira pela qual a coluna se deforma. Combinando as Equações (27.12) e (27.16), obtemos

$$P = \frac{n^2\pi^2 EI}{L^2} \qquad \text{para } n = 1, 2, 3, \ldots \tag{27.17}$$

Isso pode ser pensado como *cargas de deformação*, pois elas representam o nível no qual a coluna se move em cada configuração sucessiva de deformação. No sentido prático, em geral, é o primeiro valor que tem interesse, porque a falha geralmente acontecerá quando a coluna se deformar inicialmente. Assim, uma carga crítica pode ser definida como

$$P = \frac{\pi^2 EI}{L^2}$$

a qual é formalmente conhecida como *fórmula de Euler*.

EXEMPLO 27.5 Análise de Autovalores de uma Coluna com Carga Axial

Enunciado do Problema. Uma coluna de madeira sob carga axial tem as seguintes características: $E = 10 \times 10^9$ Pa, $I = 1{,}25 \times 10^{-5}$ m^4, e $L = 3$ m. Determine os primeiros oito autovalores e as correspondentes cargas de deformação.

Solução. As Equações (27.16) e (27.17) podem ser usadas para calcular

n	*p*, m^{-2}	*P*, kN
1	1,0472	137,078
2	2,0944	548,311
3	3,1416	1.233,701
4	4,1888	2.193,245
5	5,2360	3.426,946
6	6,2832	4.934,802
7	7,3304	6.716,814
8	8,3776	8.772,982

A carga de deformação crítica é, portanto, 137,078 kN.

Embora as soluções analíticas do tipo que se obteve anteriormente sejam úteis, elas, em geral, são difíceis ou impossíveis de obter. Isso normalmente é verdade quando se trata de sistemas complicados ou daqueles com propriedades heterogêneas. Em tais casos, os métodos numéricos do tipo descrito a seguir são a única alternativa prática.

27.2.4 Método Polinomial

A Equação (27.11) pode ser resolvida numericamente substituindo a segunda derivada por uma aproximação por diferença finita dividida centrada (Figura 23.3). Dessa forma,

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + p^2 y_i = 0$$

o que pode ser expresso por

$$y_{i-1} - (2 - h^2 p^2) y_i + y_{i+1} = 0 \tag{27.18}$$

Escrevendo essa equação para uma série de nós ao longo do eixo da coluna, obtemos um sistema homogêneo de equações. Por exemplo, se a coluna for dividida em cinco segmentos (isto é, quatro nós interiores), o resultado é

$$\begin{bmatrix} (2 - h^2 p^2) & -1 & 0 & 0 \\ -1 & (2 - h^2 p^2) & -1 & 0 \\ 0 & -1 & (2 - h^2 p^2) & -1 \\ 0 & 0 & -1 & (2 - h^2 p^2) \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{Bmatrix} = 0 \tag{27.19}$$

A expansão do determinante do sistema fornece um polinômio, cujas raízes são os autovalores. Essa abordagem, chamada *método polinomial*, é adotada no seguinte exemplo.

EXEMPLO 27.6 Método Polinomial

Enunciado do Problema. Use o método polinomial para determinar os autovalores para a coluna com carga axial do Exemplo 27.5, usando (*a*) um, (*b*) dois, (*c*) três e (*d*) quatro nós interiores.

Solução.

(a) Escrevendo a Equação (27.18) para um nó interior, obtemos ($h = 3/2$)

$$-(2 - 2{,}25p^2) y_1 = 0$$

Logo, nesse caso simples, o autovalor é analisado igualando-se o determinante a zero

$$2 - 2{,}25p^2 = 0$$

e resolvendo por $p = \pm 0{,}9428$, o que é cerca de 10% menor que o valor exato de 1,0472 obtido no Exemplo 27.4.

(**b**) Para dois nós interiores ($h = 3/3$), a Equação (27.18) é escrita como

$$\begin{bmatrix} (2-p^2) & -1 \\ -1 & (2-p^2) \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} = 0$$

A expansão do determinante fornece

$$(2-p^2)^2 - 1 = 0$$

que pode ser resolvido por $p = \pm 1$ e $\pm 1{,}73205$. Logo, o primeiro autovalor é agora baixo por aproximadamente 4,5% e é obtido um segundo autovalor que é baixo por cerca de 17%.

(**c**) Para três pontos interiores ($h = 3/4$), a Equação (27.18) fornece

$$\begin{bmatrix} 2-0{,}5625p^2 & -1 & 0 \\ -1 & 2-0{,}5625p^2 & -1 \\ 0 & -1 & 2-0{,}5625p^2 \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \end{Bmatrix} = 0 \qquad \text{(E27.6.1)}$$

O determinante pode ser igualado a zero e expandido para fornecer

$$(2-0{,}5625p^2)^3 - 2(2-0{,}5625p^2) = 0$$

Para essa equação ser válida, $2 - 0{,}5625p^2 = 0$ ou $2 - 0{,}5625p^2 = \sqrt{2}$. Portanto, podemos determinar que os três primeiros autovalores são

$$p = \pm 1{,}0205 \qquad |\varepsilon_t| = 2{,}5\%$$
$$p = \pm 1{,}8856 \qquad |\varepsilon_t| = 10\%$$
$$p = \pm 2{,}4637 \qquad |\varepsilon_t| = 22\%$$

(**d**) Para quatro pontos interiores ($h = 3/5$), o resultado é a Equação (27.19) com $2 - 0{,}36p^2$ na diagonal. Igualando-se o determinante a zero e expandindo, obtemos

$$(2-0{,}36p^2)^4 - 3(2-0{,}36p^2)^2 + 1 = 0$$

que pode ser resolvido para determinar os primeiros quatro autovalores

$$p = \pm 1{,}0301 \qquad |\varepsilon_t| = 1{,}6\%$$
$$p = \pm 1{,}9593 \qquad |\varepsilon_t| = 6{,}5\%$$
$$p = \pm 2{,}6967 \qquad |\varepsilon_t| = 14\%$$
$$p = \pm 3{,}1702 \qquad |\varepsilon_t| = 24\%$$

TABELA 27.2 Os resultados da aplicação do método polinomial a uma coluna com carga axial. Os números entre parênteses representam o valor absoluto do erro relativo porcentual verdadeiro.

		Método Polinomial			
Autovalor	**Verdadeiro**	***h* = 3/2**	***h* = 3/3**	***h* = 3/4**	***h* = 3/5**
1	1,0472	0,9428 (10%)	1,0000 (4,5%)	1,0205 (2,5%)	1,0301 (1,6%)
2	2,0944		1,7321 (21%)	1,8856 (10%)	1,9593 (65%)
3	3,1416			2,4637 (22%)	2,6967 (14%)
4	4,1888				3,1702 (24%)

A Tabela 27.2, que resume os resultados desse exemplo, ilustra alguns aspectos fundamentais do método polinomial. Conforme a divisão em segmentos se torna mais fina, são determinados autovalores adicionais e os valores estabelecidos anteriormente se tornam progressivamente mais acurados. Assim, essa abordagem é mais adequada para os casos nos quais são necessários os autovalores mais baixos.

27.2.5 O Método da Potência

O *método da potência* é uma abordagem iterativa que pode ser usada para determinar o maior autovalor. Com pequenas modificações, também pode ser usado para determinar o menor valor e valores intermediários. Ele tem a vantagem adicional de que o autovetor correspondente é obtido como um subproduto do método.

Determinação do Maior Autovalor. Para implementar o método da potência, o sistema que está sendo analisado deve ser expresso na forma

$$[A]\{X\} = \lambda\{X\} \tag{27.20}$$

Como ilustrado no exemplo seguinte, a Equação (27.20) forma a base para uma técnica de solução iterativa que eventualmente fornece o maior autovalor e o seu autovetor associado.

EXEMPLO 27.7 Método da Potência para o Maior Autovalor

Enunciado do Problema. Use o método da potência para determinar o maior autovalor para a parte (**c**) do Exemplo 27.6.

Solução. O sistema, primeiro, é escrito na forma da Equação (27.20),

$$\begin{aligned} 3{,}556x_1 - 1{,}778x_2 &= \lambda x_1 \\ -1{,}778x_1 + 3{,}556x_2 - 1{,}778x_3 &= \lambda x_2 \\ -1{,}778x_2 + 3{,}556x_3 &= \lambda x_3 \end{aligned}$$

Então, supondo que os x's no lado esquerdo da equação são iguais a 1,

$$\begin{aligned} 3{,}556(1) - 1{,}778(1) &= 1{,}778 \\ -1{,}778(1) + 3{,}556(1) - 1{,}778(1) &= 0 \\ -1{,}778(1) + 3{,}556(1) &= 1{,}778 \end{aligned}$$

A seguir, o lado direito é normalizado por 1,778 para tornar o maior elemento igual a

$$\begin{Bmatrix} 1{,}778 \\ 0 \\ 1{,}778 \end{Bmatrix} = 1{,}778 \begin{Bmatrix} 1 \\ 0 \\ 1 \end{Bmatrix}$$

Desse modo, a primeira estimativa do autovalor é 1,778. Essa iteração pode ser expressa concisamente na forma matricial como

$$\begin{bmatrix} 3{,}556 & -1{,}778 & 0 \\ -1{,}778 & 3{,}556 & -1{,}778 \\ 0 & -1{,}778 & 3{,}556 \end{bmatrix} \begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 1{,}778 \\ 0 \\ 1{,}778 \end{Bmatrix} = 1{,}778 \begin{Bmatrix} 1 \\ 0 \\ 1 \end{Bmatrix}$$

A próxima iteração consiste em multiplicar $[A]$ por $\lfloor 1 \quad 0 \quad 1 \rfloor^T$ para obter

$$\begin{bmatrix} 3{,}556 & -1{,}778 & 0 \\ -1{,}778 & 3{,}556 & -1{,}778 \\ 0 & -1{,}778 & 3{,}556 \end{bmatrix} \begin{Bmatrix} 1 \\ 0 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 3{,}556 \\ -3{,}556 \\ 3{,}556 \end{Bmatrix} = 3{,}556 \begin{Bmatrix} 1 \\ -1 \\ 1 \end{Bmatrix}$$

Portanto, a estimativa do autovalor na segunda iteração é 3,556, que pode ser usada para determinar a estimativa de erro

$$|\varepsilon_a| = \left| \frac{3{,}556 - 1{,}778}{3{,}556} \right| 100\% = 50\%$$

O processo pode, então, ser repetido.

Terceira iteração:

$$\begin{bmatrix} 3{,}556 & -1{,}778 & 0 \\ -1{,}778 & 3{,}556 & -1{,}778 \\ 0 & -1{,}778 & 3{,}556 \end{bmatrix} \begin{Bmatrix} 1 \\ -1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 5{,}334 \\ -7{,}112 \\ 5{,}334 \end{Bmatrix} = -7{,}112 \begin{Bmatrix} -0{,}75 \\ 1 \\ -0{,}75 \end{Bmatrix}$$

em que $|\varepsilon_a| = 150\%$ (que é alto por causa da mudança de sinal).

Quarta iteração:

$$\begin{bmatrix} 3{,}556 & -1{,}778 & 0 \\ -1{,}778 & 3{,}556 & -1{,}778 \\ 0 & -1{,}778 & 3{,}556 \end{bmatrix} \begin{Bmatrix} -0{,}75 \\ 1 \\ -0{,}75 \end{Bmatrix} = \begin{Bmatrix} -4{,}445 \\ 6{,}223 \\ -4{,}445 \end{Bmatrix} = 6{,}223 \begin{Bmatrix} -0{,}714 \\ 1 \\ -0{,}714 \end{Bmatrix}$$

em que $|\varepsilon_a| = 214\%$ (novamente aumentado por causa da mudança de sinal).

Quinta iteração:

$$\begin{bmatrix} 3{,}556 & -1{,}778 & 0 \\ -1{,}778 & 3{,}556 & -1{,}778 \\ 0 & -1{,}778 & 3{,}556 \end{bmatrix} \begin{Bmatrix} -0{,}714 \\ 1 \\ -0{,}714 \end{Bmatrix} = \begin{Bmatrix} -4{,}317 \\ 6{,}095 \\ -4{,}317 \end{Bmatrix} = 6{,}095 \begin{Bmatrix} -0{,}708 \\ 1 \\ -0{,}708 \end{Bmatrix}$$

Assim, o fator normalizador está convergindo para o valor 6,070 ($= 2{,}4637^2$) obtido na parte (**c**) do Exemplo 27.6.

Observe que existem alguns casos nos quais o método da potência convergirá para o segundo maior autovalor em vez do maior. James, Smith e Wolford (1985) forneceram uma ilustração de tal caso. Outros casos especiais são discutidos em Fadeev e Fadeeva (1963).

Determinação do Menor Autovalor. Freqüentemente, existem casos em engenharia nos quais estamos interessados na determinação do menor autovalor. Esse era o caso para a haste na Figura 27.7, em que o menor autovalor podia ser usado para identificar uma carga fletora crítica. Isso pode ser feito aplicando-se o método da potência à matriz inversa de $[A]$. Nesse caso, o método da potência convergirá para o maior valor de $1/\lambda$ — em outras palavras, para o menor valor de λ.

EXEMPLO 27.8 Método da Potência para o Menor Autovalor

Enunciado do Problema. Utilize o método da potência para determinar o menor autovalor para a parte (**c**) do Exemplo 27.6.

Solução. Depois de dividir a Equação (E27.6.1) por h^2 ($= 0{,}5625$), sua matriz inversa pode ser calculada como

$$[A]^{-1} = \begin{bmatrix} 0{,}422 & 0{,}281 & 0.141 \\ 0{,}281 & 0{,}562 & 0{,}281 \\ 0{,}141 & 0{,}281 & 0{,}422 \end{bmatrix}$$

Usando a mesma forma que no Exemplo 27.7, o método da potência pode ser aplicado à matriz.

Primeira iteração:

$$\begin{bmatrix} 0{,}422 & 0{,}281 & 0{,}141 \\ 0{,}281 & 0{,}562 & 0{,}281 \\ 0{,}141 & 0{,}281 & 0{,}422 \end{bmatrix} \begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 0{,}884 \\ 1{,}124 \\ 0{,}884 \end{Bmatrix} = 1{,}124 \begin{Bmatrix} 0{,}751 \\ 1 \\ 0{,}751 \end{Bmatrix}$$

Segunda iteração:

$$\begin{bmatrix} 0{,}422 & 0{,}281 & 0{,}141 \\ 0{,}281 & 0{,}562 & 0{,}281 \\ 0{,}141 & 0{,}281 & 0{,}422 \end{bmatrix} \begin{Bmatrix} 0{,}751 \\ 1 \\ 0{,}751 \end{Bmatrix} = \begin{Bmatrix} 0{,}704 \\ 0{,}984 \\ 0{,}704 \end{Bmatrix} = 0{,}984 \begin{Bmatrix} 0{,}715 \\ 1 \\ 0{,}715 \end{Bmatrix}$$

onde $|\varepsilon_a| = 14{,}6\%$.

Terceira iteração:

$$\begin{bmatrix} 0{,}422 & 0{,}281 & 0{,}141 \\ 0{,}281 & 0{,}562 & 0{,}281 \\ 0{,}141 & 0{,}281 & 0{,}422 \end{bmatrix} \begin{Bmatrix} 0{,}715 \\ 1 \\ 0{,}715 \end{Bmatrix} = \begin{Bmatrix} 0{,}684 \\ 0{,}964 \\ 0{,}684 \end{Bmatrix} = 0{,}964 \begin{Bmatrix} 0{,}709 \\ 1 \\ 0{,}709 \end{Bmatrix}$$

onde $|\varepsilon_a| = 4\%$.

Dessa maneira, depois de apenas três iterações, o resultado está convergindo para o valor 0,9602, que é o inverso do menor autovalor, $1{,}0205 (= \sqrt{1/0{,}9602})$, obtido no Exemplo 27.6*c*.

Determinação de Autovalores Intermediários. Depois de encontrar o maior autovalor, é possível determinar o próximo mais alto substituindo a matriz original por uma que inclua apenas os autovalores restantes. O processo de remover o maior autovalor conhecido é chamado *deflação*. A técnica delineada aqui, o *método de Hotelling*, foi desenvolvida para matrizes simétricas. Isso ocorre porque ele explora a ortogonalidade dos autovetores de tais matrizes, o que pode ser expresso por

$$\{X\}_i^T \{X\}_j = \begin{cases} 0 & \text{para } i \neq j \\ 1 & \text{para } i = j \end{cases} \tag{27.21}$$

onde as componentes do autovetor $\{X\}$ foram normalizadas de modo que $\{X\}^T\{X\} = 1$, isto é, de forma que a soma dos quadrados das componentes seja igual a 1. Isso pode ser alcançado dividindo cada elemento pelo fator de normalização

$$\sqrt{\sum_{k=1}^{n} x_k^2}$$

Agora, uma nova matriz $[A]_2$ pode ser calculada por

$$[A]_2 = [A]_1 - \lambda_1 \{X\}_1 \{X\}_1^T \tag{27.22}$$

onde $[A]_1$ é a matriz original e λ_1 é o maior autovalor. Se o método da potência for aplicado a essa matriz, o processo iterativo convergirá para o segundo maior autovalor, λ_2. Para mostrar isso, inicialmente, multiplique à direita a Equação (27.22) por $\{X\}_1$,

$$[A]_2\{X\}_1 = [A]_1\{X\}_1 - \lambda_1 \{X\}_1 \{X\}_1^T \{X\}_1$$

O uso do princípio da ortogonalidade transforma essa equação em

$$[A]_2\{X\}_1 = [A]_1\{X\}_1 - \lambda_1\{X\}_1$$

onde o lado direito é igual a zero pela Equação (27.20). Logo, $[A]_2\{X\}_1 = 0$. Conseqüentemente, $\lambda = 0$ e $\{X\} = \{X\}_1$ é uma solução de $[A]_2\{X\} = \lambda\{X\}$. Em outras palavras, $[A]_2$ tem autovalores $0, \lambda_2, \lambda_3, \ldots, \lambda_n$. O maior autovalor, λ_1, foi substituído por um 0 e, portanto, o método da potência convergirá para o próximo maior λ_2.

O processo anterior pode ser repetido gerando-se uma nova matriz $[A]_3$ etc. Ainda que em teoria esse processo possa ser continuado para determinar os autovalores restantes, ele é limitado pelo fato de que os erros nos autovetores são propagados ao longo de cada passo. Assim, ele só é útil na determinação de alguns dos maiores autovalores. Embora algumas vezes isso seja um defeito, tal informação é precisamente o que é necessário em muitos problemas de engenharia.

27.2.6 Outros Métodos

Uma grande variedade de métodos adicionais está disponível para a resolução de problemas de autovalores. A maioria é baseada em processos de dois passos. O primeiro passo envolve a transformação da matriz original em uma forma mais simples (por exemplo, tridiagonal) que retém todos os autovalores originais. A seguir, métodos iterativos são usados para determinar esses autovalores.

Muitas dessas abordagens foram desenvolvidas para tipos especiais de matrizes. Em particular, diversas técnicas são dedicadas a sistemas simétricos. Por exemplo, o *método de Jacobi* transforma uma matriz simétrica em uma matriz diagonal eliminando os termos fora da diagonal de forma sistemática. Infelizmente, o método exige um número infinito de operações, porque a remoção de cada elemento não-nulo, em geral, cria um novo valor não-nulo em um elemento anteriormente nulo. Embora seja necessário um tempo infinito para eliminar todos os elementos não-nulos fora da diagonal, a matriz tenderá, eventualmente, a uma forma diagonal. Logo, a abordagem é iterativa, já que é repetida até que os termos fora da diagonal sejam "suficientemente" pequenos.

O *método de Given* também envolve a transformação de uma matriz simétrica em uma forma mais simples. Entretanto, em contraste com o método de Jacobi, a forma mais simples é tridiagonal. Além disso, ele difere no fato de que os zeros que são criados nas posições fora da diagonal são mantidos. Conseqüentemente, ele é finito e, portanto, mais eficiente que o método de Jacobi.

O *método de Householder* também transforma uma matriz simétrica em uma forma tridiagonal. Trata-se de um método finito e é mais eficiente que a abordagem de Given, pois reduz a zero linhas e colunas completas de elementos fora da diagonal.

Uma vez que um sistema tridiagonal seja obtido pelo método de Given ou de Householder, os passos remanescentes envolvem a determinação dos autovalores. Uma maneira direta de fazer isso é expandir o determinante. O resultado é uma seqüência de polinômios que podem ser calculados iterativamente para determinar os autovalores.

Além das matrizes simétricas, existem também técnicas que estão disponíveis quando todos os autovalores de uma matriz geral forem necessários. Estas incluem o *método LR* de Rutishauser e o *método QR* de Francis. Embora o método QR seja menos eficiente, em geral, ele é a abordagem preferida, porque é mais estável. Como tal, ele é considerado o melhor método de solução de propósito geral.

Finalmente, deveria ser mencionado que as técnicas descritas anteriormente são geralmente usadas em conjunto para aproveitar seus respectivos pontos fortes. Por exemplo, os métodos de Given e Householder podem ser aplicados a sistemas não-simétricos. O resultado não será tridiagonal, mas, em vez disso, um tipo especial chamado *forma de Hessenberg*. Uma abordagem é explorar a velocidade do método de Householder, empregando-a para transformar a matriz para essa forma e, então, usando o algoritmo estável QR para determinar os autovalores. Informação sobre estas e outras questões relacionadas a autovalores podem ser encontradas em Ralston e Rabinowitz (1978), Wilkinson (1965) Fadeev e Fadeeva (1963) e Householder (1953, 1964). Códigos computacionais podem ser enconrados em diversas fontes, incluindo Press at al. (1992). Rice (1983) discute pacotes de software disponíveis.

27.3 EDOS E AUTOVALORES COM BIBLIOTECAS E PACOTES

As bibliotecas e os pacotes de software têm grandes recursos para a resolução de EDOs e a determinação de autovalores. Esta seção esboça algumas formas nas quais eles podem ser aplicados para esse propósito.

27.3.1 Excel

Os recursos diretos do Excel para a resolução de problemas de autovalores e EDOs são limitados. Entretanto, se for feita alguma programação, por exemplo, eles podem ser combinados com ferramentas de visualização e otimização do Excel para implementar algumas aplicações interessantes. A Seção 28.1 fornece um exemplo de como o Excel Solver pode ser usado para estimativas de parâmetros de uma EDO.

27.3.2 MATLAB

Como poderia ser esperado, o pacote padrão do software MATLAB tem recursos excelentes para a determinação de autovalores e autovetores. Entretanto, ele também tem funções de biblioteca para a resolução de EDOs. As rotinas padrão de resolução de EDOs incluem duas funções que implementam o método de Runge-Kutta Fehlberg adaptativo no tamanho do passo (lembre-se da Seção 25.5.2). Elas são a **ODE23**, a qual usa fórmulas de segunda e terceira ordens para alcançar acurácia média, e a **ODE45**, a qual usa fórmulas de quarta e quinta ordens para alcançar alta acurácia. O seguinte exemplo ilustra como elas podem ser usadas para resolver um sistema de EDOs.

EXEMPLO 27.9 Usando MATLAB para Autovalores e EDOs

Enunciado do Problema. Explore como o MATLAB pode ser usado para resolver o seguinte conjunto de EDOs não-lineares de $t = 0$ a 20:

$$\frac{dx}{dt} = 1{,}2x - 0{,}6xy \qquad \frac{dy}{dt} = -0{,}8y + 0{,}3xy$$

onde $x = 2$ e $y = 1$ em $t = 0$. Como veremos no próximo capítulo (Seção 28.2), tais equações são conhecidas como *equações predador-presa*.

Solução. Antes de obter uma solução com o MATLAB, você deve usar um processador de texto para criar um M-file contendo o lado direito das EDOs. Esse M-file será então acessado pela rotina para a resolução de EDOs [em que $x = y(1)$ e $y = y(2)$]:

```
function yp = predprey(t,y)
yp = [1.2*y(1)-0.6*y(1)*y(2);-0.8*y(2)+0.3*y(1)*y(2)];
```

Armazenamos esse M-file sob o nome: predprey.m.

A seguir, iniciamos o MATLAB e inserimos os seguintes comandos para especificar o intervalo de integração e as condições iniciais:

```
>> tspan = [0,20];
>> y0=[2,1];
```

A rotina pode ser chamada por

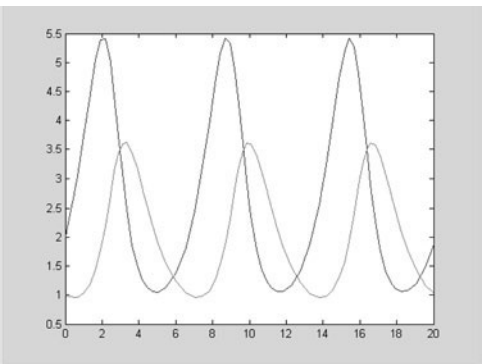
```
>> [t,y]=ode23('predprey',tspan,y0);
```

Esse comando, então, resolverá as equações diferenciais em predprey.m no intervalo definido por tspan usando as condições iniciais encontradas em y0. Os resultados podem ser exibidos simplesmente digitando

```
>> plot(t,y)
```

o que produz a Figura 27.9.

FIGURA 27.9
Solução do modelo predador-presa com o MATLAB.



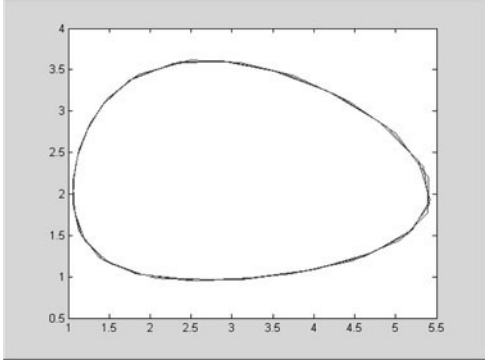


FIGURA 27.10
Gráfico no espaço de estados do modelo predador-presa com o MATLAB.

Além disso, também é instrutivo gerar um gráfico no espaço de estados, isto é, um gráfico das variáveis dependentes uma em função da outra por

```
>> plot(y(:,1),y(:,2))
```

o que produz a Figura 27.10.

O MATLAB também tem diversas funções projetadas para sistemas rígidos. Estas incluem a ODE15S e a ODE23S. Como no seguinte exemplo, elas são bem-sucedidas quando as funções padrão falham.

EXEMPLO 27.10 MATLAB para EDOs Rígidas

Enunciado do Problema. A *equação de Van der Pol* pode ser escrita como

$$\frac{dy_1}{dt} = y_2$$

$$\frac{dy_2}{dt} = \mu(1 - y_1^2)y_2 - y_1$$

Conforme o parâmetro μ se torna grande, o sistema se torna progressivamente mais rígido. Dadas as condições iniciais, $y_1(0) = y_2(0) = 1$, use o MATLAB para resolver os seguintes casos.

(a) Para $\mu = 1$, use ODE45 para resolver de $t = 0$ a 20.
(b) Para $\mu = 1000$, use ODE23S para resolver de $t = 0$ a 3000.

Solução.

(a) Um M-file pode ser criado para armazenar as equações diferenciais,

```
function yp = vanderpol(t,y)
yp=[y(2);1*(1-y(1)^2)*y(2)-y(1)];
```

Então, como no Exemplo 27.9, ODE45 pode ser chamada e o resultado mostrado em um gráfico (Figura 27.11),

```
>> tspan=[0,20];
>> y0=[1,1];
>> [t,y]=ode45('vanderpol',tspan,y0);
>> plot(t,y(:,1))
```

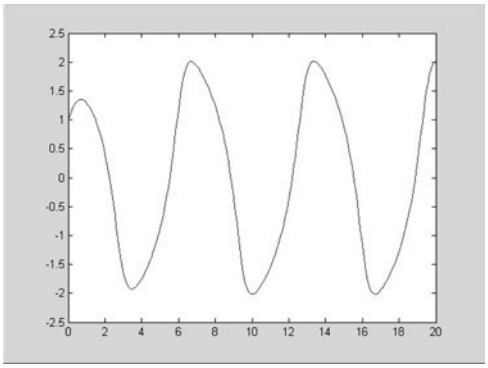


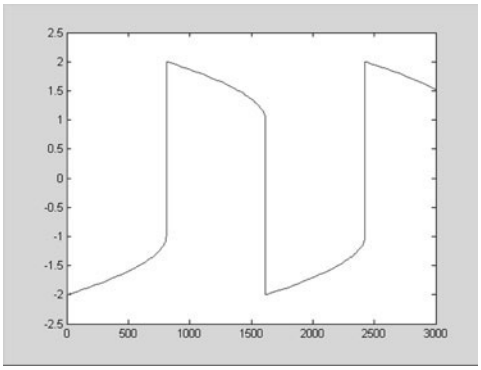
FIGURA 27.11
Forma não rígida da equação de Van der Pol resolvida com a função ODE45 do MATLAB.

(b) Se uma rotina-padrão como a ODE45 for usada para o caso rígido ($\mu = 1.000$), ela falhará miseravelmente (tente, se quiser). Entretanto, a ODE23S funciona de modo eficiente. Depois de revisar o M-file para refletir o novo valor de μ, a solução pode ser obtida e representada graficamente (Figura 27.12),

```
>> tspan=[0,3000];
>> y0=[1,1];
>> [t,y]=ode23S('vanderpol',tspan,y0);
>> plot(t,y(:,1))
```

Observe como essa solução tem quinas muito mais bruscas que no caso (**a**). Essa é uma manifestação visual da "rigidez" da solução.

FIGURA 27.12
Forma rígida da equação de Van der Pol resolvida com a função ODE23S do MATLAB.



Para autovalores, os recursos também são muito fáceis de aplicar. Lembre-se de que, em nossa discussão de sistemas rígidos no Capítulo 26, em que apresentamos o sistema rígido definido pela Equação (26.6), tais EDOs lineares podem ser escritas como um problema de alto valor da forma

$$\begin{bmatrix} 5-\lambda & -3 \\ -100 & 301-\lambda \end{bmatrix} \begin{Bmatrix} e_1 \\ e_2 \end{Bmatrix} = \{0\}$$

onde λ e $\{e\}$ são o autovalor e o autovetor, respectivamente.

O MATLAB pode então ser empregado para calcular tanto os autovalores (d) quanto os autovetores (v) com os seguintes comandos simples:

```
>> a=[5 -3;-100 301];
>> [v,d]=eig(a)

v =
   -0.9477     0.0101
   -0.3191    -0.9999

d =
    3.9899           0
         0     302.0101
```

Assim, vemos que os autovalores são bastante diferentes em módulo, o que é típico de um sistema rígido.

Os autovalores podem ser interpretados reconhecendo que a solução geral de um sistema de EDOs pode ser representada como uma soma de exponenciais. Por exemplo, a solução para o caso presente seria da forma

$$y_1 = c_{11}e^{-3,9899t} + c_{12}e^{-302,0101t}$$
$$y_2 = c_{21}e^{-3,9899t} + c_{22}e^{-302,0101t}$$

onde c_{ij} é a parte da condição inicial para y_i que está associada com o j-ésimo autovalor. Deve ser observado que os c's podem ser calculados a partir das condições iniciais e dos autovetores. Qualquer bom livro de equações diferenciais, por exemplo, Boyce e DiPrima (1992), fornecerá uma explicação de como isso pode ser feito.

Como no caso presente todos os autovalores são positivos (e, portanto, negativos na função exponencial), a solução consiste em uma série de exponenciais decrescentes. A que tem o maior autovalor (nesse caso, 302,0101) ditaria o tamanho do passo se uma técnica de solução explícita fosse usada.

27.3.3 IMSL

O IMSL tem diversas rotinas para a resolução de EDOs e determinação de autovalores (Tabela 27.3). Na presente discussão, nos concentraremos na rotina IVPRK. Essa rotina integra um sistema de EDOs usando o método de Runge-Kutta.

TABELA 27.3 Rotinas IMSL para resolver EDOs e determinar autovalores.

Categoria	Rotinas	Recursos
Equações Diferenciais Ordinárias de Primeira Ordem		
Solução de Problemas de Valor Inicial para EDOs		
	IVPRK	Método de Runge-Kutta
	IVPAG	Método de Adams ou Gear
Solução de Problemas de Contorno para EDOs		
	BVPFD	Método de diferenças finitas
	BVPMS	Método *shooting* múltiplo
Solução de Sistemas Diferenciais Algébricos		
Autovalores e (Opcionalmente) Autovetores de $Ax = \lambda x$		
Problema geral real $Ax = \lambda x$		
	EVLRG	Todos os autovalores
	EVCRG	Todos os autovalores e autovetores
	EPIRG	Índice de desempenho
Problema geral complexo $Ax = \lambda x$		
Problema simétrico real $Ax = \lambda x$		
Matrizes simétricas de banda reais no modo de armazenagem por banda		
Matrizes hermitianas complexas		
Matrizes de Hessenberg superiores reais		
Matrizes de Hessenberg superiores complexas		
Autovalores e (opcionalmente) autovetores de $Ax = \lambda Bx$		
Problema geral real $Ax = \lambda Bx$		
Problema geral complexo $Ax = \lambda Bx$		
Problema simétrico real $Ax = \lambda Bx$		

A IVPRK é implementada pelo seguinte comando CALL:

```
CALL IVPRK (IDO, N, FCN, T, TEND, TOL, PARAM, Y)
```

onde

IDO = Flag indicando o estado do cálculo. Normalmente, a chamada inicial é feita com IDO = 1. A rotina então faz IDO = 2 e esse valor é usado para todas as chamadas, exceto a última que é feita quando IDO = 3. Essa chamada final é feita para liberar espaço de trabalho, que foi automaticamente alocado pela chamada inicial IDO = 1. Nenhuma integração é feita nessa chamada final.

N = Número de equações diferenciais. (Entrada)

FCN = Sub-rotina fornecida pelo usuário para calcular funções.

T = Variável independente. (Entrada/Saída) Na entrada, T contém o valor inicial. Na saída, T é substituído por TEND, a menos que condições de erro tenham ocorrido.

TEND = Valor de t em que a solução é necessária. (Entrada) O valor TEND pode ser menor que o valor inicial de t.

TOL = Tolerância para o controle de erro. (Entrada) É feita uma tentativa para controlar a norma do erro local de modo que o erro global seja proporcional a TOL.

PARAM = Um vetor em ponto flutuante de tamanho 50, contendo parâmetros opcionais.

Y = Vetor de tamanho N das variáveis dependentes. (Entrada/Saída). Na entrada, Y contém os valores iniciais. Na saída, Y possui a solução aproximada.

A sub-rotina FCN deve ser escrita para armazenar as equações diferenciais. Ela deve ter a forma geral,

```
subroutine fcn (n, t, y, yprime)
integer    n
real       t, y(n), yprime(n)
yprime(1) = . . .
yprime(2) = . . .
return
end
```

sendo "yprime(*i*) = ..." o local onde a *i*-ésima EDO é escrita. A FCN deve ser declarada externa no programa que a chama.

EXEMPLO 27.11 Usando IMCL para Resolver EDOs

Enunciado do Problema. Use IVPRK para resolver as mesmas EDOs predador-presa que no Exemplo 27.9.

Solução. Um exemplo de um programa principal em Fortran 90 e de função usando IVPRK para esse problema pode ser escrito como

```
Program PredPrey
USE msimsl
INTEGER :: mxparm, n
PARAMETER (mxparm=50, n=2)
INTEGER :: ido, istep, nout
REAL :: param(mxparm), t, tend, tol, y(n)
EXTERNAL fcn
CALL UMACH (2, nout)
t = 0.0
y(1) = 2.0
y(2) = 1.0
tol = 0.0005
CALL SSET (mxparm, 0.0, param, 1)
```

SOFTWARE

```
param(10) = 1.0
PRINT '(4X, "istep", 5X, "Time", 9X, "Y1", 11X, "Y2")'
ido = 1
istep = 0
WRITE (nout,'(I6,3F12.3)') istep, t, y
DO
  istep = istep + 1
  tend = istep
  CALL IVPRK (ido, n, fcn, t, tend, tol, param, y)
  IF (istep .GT. 10) EXIT
  WRITE (nout,'(I6,3F12.3)') istep, t, y
  IF (istep .EQ. 10) ido = 3
END DO
END PROGRAM

SUBROUTINE fcn (n, t, y, yprime)
IMPLICIT NONE
INTEGER :: n
REAL :: t, y(n), yprime(n)
yprime(1) = 1.2*y(1) - 0.6*y(1)*y(2)
yprime(2) = -0.8*y(2) + 0.3*y(1)*y(2)
END SUBROUTINE
```

Um exemplo de execução é

```
istep     time      y1       y2
    0        .000    2.000    1.000
    1       1.000    3.703    1.031
    2       2.000    5.433    1.905
    3       3.000    3.390    3.533
    4       4.000    1.407    3.073
    5       5.000    1.048    1.951
    6       6.000    1.367    1.241
    7       7.000    2.393     .959
    8       8.000    4.344    1.161
    9       9.000    5.287    2.421
   10      10.000    2.561    3.624
```

PROBLEMAS

27.1 Um balanço de calor estacionário para uma haste pode ser representado por

$$\frac{d^2T}{dx^2} - 0{,}15T = 0$$

Obtenha uma solução analítica para uma haste de 10 m com $T(0) = 240$ e $T(10) = 150$.

27.2 Use o método *shooting* para resolver o Problema 27.1.

27.3 Use a abordagem por diferenças finitas com $\Delta x = 1$ para resolver o Problema 27.1.

27.4 Use o método *shooting* para resolver

$$7\frac{d^2y}{dx^2} - 2\frac{dy}{dx} - y + x = 0$$

com condições de contorno $y(0) = 5$ e $y(20) = 8$.

27.5 Resolva o Problema 27.4 com a abordagem por diferenças finitas, usando $\Delta x = 2$.

27.6 Use o método *shooting* para resolver

$$\frac{d^2T}{dx^2} - 1 \times 10^{-7}(T + 273)^4 + 4(150 - T) = 0 \qquad \text{(P27.6)}$$

Obtenha uma solução para as condições de contorno: $T(0) = 200$ e $T(0{,}5) = 100$.

27.7 As equações diferenciais como aquela que resolvemos no Problema 27.6 podem ser simplificadas linearizando seus termos não-lineares. Por exemplo, uma expansão em série de Taylor de primeira ordem pode ser usada para linearizar o termo de quarto grau na Equação (P27.6) como

$$1 \times 10^{-7}(T + 273)^4 = 1 \times 10^{-7}(T_b + 273)^4 + 4 \times 10^{-7}(T_b + 273)^3(T - T_b)$$

em que T_b é a temperatura base em torno da qual o termo é linearizado. Substitua essa relação na Equação (P27.6) e então resolva a equação linear resultante com a abordagem por diferenças finitas. Utilize $T_b = 150$ e $\Delta x = 0{,}01$ para obter sua solução.

27.8 Repita o Exemplo 27.4, mas para três massas. Produza um gráfico como o da Figura 27.6 para identificar os modos principais de vibração. Mude todos os k's para 240.
27.9 Repita o Exemplo 27.6, mas para cinco pontos interiores ($h = 3/6$).
27.10 Use menores para expandir o determinante de

$$\begin{bmatrix} 2-\lambda & 8 & 10 \\ 8 & 4-\lambda & 5 \\ 10 & 5 & 7-\lambda \end{bmatrix}$$

27.11 Use o método da potência para determinar o maior autovalor e o autovetor correspondente para o Problema 27.10.
27.12 Utilize o método da potência para determinar o menor autovalor e o autovetor correspondente para o Problema 27.10.
27.13 Desenvolva um programa computacional amigável ao usuário para implementar o método *shooting* para uma EDO linear de segunda ordem. Teste seu programa duplicando o Exemplo 27.1.
27.14 Use o programa desenvolvido no Problema 27.13 para resolver os Problemas 27.2 e 27.4.
27.15 Desenvolva um programa computacional amigável ao usuário para implementar a abordagem por diferenças finitas para resolver uma EDO linear de segunda ordem. Teste-o, duplicando o Exemplo 27.3.
27.16 Utilize o programa desenvolvido no Problema 27.15 para resolver os Problemas 27.3 e 27.5.
27.17 Desenvolva um programa amigável ao usuário para determinar o maior autovalor com o método da potência. Teste-o, duplicando o Exemplo 27.7.
27.18 Desenvolva um programa amigável ao usuário para determinar o menor autovalor com o método da potência. Teste-o, duplicando o Exemplo 27.8.
27.19 Use o Excel Solver para resolver diretamente (isto é, sem linearizar) o Problema 27.6, usando a abordagem por diferenças finitas. Utilize $\Delta x = 0{,}1$ para obter sua solução.
27.20 Use o MATLAB para integrar o seguinte par de EDOs de $t = 0$ a 100:

$$\frac{dy_1}{dt} = 0{,}35y_1 - 1{,}6y_1y_2 \qquad \frac{dy_2}{dt} = 0{,}04y_1y_2 - 0{,}15y_2$$

em que $y_1 = 1$ e $y_2 = 0{,}05$ em $t = 0$. Desenvolva um gráfico no espaço de estados (y_1 em função de y_2) para seus resultados.
27.21 A seguinte equação diferencial foi usada na Seção 8.4 para analisar as vibrações do amortecedor de um automóvel:

$$1{,}25 \times 10^6 \frac{d^2x}{dt^2} + 1 \times 10^7 \frac{dx}{dt} + 1{,}5 \times 10^9 x = 0$$

Transforme essa equação em um par de EDOs. (**a**) Use o MATLAB para resolver essas equações de $t = 0$ a 0,4 para o caso em que $x = 0{,}5$ e $dx/dt = 0$ em $t = 0$. (**b**) Use o MATLAB para determinar os autovalores e os autovetores do sistema.
27.22 Utilize o IMSL para integrar

(a)
$$\frac{dx}{dt} = ax - bxy$$
$$\frac{dy}{dt} = -cy + dxy$$

em que $a = 1{,}5$, $b = 0{,}7$, $c = 0{,}9$ e $d = 0{,}4$. Empregue condições iniciais $x = 2$ e $y = 1$ e integre de $t = 0$ a 30.

(b)
$$\frac{dx}{dt} = -\sigma x + \sigma y$$
$$\frac{dy}{dt} = rx - y - xz$$
$$\frac{dz}{dt} = -bz + xy$$

em que $\sigma = 10$, $b = 2{,}666667$ e $r = 28$. Empregue condições iniciais $x = y = z = 5$ e integre de $t = 0$ a 20.
27.23 Use diferenças finitas para resolver o problema de contorno da equação diferencial ordinária

$$\frac{d^2u}{dx^2} + 6\frac{du}{dx} - u = 2$$

com condições de contorno $u(0) = 10$ e $u(2) = 1$. Trace seus resultados para u em função de x. Utilize $\Delta x = 0{,}1$.
27.24 Resolva, usando o método de diferenças finitas, a EDO adimensional que descreve a distribuição de temperaturas em uma haste circular com uma fonte de calor interna S

$$\frac{d^2T}{dr^2} + \frac{1}{r}\frac{dT}{dr} + S = 0$$

no intervalo $0 \le r \le 1$, com condições de contorno

$$T(r = 1) = 1 \qquad \left.\frac{dT}{dr}\right|_{r=0} = 0$$

para $S = 1$, 10 e 20 K/m^2. Trace a temperatura em função do raio.
27.25 Deduza o conjunto de equações diferenciais para o sistema de três massas e quatro molas (Figura P27.25) que descreve seu movimento no tempo. Escreva as três equações diferenciais na forma matricial.

[Vetor aceleração]+ matriz[k/m][vetor deslocamento x] = 0

Observe que cada equação foi dividida pela massa. Determine os autovalores e as freqüências naturais para os seguintes valores das massas e das constantes das molas: $k_1 = k_4 = 15$ N/m, $k_2 = k_3 = 35$ N/m e $m_1 = m_2 = m_3 = 1{,}5$ kg.

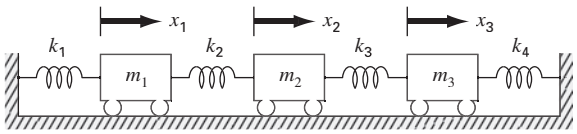


Figura P27.25

27.26 Considere o sistema massa-mola na Figura P27.26. As freqüências para as vibrações das massas podem ser determinadas por meio dos autovalores e aplicando $M\ddot{x} + kx = 0$, o que produz

$$\begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix} \begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \\ \ddot{x}_3 \end{Bmatrix} + \begin{bmatrix} 2k & -k & -k \\ -k & 2k & -k \\ -k & -k & 2k \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix}$$

Tentando $x = x_0 e^{i\omega t}$ como solução, obtemos a seguinte matriz:

$$\begin{bmatrix} 2k - m_1\omega^2 & -k & -k \\ -k & 2k - m_2\omega^2 & -k \\ -k & -k & 2k - m_3\omega^2 \end{bmatrix} \begin{Bmatrix} x_{01} \\ x_{02} \\ x_{03} \end{Bmatrix} e^{i\omega t} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix}$$

Utilize o comando `eig` do MATLAB para determinar os autovalores da matriz $k - m\omega^2$ anterior. A seguir, use esses autovalores para determinar as freqüências (ω). Faça $m_1 = m_2 = m_3 = 1$ kg e $k = 2$ N/m.

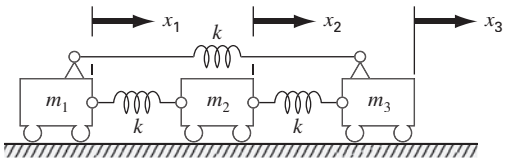


Figura P27.26

27.27 A seguinte EDO não-linear parasítica foi sugerida por Hornbeck (1975):

$$\frac{dy_1}{dt} = 5(y_1 - t^2)$$

Se a condição inicial for $y_1(0) = 0{,}08$, obtenha a solução de $t = 0$ a 5:

(a) Analiticamente.

(b) Empregando o método RK de quarta ordem com um passo constante de 0,03125.

(c) Usando a função ODE45 do MATLAB.

(d) Utilizando a função ODE23S do MATLAB.

(e) Usando a função ODE23TB do MATLAB.

Apresente seus resultados na forma gráfica.

27.28 Uma haste aquecida com uma fonte de calor uniforme pode ser modelada pela equação de Poisson:

$$\frac{d^2T}{dx^2} = -f(x)$$

Dada uma fonte de calor $f(x) = 25$ e as condições de contorno $T(x = 0) = 40$ e $T(x = 10) = 200$, determine a distribuição de temperatura com (**a**) o método *shooting* e (**b**) o método de diferenças finitas ($\Delta x = 2$).

27.29 Repita o Problema 27.28, mas com a seguinte fonte de calor: $f(x) = 0{,}12x^3 - 2{,}4x^2 + 12x$.

CAPÍTULO 28

Estudo de Casos: Equações Diferenciais Ordinárias

O propósito deste capítulo é resolver algumas equações diferenciais ordinárias usando os métodos numéricos apresentados na Parte Sete. As equações originam-se em aplicações práticas da engenharia. Muitas dessas aplicações resultam em equações diferenciais não-lineares que não podem ser resolvidas através de técnicas analíticas. Portanto, os métodos numéricos em geral são necessários. Logo, as técnicas para a solução numérica de equações diferenciais ordinárias são recursos fundamentais que caracterizam a boa prática da engenharia. Os problemas neste capítulo ilustram alguns prós e contras associados com os diversos métodos desenvolvidos na Parte Sete.

A Seção 28.1 vem de um contexto de problema da engenharia química. Ele ilustra como o comportamento transiente de reatores químicos pode ser simulado. Também mostra como a otimização pode ser usada para obter estimativa de parâmetros para EDOs.

As Seções 28.2 e 28.3, as quais vêm das engenharias civil e elétrica, respectivamente, tratam da solução de sistemas de equações. Em ambos os casos, é necessária alta acurácia e, como conseqüência, é utilizado um esquema RK de quarta ordem. Além disso, a aplicação em engenharia elétrica também trata da determinação de autovalores.

A Seção 28.4 usa diversas abordagens diferentes para investigar o comportamento de um pêndulo oscilante. Esse problema também utiliza duas equações simultâneas. Um aspecto importante desse problema é que ele ilustra como os métodos numéricos permitem que os efeitos não-lineares sejam incorporados facilmente na análise da engenharia.

28.1 USANDO EDOS PARA ANALISAR A RESPOSTA TRANSIENTE DE UM REATOR (ENGENHARIA QUÍMICA/BIOENGENHARIA)

Fundamentos. Na Seção 12.1, analisamos o estado estacionário de uma série de reatores. Além dos cálculos estacionários, poderíamos também nos interessar pela resposta transiente de um reator de mistura. Para fazer isso, temos de desenvolver uma expressão matemática para o termo de acumulação na Equação (12.1).

A acumulação representa a variação na massa contida no reator pela variação no tempo. Para um sistema com volume constante, ela pode ser formulada de modo simples como

$$\text{Acumulação} = V\frac{dc}{dt} \tag{28.1}$$

em que V é o volume e c, a concentração. Logo, uma formulação matemática para a acumulação é o volume vezes a derivada de c com relação a t.

Nessa aplicação, vamos incorporar o termo de acumulação no esquema geral de balanço de massa que desenvolvemos na Seção 12.1. Vamos, então, usá-lo para simular a dinâmica de um único reator e de um sistema de reatores. No último caso, vamos mostrar como os autovalores do sistema podem ser determinados e fornecer uma visão geral de sua dinâmica. Finalmente, vamos ilustrar como a otimização pode ser usada para obter estimativas dos parâmetros nos modelos de balanço de massa.

Solução. As Equações (28.1) e (12.1) podem ser usadas para representar o balanço de massa para um único reator, como mostrado na Figura 28.1:

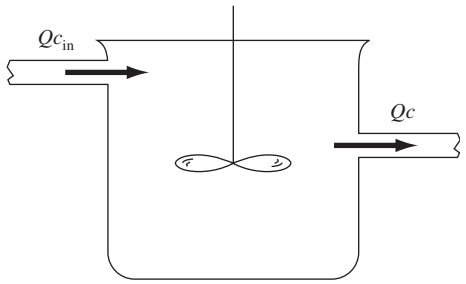


FIGURA 28.1
Um único reator de mistura, com escoamento de entrada e outro de saída.

$$V\frac{dc}{dt} = Qc_{in} - Qc \tag{28.2}$$

Acumulação = entradas − saídas

A Equação (28.2) pode ser utilizada para determinar soluções transientes ou dependentes do tempo para o reator. Por exemplo, se $c = c_0$ em $t = 0$, o cálculo pode ser empregado para resolver analiticamente a Equação (28.2) por

$$c = c_{in}\left(1 - e^{-(Q/V)t}\right) + c_0 e^{-(Q/V)t}$$

Se $c_{in} = 50$ mg/m^3, $Q = 5$ m^3/min, $V = 100$ m^3 e $c_0 = 10$ mg/m^3, a equação é

$$c = 50(1 - e^{-0{,}05t}) + 10e^{-0{,}05t}$$

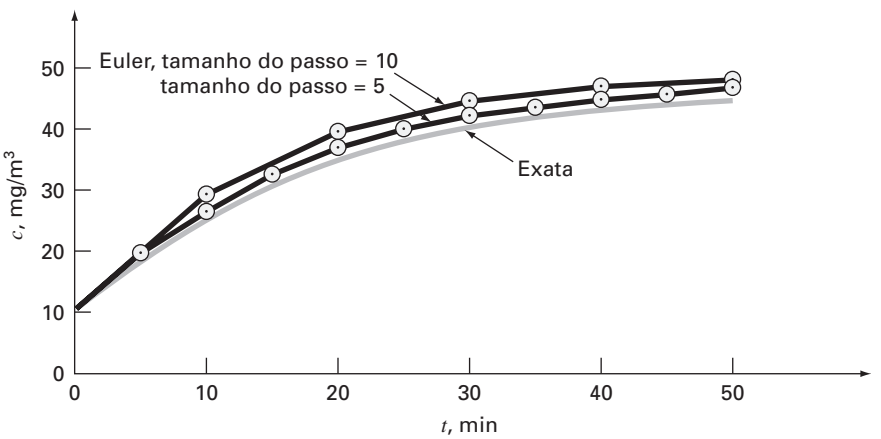
A Figura 28.2 mostra esta solução analítica exata.

O método de Euler fornece uma abordagem alternativa para a solução da Equação (28.2). A Figura 28.2 inclui duas soluções com diferentes tamanhos de passo. Conforme o tamanho do passo diminui, a solução numérica converge para a solução analítica. Nesse caso, o método numérico pode ser usado para verificar o resultado analítico.

Além de verificar os resultados de uma solução analítica, a solução numérica acrescentou valores nas situações em que as soluções analíticas são impossíveis ou são tão difíceis que não são práticas. Por exemplo, além de um único reator, os métodos numéricos têm utilidade para simular a dinâmica de um sistema de reatores. Como uma ilustração, podem ser escritas EDOs para os cinco reatores acoplados na Figura 12.3. O balanço de massa para o primeiro reator pode ser escrito como

$$V_1\frac{dc_1}{dt} = Q_{01}c_{01} + Q_{31}c_3 - Q_{12}c_1 - Q_{15}c_1$$

FIGURA 28.2
Gráfico das soluções analítica e numérica da Equação (28.2). As soluções numéricas foram obtidas pelo método de Euler usando diferentes tamanhos de passo.



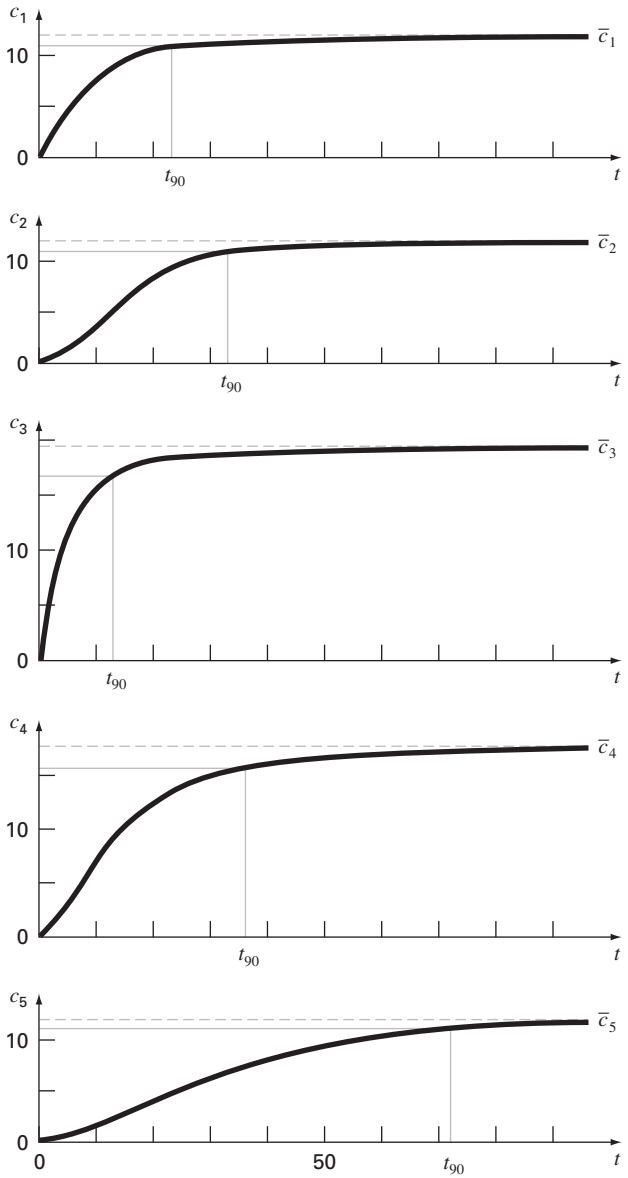


FIGURA 28.3
Gráficos da resposta transiente ou dinâmica da rede de reatores da Figura 12.3. Observe que todos os reatores eventualmente se aproximam de suas concentrações estacionárias calculadas previamente na Seção 12.1. Além disso, o tempo até o estado estacionário é parametrizado pela resposta no tempo de 90%, t_{90}.

ou, substituindo parâmetros (observe que $Q_{01}c_{01} = 50$ mg/min, $Q_{03}c_{03} = 160$ mg/min, $V_1 = 50$ m^3, $V_2 = 20$ m^3, $V_3 = 40$ m^3, $V_4 = 80$ m^3 e $V_5 = 100$ m^3),

$$\frac{dc_1}{dt} = -0{,}12c_1 + 0{,}02c_3 + 1$$

De forma análoga, os balanços para os outros reatores podem ser escritos como

$$\frac{dc_2}{dt} = 0{,}15c_1 - 0{,}15c_2$$

$$\frac{dc_3}{dt} = 0{,}025c_2 - 0{,}225c_3 + 4$$

$$\frac{dc_4}{dt} = 0{,}1c_3 - 0{,}1375c_4 + 0{,}025c_5$$

$$\frac{dc_5}{dt} = 0{,}03c_1 + 0{,}01c_2 - 0{,}04c_5$$

Suponha que, em $t = 0$, todas as concentrações nos reatores sejam nulas. Calcule como suas concentrações aumentarão na próxima hora.

As equações podem ser integradas com o método RK de quarta ordem para sistemas de equação e os resultados estão descritos na Figura 28.3. Observe que cada um dos reatores mostra uma resposta transiente diferente à introdução do composto químico. Essas respostas podem ser parametrizadas por uma resposta no tempo de 90%, t_{90}, que mede o tempo necessário para cada reator atingir 90% do seu nível final estacionário. Os tempos variam de cerca de 10 min para o reator 3 até cerca de 70 min para o reator 5. Os tempos de resposta dos reatores 4 e 5 apresentam interesse particular, pois as duas correntes de escoamento de saída ocorrem a partir desses tanques. Logo, um engenheiro químico projetando o sistema pode mudar os escoamentos ou os volumes dos reatores para acelerar a resposta desses tanques, enquanto ainda mantém as saídas desejadas. Os métodos numéricos do tipo descrito nesta parte do livro podem ser úteis nesses cálculos de projeto.

Uma percepção melhor das características da resposta do sistema pode ser desenvolvida pelo cálculo de seus autovalores. Primeiro, o sistema de EDOs pode ser escrito como um problema de autovalores como

$$\begin{bmatrix} 0{,}12-\lambda & 0 & -0{,}02 & 0 & 0 \\ -0{,}15 & 0{,}15-\lambda & 0 & 0 & 0 \\ 0 & -0{,}025 & 0{,}225-\lambda & 0 & 0 \\ 0 & 0 & -0{,}1 & 0{,}1375-\lambda & -0{,}025 \\ -0{,}03 & -0{,}01 & 0 & 0 & 0{,}04-\lambda \end{bmatrix} \begin{Bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{Bmatrix} = \{0\}$$

em que λ e $\{e\}$ são o autovalor e o autovetor, respectivamente.

Um pacote como o software MATLAB pode ser utilizado para gerar de forma muito conveniente os autovalores e os autovetores,

```
>> a=[0.12 0.0 -0.02 0.0 0.0;-.15 0.15 0.0 0.0 0.0;0.0
-0.025 0.225 0.0 0.0; 0.0 0.0 -.1 0.1375 -0.025;-0.03
-0.01 0.0 0.0 0.04];

>> [e,l]=eig(a)
e =
         0         0   -0.1228   -0.1059    0.2490
         0         0    0.2983    0.5784    0.8444
         0         0    0.5637    0.3041    0.1771
    1.0000    0.2484   -0.7604   -0.7493    0.3675
         0    0.9687    0.0041   -0.0190   -0.2419

l =
    0.1375         0         0         0         0
         0    0.0400         0         0         0
         0         0    0.2118         0         0
         0         0         0    0.1775         0
         0         0         0         0    0.1058
```

Os autovalores podem ser interpretados reconhecendo que a solução geral para um sistema de EDOs pode ser representado como a soma de exponenciais. Por exemplo, para o reator 1, a solução geral seria da forma

$$c_1 = c_{11}e^{-\lambda_1 t} + c_{12}e^{-\lambda_2 t} + c_{13}e^{-\lambda_3 t} + c_{14}e^{-\lambda_4 t} + c_{15}e^{-\lambda_5 t}$$

onde c_{ij} é a parte da condição inicial para o reator i que é associada com o autovalor j. Assim, como no caso presente todos os autovalores são positivos (e, portanto, negativos na função exponencial), a solução consiste em uma série de exponenciais decrescentes. A que tem o menor autovalor (em nosso caso, 0,04) será a mais lenta. Em alguns casos, o

engenheiro fazendo essa análise poderia ser capaz de relacionar esse autovalor aos parâmetros do sistema. Por exemplo, a razão do escoamento de saída do reator 5 por seu volume é $(Q_{55} + Q_{54})/V_5 = 4/100 = 0{,}04$. Tal informação pode então ser usada para modificar o desempenho dinâmico do sistema.

O tópico final que gostaríamos de revisar no presente contexto é a *estimativa de parâmetro*. Uma área em que isso ocorre frequentemente é na *cinética de reações*, isto é, na quantificação das taxas de reações químicas.

Um exemplo simples é descrito na Figura 28.4. Uma série de béqueres é montada contendo um composto químico que decai com o tempo. Em certos intervalos de tempo, a concentração em um dos béqueres é medida e anotada. Assim, o resultado é uma tabela de tempos e concentrações.

Um modelo que é comumente usado para descrever tais dados é

$$\frac{dc}{dt} = -kc^n \tag{28.3}$$

onde k é a taxa de reação e n é a ordem da reação. Os engenheiros químicos usam os dados de tempo-concentração do tipo descrito na Figura 28.4 para obter estimativas de k e n. Uma forma de fazer isso é escolher arbitrariamente valores para os parâmetros e então resolver a Equação (28.3) numericamente. Os valores previstos da concentração podem ser comparados com as concentrações medidas e pode ser feita uma avaliação do ajuste. Se o ajuste for considerado inadequado (por exemplo, pelo exame do gráfico ou de uma medida estatística como a soma do quadrado dos resíduos), os valores são então ajustados e o procedimento repetido até que seja atingido um ajuste decente.

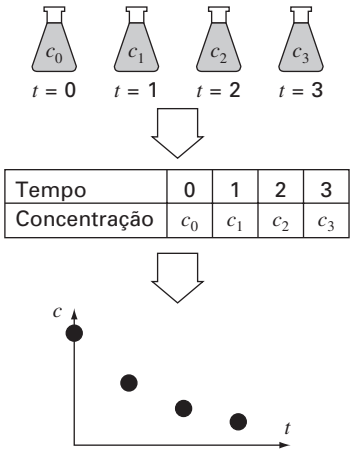
Os seguintes dados podem ser ajustados dessa forma:

t, d	0	1	3	5	10	15	20
c, mg/L	12	10,7	9	7,1	4,6	2,5	1,8

A solução desse problema é mostrada na Figura 28.5. A planilha do Excel foi usada para fazer os cálculos.

Tentativas iniciais para a taxa de reação e para a ordem foram inseridas nas células B3 e B4, respectivamente, e o tamanho do passo para os cálculos numéricos foi digitado na célula B5. Nesse caso, uma coluna para os instantes dos cálculos foi inserida na coluna A, começando em 0 (célula A7) e terminando em 20 (célula A27). Os coeficientes de k_1 a k_4 do método RK de quarta ordem são então calculados no bloco B7 a E27. Estes, então, são usados para determinar as concentrações previstas (os valores de c_p) na coluna F.

FIGURA 28.4
Uma experiência simples para coletar dados sobre a taxa de decaimento de um composto químico que decai com o tempo (reimpresso de Chapra, 1997).



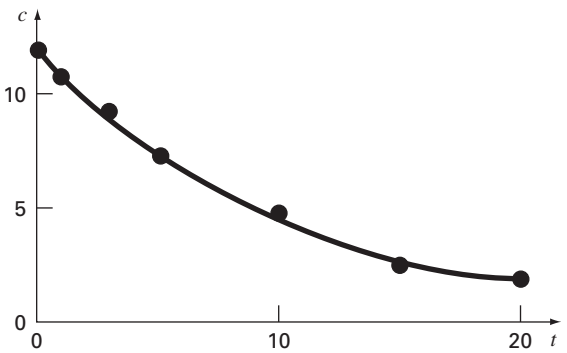
	A	B	C	D	E	F	G	H
1	Ajuste dos dados de taxa de reação com a							
2	abordagem integral/mínimos quadrados							
3	k	0,091528						
4	n	1,044425						
5	dt	1						
6	t	k1	k2	k3	k4	cp	cm	(cp-cm)^2
7	0	−1,22653	−1,16114	−1,16462	−1,10248	12	12	0
8	1	−1,10261	−1,04409	−1,04719	−0,99157	10,83658	10,7	0,018653
9	2	−0,99169	−0,93929	−0,94206	−0,89225	9,790448		
10	3	−0,89235	−0,84541	−0,84788	−0,80325	8,849344	9	0,022697
11	4	−0,80334	−0,76127	−0,76347	−0,72346	8,002317		
12	5	−0,72354	−0,68582	−0,68779	−0,65191	7,239604	7,1	0,019489
13	6	−0,65198	−0,61814	−0,61989	−0,5877	6,552494		
14	7	−0,58776	−0,55739	−0,55895	−0,53005	5,933207		
15	8	−0,53011	−0,50283	−0,50424	−0,47828	5,374791		
16	9	−0,47833	−0,45383	−0,45508	−0,43175	4,871037		
17	10	−0,4318	−0,40978	−0,4109	−0,38993	4,416389	4,6	0,033713
18	11	−0,38997	−0,37016	−0,37117	−0,35231	4,005877		
19	12	−0,35234	−0,33453	−0,33543	−0,31846	3,635053		
20	13	−0,31849	−0,30246	−0,30326	−0,28798	3,299934		
21	14	−0,28801	−0,27357	−0,2743	−0,26054	2,996949		
22	15	−0,26056	−0,24756	−0,24821	−0,23581	2,7229	2,5	0,049684
23	16	−0,23583	−0,22411	−0,22469	−0,21352	2,474917		
24	17	−0,21354	−0,20297	−0,20349	−0,19341	2,250426		
25	18	−0,19343	−0,18389	−0,18436	−0,17527	2,047117		
26	19	−0,17529	−0,16668	−0,16711	−0,1589	1,862914		
27	20	−0,15891	−0,15115	−0,15153	−0,14412	1,695953	1,8	0,010826
28								
29							SSR =	0,155062

FIGURA 28.5
A aplicação de uma planilha e de métodos numéricos para determinar a ordem e o coeficiente de reação a partir de dados de reação. Essa aplicação foi feita com a planilha Excel.

Os valores medidos (c_m) são inseridos na coluna G ao lado do valor previsto correspondente. Estes são então usados em conjunto com os valores previstos para calcular os resíduos quadráticos na coluna H. Esses valores são então somados na célula H29.

Neste ponto, o Excel Solver pode ser usado para determinar os melhores valores dos parâmetros. Uma vez que você tenha acessado o Solver, este lhe pedirá uma célula alvo ou solução (H29), perguntará se você quer maximizar ou minimizar a célula alvo (minimizar) e perguntará quais as células devem ser variadas (B3 a B4). Você então ativa o algoritmo [***s***(olve)] e os resultados serão como na Figura 28.5. Como mostrado, os valores nas células B3 a B4 ($k = 0{,}0915$ e $n = 1{,}044$) minimizam a soma dos quadrados dos resíduos (SSR = 0,155) entre os dados previstos e medidos. Um gráfico do ajuste juntamente com os dados é mostrado na Figura 28.6.

FIGURA 28.6
Gráfico do ajuste gerado com a abordagem integração/mínimos quadrados.



28.2 MODELOS PREDADOR-PRESA E CAOS (ENGENHARIA CIVIL/AMBIENTAL)

Fundamentos. Os engenheiros ambientais tratam diversos problemas envolvendo sistemas de equações diferenciais ordinárias não-lineares. Nesta seção, nos concentraremos em duas dessas aplicações. A primeira está relacionada com os assim chamados modelos predador-presa que são usados para estudar a variação cíclica de nutrientes e poluentes tóxicos na cadeia alimentar aquática e nos sistemas de tratamento biológicos. A segunda são equações que derivam da dinâmica dos fluidos e que são utilizadas para simular a atmosfera. Além de sua aplicação óbvia à previsão do tempo, tais equações também têm sido usadas para estudar a poluição do ar e a variação global do clima.

Os *modelos predador-presa* foram desenvolvidos independentemente no começo do século XX pelo matemático italiano Vito Volterra e pelo biólogo norte-americano Alfred J. Lotka. Essas equações são, em geral, denominadas *equações de Lotka-Volterra*. O exemplo mais simples é o seguinte par de EDOs:

$$\frac{dx}{dt} = ax - bxy \tag{28.4}$$

$$\frac{dy}{dt} = -cy + dxy \tag{28.5}$$

onde x e y são os números de presas e predadores, respectivamente, a é a taxa de crescimento da presa, c é a taxa de morte do predador e b e d são taxas caracterizando o efeito da interação predador-presa na morte da presa e no crescimento do predador, respectivamente. Os termos multiplicativos (isto é, aqueles envolvendo xy) são os que tornam tais equações não-lineares.

Um exemplo de modelo simples baseado na dinâmica dos fluidos atmosféricos são as equações de Lorenz, desenvolvidas pelo meteorologista norte-americano Edward Lorenz,

$$\frac{dx}{dt} = -\sigma x + \sigma y \tag{28.6}$$

$$\frac{dy}{dt} = rx - y - xz \tag{28.7}$$

$$\frac{dz}{dt} = -bz + xy \tag{28.8}$$

Lorenz desenvolveu essas equações para relacionar a intensidade de movimento do fluido atmosférico, x, as variações de temperatura y e z nas direções horizontal e vertical, respectivamente. Como no modelo predador-presa, vemos que a não-linearidade está localizada em termos multiplicativos simples (xz e xy).

Use métodos numéricos para obter soluções para essas equações. Faça gráficos dos seus resultados para visualizar como as variáveis dependentes variam no tempo. Além disso, faça um gráfico das variáveis dependentes uma em função da outra para ver se emergem quaisquer padrões interessantes.

Solução. Use os seguintes valores dos parâmetros para a simulação pedador-presa: $a = 1{,}2$, $b = 0{,}6$, $c = 0{,}8$ e $d = 0{,}3$. Use condições iniciais $x = 2$ e $y = 1$ e integre de $t = 0$ a 30. Utilizaremos o método RK de quarta ordem com dupla precisão para obter as soluções.

Os resultados usando um tamanho de passo de 0,1 estão mostrados na Figura 28.7. Observe que emerge um padrão cíclico. Assim, como a população de predadores é inicialmente pequena, a presa cresce exponencialmente. Em certo ponto, a presa se torna tão numerosa que a população de predadores começa a crescer. Eventualmente, o crescimento dos predadores faz que a presa diminua. Essa diminuição, por sua vez, leva a uma diminuição dos predadores. Eventualmente, o processo se repete. Observe que, como esperado, o pico dos predadores ocorre um pouco depois do da presa. Além disso, note que o processo tem um período fixo, isto é, ele se repete em um mesmo intervalo de tempo.

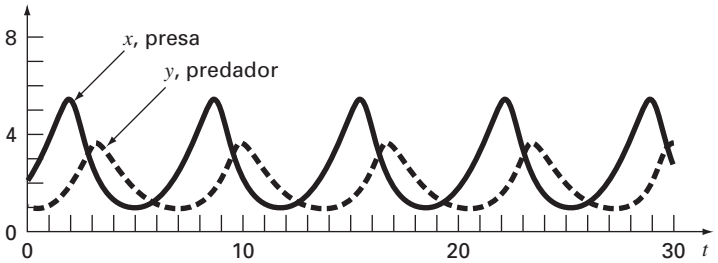


FIGURA 28.7
Representação no domínio do tempo do número de presas e de predadores no modelo de Lotka-Volterra.

Agora, se os parâmetros usados para simular a Figura 28.7 forem mudados, embora o padrão geral permaneça o mesmo, a intensidade dos picos, atrasos e períodos iriam mudar. Assim, existe um número infinito de ciclos que poderiam ocorrer.

Uma representação em espaço de estados é útil para discernir a estrutura subjacente do modelo. Em vez de traçar o gráfico de x e y em função de t, podemos traçar x *versus* y. Como tal gráfico ilustra a maneira pela qual as variáveis de estado x e y interagem, e este é conhecido como *representação no espaço de estados*.

A Figura 28.8 mostra a representação no espaço de estados para o caso que estamos estudando. Assim, a interação entre o predador e a presa definiu uma órbita fechada percorrida no sentido anti-horário. Observe que existe um ponto crítico ou de repouso no centro da órbita. A localização exata desse ponto pode ser determinada igualando-se as Equações (28.4) e (28.5) ao estado estacionário ($dy/dt = dx/dt = 0$) e encontrando $(x, y) = (0, 0)$ e $(c/d, a/b)$. O primeiro é o resultado trivial, em que, se começarmos sem nenhum predador e nenhuma presa, nada acontecerá. O último é a saída mais interessante, em que, se as condições inicias forem escolhidas como $x = c/d$ e $y = a/b$, as derivadas serão nulas e as populações permanecerão constantes.

Agora, vamos usar a mesma abordagem para investigar as trajetórias das equações de Lorenz com os seguintes valores de parâmetro: $\sigma = 10$, $b = 2{,}666667$ e $r = 28$. Use condições iniciais $x = y = z = 5$ e integre de $t = 0$ a 20. Novamente, vamos usar o método RK de quarta ordem de dupla precisão para obter soluções.

Os resultados mostrados na Figura 28.9 são bastante diferentes das equações de Lotka-Volterra. A variável x parece estar seguindo um padrão quase aleatório de oscilações, variando de valores negativos para valores positivos. Entretanto, mesmo que os padrões pareçam aleatórios, a freqüência de oscilação e as amplitudes parecem bastante consistentes.

Outra característica interessante pode ser ilustrada variando ligeiramente a condição inicial para x (de 5 para 5,001). Os resultados estão superpostos como a linha tracejada na Figura 28.9. Embora as soluções permaneçam juntas por certo tempo, depois de cerca de $t = 12{,}5$, elas divergem significativamente. Dessa forma, podemos ver que as equações de Lorenz são bastante sensíveis às condições iniciais. No estudo inicial, isso levou Lorenz à conclusão de que previsão de tempo de longo alcance poderia ser impossível!

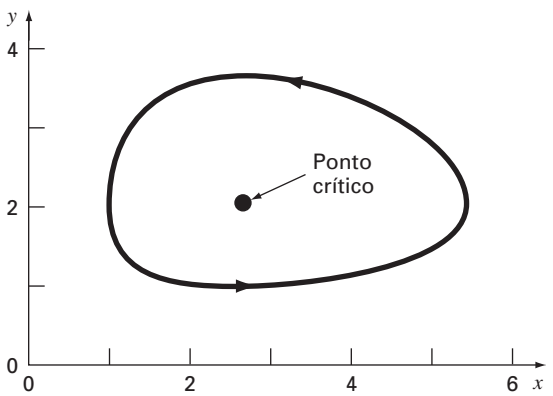


FIGURA 28.8
Representação no espaço de estados do modelo de Lotka-Volterra.

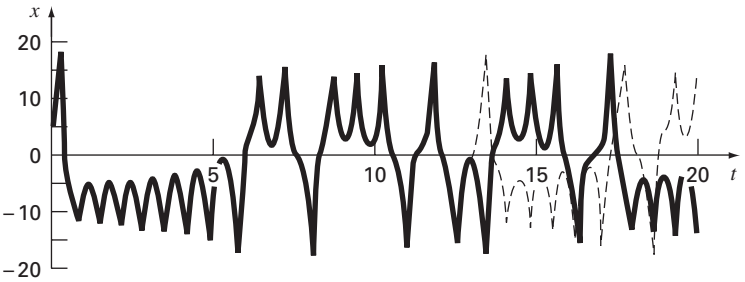


FIGURA 28.9
Representação no domínio do tempo de x em função de t para as equações de Lorenz. A linha sólida é para as condições iniciais (5, 5, 5). A linha pontilhada é para o caso da condição inicial para x ser ligeiramente perturbada [(5,001, 5, 5)].

Finalmente, vamos analisar os gráficos no espaço de estados. Como se trata de três variáveis independentes, estamos limitados a projeções. A Figura 28.10 mostra projeções nos planos *xy* e *xz*. Observe como é aparente uma estrutura quando observada da perspectiva do espaço de estados. A solução forma órbitas em torno do que parecem ser pontos críticos. Esses pontos são chamados *atratores estranhos* no jargão dos matemáticos que estudam tais sistemas não-lineares.

Soluções do tipo das que exploramos para as equações de Lorenz são conhecidas como soluções *caóticas*. O estudo do caos e de sistemas não-lineares atualmente representa uma área excitante de análise que tem implicações tanto na matemática quanto na ciência e na engenharia.

De uma perspectiva numérica, o ponto principal é a sensibilidade de tais soluções às condições iniciais. Assim, algoritmos numéricos diferentes, precisão do computador e tamanho de passos de integração podem todos ter um impacto grande na solução numérica resultante.

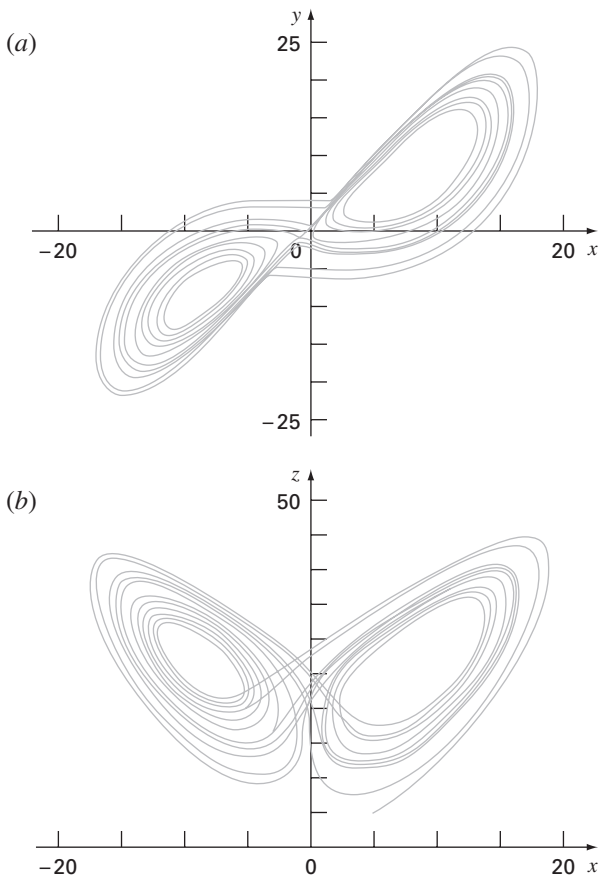


FIGURA 28.10
Representação no espaço de estados das equações de Lorenz. (*a*) Projeção *xy* e (*b*) projeção *xz*.

28.3 SIMULANDO CORRENTES TRANSIENTES PARA UM CIRCUITO ELÉTRICO (ENGENHARIA ELÉTRICA)

Fundamentos. Os circuitos elétricos nos quais a corrente varia no tempo, em vez de serem constantes, são comuns. Uma corrente transiente é estabelecida no laço à direita do circuito mostrado na Figura 28.11 quando a chave é fechada repentinamente.

As equações que descrevem o comportamento transiente do circuito na Figura 28.11 são baseadas na lei Kirchhoff, que afirma que a soma algébrica das quedas de voltagem em um laço fechado é nula (lembre-se da Seção 8.3). Logo,

$$L\frac{di}{dt} + Ri + \frac{q}{C} - E(t) = 0 \tag{28.9}$$

em que $L(di/dt)$ é a queda de voltagem no indutor, L é a indutância (H), R é a resistência (Ω), q é a carga no capacitor (C), C é a capacitância (F), $E(t)$ é a fonte de voltagem dependente do tempo (V) e

$$i = \frac{dq}{dt} \tag{28.10}$$

As Equações (28.9) e (28.10) são um par de equações diferenciais lineares de primeira ordem que podem ser resolvidas analiticamente. Por exemplo, se $E(t) = E_0 \operatorname{sen} \omega t$ e $R = 0$,

$$q(t) = \frac{-E_0}{L(p^2 - \omega^2)}\frac{\omega}{p}\operatorname{sen} pt + \frac{E_0}{L(p^2 - \omega^2)}\operatorname{sen} \omega t \tag{28.11}$$

onde $p = 1/\sqrt{LC}$. Os valores de q e dq/dt são nulos para $t = 0$. Use uma abordagem numérica para resolver as Equações (28.9) e (28.10) e compare os resultados com a Equação (28.11).

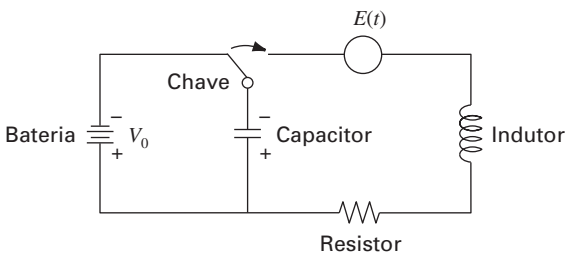


FIGURA 28.11
Um circuito elétrico no qual a corrente varia com o tempo.

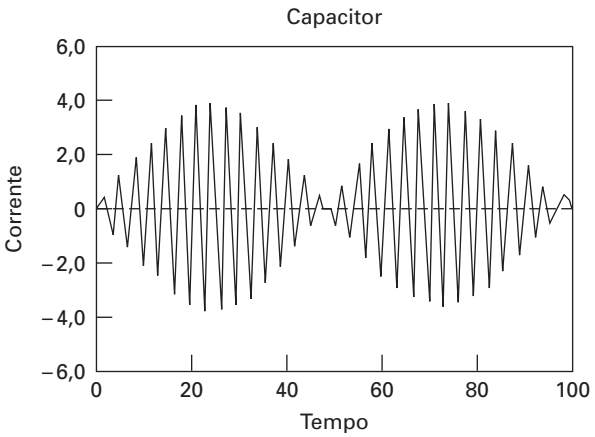


FIGURA 28.12
Tela do computador mostrando um gráfico da função representada pela Equação (28.11).

Solução. Esse problema envolve um intervalo de integração bastante grande e exige o uso de um esquema altamente acurado para resolver a equação diferencial se forem esperados bons resultados. Vamos supor que $L = 1$ H, $E_0 = 1$ V, $C = 0{,}25$ C, e $\omega^2 = 3{,}5$ s^2. Isso resulta em $p = 2$ e a Equação (28.11) se torna

$$q(t) = -1{,}8708 \text{ sen } (2t) + 2 \text{ sen } (1{,}8708t)$$

para a solução analítica. O gráfico dessa função está ilustrado na Figura 28.12. A natureza de variação rápida da função coloca uma exigência severa em qualquer procedimento numérico para encontrar $q(t)$. Além disso, como a função exibe uma natureza periódica que varia lentamente, bem como uma componente que varia rapidamente, são necessários intervalos longos de integração para retratar a solução. Assim, esperamos que um método de alta ordem seja preferido para esse problema.

Entretanto, podemos tentar tanto os métodos de Euler e de *RK* de quarta ordem e comparar os resultados. Usando um tamanho de passo de 0,1 s, obtemos um valor de q em $t = 10$ s de $-6{,}638$ com o método de Euler e um valor de $-1{,}9897$ com o método RK de quarta ordem. Podemos comparar esses resultados com uma solução exata de $-1{,}996$C.

A Figura 28.13 mostra os resultados da integração de Euler a cada 1,0 s comparados com a solução exata. Observe que apenas um em cada dez pontos de saída está no gráfico. Parece que o erro global aumenta conforme t aumenta. Esse comportamento divergente se intensifica se t tende a infinito.

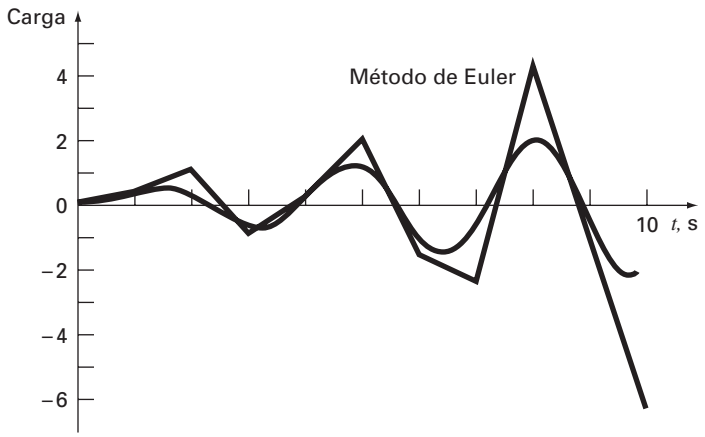
Além disso, para simular diretamente a resposta transiente de uma rede, os métodos numéricos também podem ser usados para determinar seus autovalores. Por exemplo, a Figura 28.14 exibe uma rede *RC* para a qual a lei da voltagem de Kirchhoff pode ser utilizada para deduzir o seguinte sistema de EDOs:

$$-L_1\frac{di_1}{dt} - \frac{1}{C_1}\int_{-\infty}^{t}(i_1 - i_2)\,dt = 0$$

$$-L_2\frac{di_2}{dt} - \frac{1}{C_2}\int_{-\infty}^{t}(i_2 - i_3)\,dt + \frac{1}{C_1}\int_{-\infty}^{t}(i_1 - i_2)\,dt = 0$$

$$-L_3\frac{di_3}{dt} - \frac{1}{C_3}\int_{-\infty}^{t} i_3\,dt + \frac{1}{C_2}\int_{-\infty}^{t}(i_2 - i_3)\,dt = 0$$

FIGURA 28.13
Resultados da integração de Euler *versus* a solução exata. Observe que apenas um em cada dez pontos de saída está no gráfico.



Observe porque representamos a queda de voltagem no capacitor

$$V_C = \frac{1}{C}\int_{-\infty}^{t} i\,dt$$

Esta é uma expressão alternativa equivalente da relação usada na Equação (28.9) e introduzida na Seção 8.3.

O sistema de EDOs pode ser derivado e reorganizado para fornecer

$$L_1\frac{d^2i_1}{dt^2} + \frac{1}{C_1}(i_1 - i_2) = 0$$

$$L_2\frac{d^2i_2}{dt^2} + \frac{1}{C_2}(i_2 - i_3) - \frac{1}{C_1}(i_1 - i_2) = 0$$

$$L_3\frac{d^2i_3}{dt^2} + \frac{1}{C_3}i_3 - \frac{1}{C_2}(i_2 - i_3) = 0$$

A comparação desse sistema com o sistema da Equação (27.5) indica uma analogia entre o sistema massa-mola e um circuito *LC*. Como feito na Equação (27.5), podemos supor que a solução tenha a forma

$$i_j = A_j \text{sen}\,(\omega t)$$

Essa solução juntamente com sua segunda derivada pode ser substituída nas EDOs simultâneas. Depois de simplificações, o resultado é

$$\begin{aligned}
\left(\frac{1}{C_1} - L_1\omega^2\right)A_1 \qquad & -\frac{1}{C_2}A_2 & & = 0\\
-\frac{1}{C_1}A_1 \qquad & +\left(\frac{1}{C_1} + \frac{1}{C_2} - L_2\omega^2\right)A_2 \qquad & -\frac{1}{C_2}A_3 \qquad & = 0\\
& -\frac{1}{C_2}A_2 \qquad & +\left(\frac{1}{C_2} + \frac{1}{C_3} - L_3\omega^2\right)A_3 & = 0
\end{aligned}$$

Assim, formulamos um problema de autovalores. Mais simplificações ocorrem no caso em que os *C*'s e os *L*'s são constantes. Para essa situação, o sistema pode ser expresso na forma matricial como

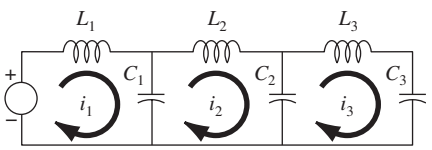
$$\begin{bmatrix} 1-\lambda & -1 & 0\\ -1 & 2-\lambda & -1\\ 0 & -1 & 2-\lambda \end{bmatrix}\begin{Bmatrix} A_1\\ A_2\\ A_3 \end{Bmatrix} = \{0\} \tag{28.12}$$

onde

$$\lambda = LC\omega^2 \tag{28.13}$$

Podem ser utilizados métodos numéricos para determinar os autovalores e autovetores. O MATLAB é particularmente conveniente nesse caso. A seguinte seção MATLAB foi desenvolvida para fazer isso:

FIGURA 28.14
Uma rede *LC*.



```
>>a=[1 -1 0; -1 2 -1; 0 -1 2]

a =

     1    -1     0
    -1     2    -1
     0    -1     2
>>[v,d]=eig(a)

v =

    0.7370    0.5910    0.3280
    0.5910   -0.3280   -0.7370
    0.3280   -0.7370    0.5910

d =

    0.1981         0         0
         0    1.5550         0
         0         0    3.2470
```

A matriz `v` consiste nos três autovetores do sistema (dispostos como colunas) e `d` é a matriz com os autovalores correspondentes na diagonal. Assim, o pacote calcula que os autovalores são $\lambda = 0{,}1981$, 1,555 e 3,247. Esses valores, por sua vez, podem ser substituídos na Equação (28.13) para determinar as freqüências naturais do sistema

$$\omega = \begin{Bmatrix} \dfrac{0{,}4450}{\sqrt{LC}} \\ \dfrac{1{,}2470}{\sqrt{LC}} \\ \dfrac{1{,}8019}{\sqrt{LC}} \end{Bmatrix}$$

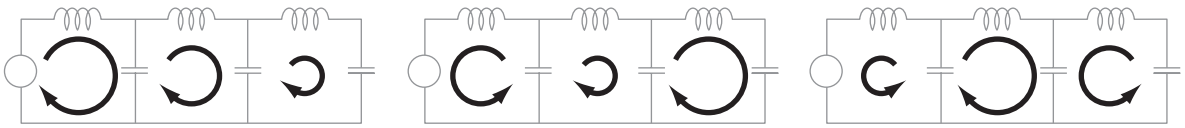
Além de fornecer as freqüências naturais, os autovalores podem ser substituídos na Equação (28.12) para fornecerem uma percepção maior do comportamento físico do sistema. Por exemplo, a substituição de $\lambda = 0{,}1981$ fornece

$$\begin{bmatrix} 0{,}8019 & -1 & 0 \\ -1 & 1{,}8019 & -1 \\ 0 & -1 & 1{,}8019 \end{bmatrix} \begin{Bmatrix} i_1 \\ i_2 \\ i_3 \end{Bmatrix} = \{0\}$$

Embora esse sistema não tenha uma única solução, ele será satisfeito se as correntes obedecerem a razões fixas, como em

$$0{,}8019i_1 = i_2 = 1{,}8019i_3 \tag{28.14}$$

FIGURA 25.15
Uma representação visual dos modos naturais de oscilação da rede *LC* da Figura 28.14. Observe que os diâmetros das setas circulares são proporcionais à intensidade da corrente em cada laço.



(*a*) $\omega = \dfrac{0{,}4451}{\sqrt{LC}}$ (*b*) $\omega = \dfrac{1{,}2470}{\sqrt{LC}}$ (*c*) $\omega = \dfrac{1{,}8019}{\sqrt{LC}}$

Dessa forma, como descrito pela Figura 28.15*a*, elas oscilam na mesma direção com intensidades diferentes. Observe que, se supusermos $i_1 = 0{,}737$, podemos usar a Equação (28.14) para calcular as outras correntes com o resultado

$$\{i\} = \begin{Bmatrix} 0{,}737 \\ 0{,}591 \\ 0{,}328 \end{Bmatrix}$$

que é a primeira coluna da matriz v, calculada com o MATLAB.

De forma análoga, o segundo autovalor $\lambda = 1{,}555$ pode ser substituído, fornecendo o resultado

$$-1{,}8018i_1 = i_2 = 2{,}247i_3$$

Como descrito na Figura 28.15*b*, o primeiro laço oscila no sentido oposto do segundo e do terceiro. Finalmente, o terceiro modo pode ser determinado como

$$-0{,}445i_1 = i_2 = -0{,}8718i_3$$

Conseqüentemente, como na Figura 28.15*c*, os primeiro e terceiro laços oscilam no sentido oposto do segundo.

28.4 O PÊNDULO OSCILANTE (ENGENHARIA MECÂNICA/AEROESPACIAL)

Fundamentos. Os engenheiros mecânicos (assim como todos os outros engenheiros) são freqüentemente confrontados por problemas relacionados com o movimento periódico de corpos livres. A abordagem da engenharia para tais problemas acaba exigindo que a posição e a velocidade do corpo sejam conhecidas como função do tempo. Essas funções do tempo invariavelmente são soluções de equações diferenciais ordinárias. As equações diferenciais são geralmente baseadas nas leis de movimento de Newton.

Como um exemplo, considere o pêndulo simples mostrado anteriormente na Figura PT7.1. A partícula de peso W está presa a uma haste sem peso de comprimento l. As únicas forças agindo na partícula são seu peso e a tensão R na haste. A posição da partícula em qualquer instante de tempo é especificada completamente em termos do ângulo θ e l.

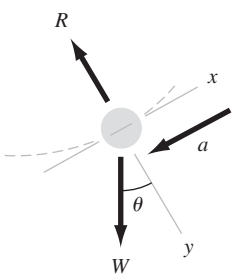
O diagrama de corpo livre na Figura 28.16 mostra as forças na partícula e a aceleração. É conveniente aplicar as leis de movimento de Newton na direção x tangente à trajetória da partícula:

$$\Sigma F = -W \operatorname{sen} \theta = \frac{W}{g} a$$

onde g é a constante gravitacional (32,2 pés/s^2) e a é a aceleração na direção x. A aceleração angular da partícula (α) se torna

$$\alpha = \frac{a}{l}$$

FIGURA 28.16
Um diagrama de corpo livre do pêndulo oscilante mostrando as forças na partícula e a aceleração.



Portanto, em coordenadas polares ($\alpha = d^2\theta/dt^2$),

$$-W \operatorname{sen} \theta = \frac{Wl}{g}\alpha = \frac{Wl}{g}\frac{d^2\theta}{dt^2}$$

ou

$$\frac{d^2\theta}{dt^2} + \frac{g}{l} \operatorname{sen} \theta = 0 \tag{28.15}$$

Essa equação aparentemente simples é uma equação diferencial não-linear de segunda ordem. Em geral, tais equações são difíceis ou impossíveis de resolver analiticamente. Você tem duas escolhas relativas de como prosseguir. Primeiro, a equação diferencial poderia ser reduzida a uma forma que pudesse ser resolvida analiticamente (lembre-se da

Seção PT7.1.1), ou, segundo, uma técnica de aproximação numérica poderia ser usada para resolver diretamente a equação diferencial. Examinaremos ambas as alternativas neste exemplo.

Solução. Prosseguindo com a primeira abordagem, observamos que uma expansão em série para sen θ é dada por

$$\operatorname{sen}\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots \tag{28.16}$$

Para pequenos deslocamentos angulares, sen θ é aproximadamente igual a θ quando expresso em radianos. Portanto, para pequenos deslocamentos, a Equação (28.15) se torna

$$\frac{d^2\theta}{dt^2} + \frac{g}{l}\theta = 0 \tag{28.17}$$

que é uma equação diferencial linear de segunda ordem. Essa aproximação é muito importante porque a Equação (28.17) é muito fácil de resolver analiticamente. A solução, baseada na teoria das equações diferenciais, é dada por

$$\theta(t) = \theta_0 \cos\sqrt{\frac{g}{l}}t \tag{28.18}$$

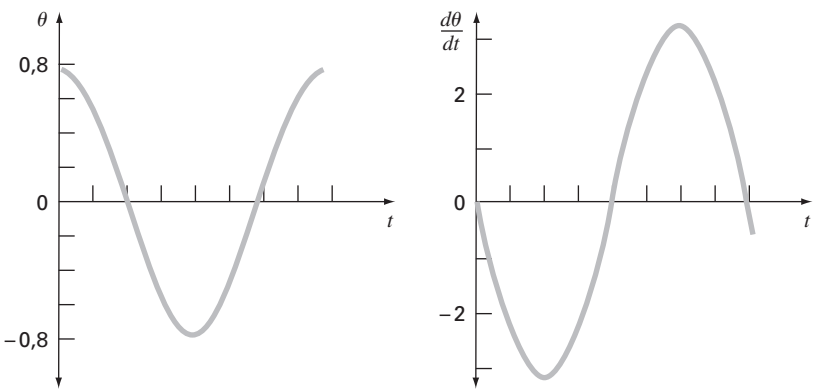
onde θ_0 é o deslocamento em $t = 0$ e na qual é suposto que a velocidade ($v = d\theta/dt$) seja zero em $t = 0$. O tempo necessário para que o pêndulo complete um ciclo de oscilação é chamado período e é dado por

$$T = 2\pi\sqrt{\frac{l}{g}} \tag{28.19}$$

A Figura 28.17 apresenta um gráfico do deslocamento θ e da velocidade $d\theta/dt$ como uma função do tempo, como calculado a partir da Equação (28.18) com $\theta_0 = \pi/4$ e $l = 2$ pés. O período, calculado a partir da Equação (28.19), é 1,5659 s.

Os cálculos anteriores são essencialmente uma solução completa do movimento do pêndulo. Entretanto, você deve considerar a acurácia de seus resultados por causa da hipótese inerente na Equação (28.17). Para avaliar a acurácia, é necessário obter uma solução numérica para a Equação (28.15), que é uma representação física mais completa do movimento. Qualquer um dos métodos discutidos nos Capítulos 25 e 26 poderia ser

FIGURA 28.17
Gráfico do deslocamento θ e da velocidade $d\theta/dt$ como uma função do tempo t, como calculado a partir da Equação (28.18). θ_0 é $\pi/4$ e o comprimento é 2 pés.



usado com esse propósito — por exemplo, os métodos de Euler e RK de quarta ordem. A Equação (28.15) deve ser transformada em duas equações de primeira ordem para ser compatível com os métodos descritos. Isso pode ser obtido como segue. A velocidade v é definida por

$$\frac{d\theta}{dt} = v \tag{28.20}$$

e, portanto, a Equação (28.15) pode ser expressa por

$$\frac{dv}{dt} = -\frac{g}{l}\text{sen}\,\theta \tag{28.21}$$

As Equações (28.20) e (28.21) são um sistema acoplado de duas equações diferenciais ordinárias. As soluções numéricas pelo método de Euler e pelo método RK de quarta ordem dão os resultados mostrados na Tabela 28.1, que compara a solução analítica para a equação linear de movimento [Equação (28.18)] na coluna (*a*) com as soluções numéricas nas colunas (*b*), (*c*) e (*d*).

Os métodos de Euler e RK de quarta ordem fornecem resultados diferentes e ambos são diferentes da solução analítica, embora o método RK de quarta ordem para o caso não-linear esteja mais próximo da solução analítica que o método de Euler. Para avaliar adequadamente a diferença entre os modelos linear e não-linear, é importante determinar a acurácia dos resultados numéricos. Pode-se conseguir isso de três maneiras. Primeira, a solução numérica de Euler é facilmente reconhecida como inadequada, pois ela supera a condição inicial em $t = 0{,}8$ s. Isso claramente viola a conservação da energia. Segunda, as colunas (*c*) e (*d*) na Tabela 28.1 mostram a solução pelo método RK de quarta ordem para tamanhos de passo de 0,05 e 0,01. Como elas diferem na quarta casa decimal, é razoável supor que a solução com tamanho de passo de 0,01 também apresente esta ordem de acurácia. Terceira, para o caso do tamanho do passo de 0,01, θ atinge um valor máximo local de 0,785385 em $t = 1{,}63$ s (não mostrado na Tabela 28.1). Isso indica que o pêndulo volta à sua posição inicial com quatro casas de acurácia com um período de 1,63 s. Essas considerações nos permitem supor com segurança que a diferença entre as colunas (*a*) e (*d*) na Tabela 28.1 representa verdadeiramente a diferença entre o modelo linear e não-linear.

Outra forma de caracterizar a diferença entre os modelos linear e não-linear é com base no período. A Tabela 28.2 mostra o período de oscilação como calculado pelo modelo linear e pelo modelo não-linear para três deslocamentos iniciais diferentes. Podemos ver que os períodos calculados ficam próximos quando θ é pequeno, pois θ é uma boa aproximação para sen θ na Equação (28.16). Essa aproximação se deteriora conforme θ se torna grande.

Essas análises são típicas dos casos que vamos encontrar rotineiramente como en-

TABELA 28.1 Comparação de uma solução analítica linear do problema do pêndulo oscilante com três soluções numéricas não-lineares.

		Soluções Numéricas Não-lineares		
Tempo, s	**Solução Analítica Linear (*a*)**	**Euler (*h* = 0,05) (*b*)**	**RK de Quarta Ordem (*h* = 0,05) (*c*)**	**RK de Quarta Ordem (*h* = 0,01) (*d*)**
0,0	0,785398	0,785398	0,785398	0,785398
0,2	0,545784	0,615453	0,566582	0,566579
0,4	−0,026852	0,050228	0,021895	0,021882
0,6	−0,583104	−0,639652	−0,535802	−0,535820
0,8	−0,783562	−1,050679	−0,784236	−0,784242
1,0	−0,505912	−0,940622	−0,595598	−0,595583
1,2	0,080431	−0,299819	−0,065611	−0,065575
1,4	0,617698	0,621700	0,503352	0,503392
1,6	0,778062	1,316795	0,780762	0,780777

TABELA 28.2 Comparação do período de um corpo oscilante calculado a partir dos modelos linear e não-linear.

Deslocamento Inicial, θ_0	Período, s	
	Modelo Linear ($T = 2\pi\sqrt{l/g}$)	**Modelo Não-linear [Solução Numérica da Eq. (28.15)]**
$\pi/16$	1,5659	1,57
$\pi/4$	1,5659	1,63
$\pi/2$	1,5659	1,85

genheiros. A utilidade das técnicas numéricas se torna particularmente significativa nos problemas não-lineares e, em muitos casos, os problemas do mundo real são não-lineares.

PROBLEMAS

Engenharia Química/Bioengenharia

28.1 Faça o primeiro cálculo da Seção 28.1, mas para o caso em que $h = 10$. Use os métodos de Heun (sem iteração) e o RK de quarta ordem para obter soluções.

28.2 Faça o segundo cálculo da Seção 28.1, mas para o sistema descrito no Problema 12.4.

28.3 Um balanço de massa para um composto químico em um reator de mistura pode ser escrito como

$$V\frac{dc}{dt} = F - Qc - kVc^2$$

onde V é o volume (12 m^3), c é a concentração (g/m^3), F é a taxa de alimentação (175 g/min), Q é a vazão (1 m^3/min) e k é uma taxa de reação de segunda ordem (0,15 m^3/g/min). Se $c(0) = 0$, resolva a EDO até que a concentração atinja um nível estável. Use o método do ponto médio ($h = 0{,}5$) e faça o gráfico de seus resultados.

Questão desafio: se ignorarmos o fato de que as concentrações devem ser positivas, encontre um intervalo de condições iniciais tal que você obteria uma trajetória muito diferente da que foi obtida com $c(0) = 0$. Relacione seus resultados com as soluções estacionárias.

28.4 Se $c_{\text{in}} = c_b(1 - e^{-0{,}12t})$, calcule a concentração no escoamento de saída de uma substância conservativa (sem reações) para um único reator de mistura como uma função do tempo. Use o método de Heun (sem iteração) para fazer os cálculos. Use valores de $c_b = 40$ mg/m^3, $Q = 6$ m^3/min, $V = 100$ m^3 e $c_0 = 20$ mg/m^3. Faça os cálculos de $t = 0$ a 100 min usando $h = 2$. Trace seus resultados com a concentração no escoamento de entrada em função do tempo.

28.5 A água do mar com uma concentração de 8000 g/m^3 é bombeada para um tanque bem misturado a uma vazão de 0,6 m^3/hr. Por causa de um defeito no trabalho do projeto, a água está evaporando do tanque a uma vazão de 0,025 m^3/hr. A solução salina deixa o tanque a uma vazão de 0,6 m^3/hr.

(a) Se o tanque continha originalmente 1 m^3 da solução de entrada, quanto tempo depois de ligar a bomba de saída o tanque secará?

(b) Use métodos numéricos para determinar a concentração de sal no tanque como uma função do tempo.

28.6 Uma esfera de gelo que tem 6 cm de diâmetro é removida de um congelador a 0° C e é colocada em uma tela de rede à temperatura ambiente $T_a = 20$°C. Qual será o diâmetro da esfera de gelo como uma função do tempo fora do congelador (supondo que toda a água que derreteu caia imediatamente pela tela)? O coeficiente de transferência de calor h para uma esfera em um quarto sem vento é cerca de 3 W/(m$^2 \cdot$ K). O fluxo de calor da esfera de gelo para o ar é dado por

$$\text{Fluxo} = \frac{q}{A} = h(T_a - T)$$

onde q é o calor e A é a área da superfície da esfera. Use um método numérico para fazer seus cálculos. Observe que o calor latente de fusão é 333 kJ/kg e que a densidade do gelo é aproximadamente 0,917 kg/m^3.

28.7 As seguintes equações definem a concentração de três reagentes:

$$\frac{dc_a}{dt} = -10c_ac_c + c_b$$
$$\frac{dc_b}{dt} = 10c_ac_c - c_b$$
$$\frac{dc_c}{dt} = -10c_ac_c + c_b - 2c_c$$

Se as condições iniciais forem $c_a = 50$, $c_b = 0$ e $c_c = 40$, encontre as concentrações para instantes entre 0 e 3 s.

28.8 O composto A difunde por meio de um tubo de 4 cm de comprimento e reage conforme ele difunde. A equação que rege a difusão com reação é

$$D\frac{d^2A}{dx^2} - kA = 0$$

Em uma extremidade do tubo, existe uma grande fonte de A a uma concentração de 0,1 M. Na outra extremidade do tubo, existe um material absorvente que absorve rapidamente qualquer A, tornando a concentração 0 M. Se $D = 1{,}5 \times 10^{-6}$ cm^2/s e $k = 5 \times 10^{-6}$ s^{-1}, qual é a concentração de A como função da distância no tubo?

28.9 Na investigação de homicídio ou de morte acidental, em geral, é importante determinar o momento da morte. De observações experimentais, sabe-se que a temperatura da superfície de um objeto varia a uma taxa proporcional à diferença entre a temperatura do objeto e a do meio que o cerca ou a temperatura ambiente. Isso é conhecido como a lei de resfriamento de Newton. Assim, se $T(t)$ for a temperatura do objeto no instante t e T_a for a temperatura constante do ambiente:

$$\frac{dT}{dt} = -K(T - T_a)$$

onde $K > 0$ é uma constante de proporcionalidade. Suponha que, no instante $t = 0$, um corpo seja descoberto e que sua temperatura medida seja T_0. Supomos que no instante da morte a temperatura do corpo, T_d, tinha o valor normal de 37° C. Suponha que a temperatura do corpo, quando foi descoberto, era de 29,5° C e que duas horas depois era de 23,5° C. A temperatura ambiente é 20° C.

(a) Determine K e o momento da morte.

(b) Resolva a EDO numericamente e trace seus resultados.

28.10 A reação $A \rightarrow B$ ocorre em dois reatores em série. Os reatores estão bem misturados, mas não estão no estado estacionário. O balanço de massa para cada reator de tanque agitado é mostrado a seguir:

$$\frac{dCA_1}{dt} = \frac{1}{\tau}(CA_0 - CA_1) - kCA_1$$

$$\frac{dCB_1}{dt} = -\frac{1}{\tau}CB_1 + kCA_1$$

$$\frac{dCA_2}{dt} = \frac{1}{\tau}(CA_1 - CA_2) - kCA_2$$

$$\frac{dCB_2}{dt} = \frac{1}{\tau}(CB_1 - CB_2) + kCA_2$$

onde CA_0 é a concentração de A na entrada do primeiro reator, CA_1 é a concentração de A na saída do primeiro reator (e na entrada do segundo), CA_2 é a concentração de A na saída do segundo reator, CB_1 é a concentração de B na saída do primeiro reator (e na entrada do segundo), CB_2 é a concentração de B no segundo reator, τ é o tempo de residência para cada reator e k é a taxa constante para a reação de A produzir B. Se CA_0 for igual a 20, encontre as concentrações de A e B em ambos os reatores durante os primeiros 10 minutos de operação. Use $k = 0{,}12$/min e $\tau = 5$ min e suponha que as condições iniciais de todas as variáveis dependentes sejam zero.

28.11 Um reator de batelada pode ser descrito pelas seguintes equações:

$$\frac{dC}{dt} = -e^{(-10/(T+273))}C$$

$$\frac{dT}{dt} = 1000e^{(-10/(T+273))}C - 10(T - 20)$$

onde C é a concentração do reagente e T, a temperatura do reator. Inicialmente, o reator está a 15 ºC e tem uma concentração do reagente C de 1,0 gmol/L. Encontre a concentração e a temperatura do reator como uma função do tempo.

28.12 O seguinte sistema é um exemplo clássico de EDOs rígidas que podem ocorrer na cinética de reações químicas.

$$\frac{dc_1}{dt} = -0{,}013c_1 - 1000c_1c_3$$

$$\frac{dc_2}{dt} = -2500c_2c_3$$

$$\frac{dc_3}{dt} = -0{,}013c_1 - 1000c_1c_3 - 2500c_2c_3$$

Resolva essas reações de $t = 0$ a 50 com condições iniciais $c_1(0) = c_2(0) = 1$ e $c_3(0) = 0$. Se você tiver acesso ao software MATLAB, use ambas as funções padrão (por exemplo, `ode45`) e rígida (por exemplo, `ode23s`) para obter suas soluções.

Engenharia Civil/Ambiental

28.13 Faça os mesmos cálculos efetuados para o sistema Lotka-Volterra na Seção 28.2, mas use (**a**) o método de Euler, (**b**) o método de Heun (sem iteragir o corretor), (**c**) o método RK de quarta ordem e (**d**) a função `ode45` do MATLAB. Em todos os casos, use precisão simples das variáveis, um tamanho de passo de 0,1 e simulação de $t = 0$ a 20. Desenvolva gráficos no espaço de estados para todos os casos.

28.14 Faça os mesmos cálculos para as equações de Lorenz na Seção 28.2, mas use (**a**) o método de Euler, (**b**) o método de Heun (sem iteragir o corretor), (**c**) o método RK de quarta ordem e (**d**) a função `ode45` do MATLAB. Em todos os casos, utilize variáveis de precisão simples e tamanho de passo de 0,1 e simulação de $t = 0$ a 20. Desenvolva gráficos no espaço de estados para todos os casos.

28.15 A seguinte equação pode ser usada para modelar a deflexão do mastro da vela de um barco sujeita à força do vento:

$$\frac{d^2y}{dz^2} = \frac{f}{2EI}(L - z)^2$$

onde f é a força do vento, E é o módulo de elasticidade, L é o comprimento do mastro e I é o momento de inércia. Calcule a deflexão se $y = 0$ e $dy/dz = 0$ em $z = 0$. Use valores de parâmetros de $f = 60$, $L = 30$, $E = 1{,}25 \times 10^8$, e $I = 0{,}05$ para seus cálculos.

28.16 Faça os mesmos cálculos efetuados no Problema 28.15, mas, em vez de usar a força do vento constante, empregue uma força que varie com a altura de acordo com (lembre-se da Seção 24.2)

$$f(z) = \frac{200z}{5 + z}e^{-2z/30}$$

28.17 Um engenheiro ambiental está interessado em obter uma estimativa da mistura que ocorre entre um lago estratificado e a baía adjacente (Figura P28.17). Um rastreador conservativo é misturado instantaneamente com a água da baía e, a seguir, a concentração do rastreador é monitorada no período seguinte em todos os três segmentos. Os valores são

t	0	2	4	6	8	12	16	20
c_1	0	15	11	7	6	3	2	1
c_2	0	3	5	7	7	6	4	2
c_3	100	48	26	16	10	4	3	2

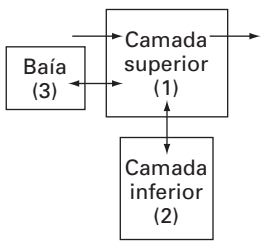


Figura P28.17

Usando balanços de massa, o sistema pode ser modelado pelas seguintes EDOs simultâneas:

$$V_1\frac{dc_1}{dt} = -Qc_1 + E_{12}(c_2 - c_1) + E_{13}(c_3 - c_1)$$

$$V_2\frac{dc_2}{dt} = E_{12}(c_1 - c_2)$$

$$V_3\frac{dc_3}{dt} = E_{13}(c_1 - c_3)$$

onde V_i é o volume do segmento i, Q é o escoamento e E_{ij} é a taxa de mistura difusiva entre os segmentos i e j. Use os dados e as equações diferenciais para obter uma estimativa dos E's se $V_1 = 1 \times 10^7$, $V_2 = 8 \times 10^6$, $V_3 = 5 \times 10^6$ e $Q = 4 \times 10^6$. Empregue o método de Euler com um tamanho de passo de 0,1 para a sua análise.

28.18 As dinâmicas de crescimento populacionais são importantes em diversos estudos de planejamento para áreas tais como de engenharia de transporte e de recursos hídricos. Um dos modelos mais simples para tal crescimento incorpora a hipótese de que a taxa de variação da população é proporcional à população existente em qualquer instante t:

$$\frac{dp}{dt} = Gp \tag{P28.18}$$

onde G é a taxa de crescimento (por ano). Esse modelo faz sentido do ponto de vista intuitivo, pois, quanto maior a população, maior o número de pais potenciais. No instante $t = 0$, uma ilha tem a população de 6.000 pessoas. Se $G = 0{,}075$ por ano, use o método de Heun (sem iteração) para prever a população em $t = 20$ anos, usando um tamanho de passo de 0,5 ano. Faça o gráfico de p em função de t em papel de gráfico padrão e semilog. Determine a inclinação da reta no gráfico do papel semilog. Discuta os resultados.

28.19 Embora o modelo no Problema 28.18 funcione adequadamente quando o crescimento populacional é ilimitado, ele não funciona quando fatores tais como escassez de alimentos, poluição e falta de espaço inibem o crescimento. Em tais casos, a própria taxa de crescimento pode ser pensada como inversamente proporcional à população. Um modelo para essa relação é

$$G = G'(p_{\max} - p) \tag{P28.19}$$

onde G' é uma taxa de crescimento dependente da população (por pessoa-ano) e $p_{\max}$ é o máximo de população sustentável. Logo, quando a população for pequena ($p \ll p_{\max}$), a taxa de crescimento será a taxa constante mais alta $G' p_{\max}$. Para tais casos, o crescimento é ilimitado e a Equação (P28.19) é essencialmente idêntica à Equação (P28.18). Entretanto, conforme a população cresce (isto é, p se aproxima de $p_{\max}$), G decresce até que em $p = p_{\max}$ ele seja zero. Portanto, o modelo prevê que, quando a população atingir o máximo nível sustentável, o crescimento será não existente e o sistema estará em um estado estacionário. A substituição da Equação (P28.19) na Equação (P28.18) fornece

$$\frac{dp}{dt} = G'(p_{\max} - p)p$$

Para a mesma ilha estudada no Problema 28.18, empregue o método de Heun (sem iteração) para prever a população em $t = 20$ anos, usando um tamanho de passo de 0,5 ano. Use valores de $G = 10^{-5}$ por pessoa-ano e $p_{\max} = 20.000$ pessoas. No instante $t = 0$, a ilha tem população de 6.000 pessoas. Faça o gráfico de p em função de t e interprete a forma da curva.

28.20 O Isle Royale National Park é um arquipélago de 210 milhas quadradas composto por uma única ilha grande e por muitas ilhas pequenas no Lago Superior. Os alces chegaram em torno de 1900 e, por volta de 1930, sua população se aproximou de 3.000, destruindo a vegetação. Em 1949, lobos cruzaram uma ponte de gelo a partir de Ontário. A partir do final da década de 50, o número de alces e de lobos foi monitorado. (Um traço indica que não há dados.)

Ano	Alces	Lobos	Ano	Alces	Lobos
1960	700	22	1972	836	23
1961	—	22	1973	802	24
1962	—	23	1974	815	30
1963	—	20	1975	778	41
1964	—	25	1976	641	43
1965	—	28	1977	507	33
1966	881	24	1978	543	40
1967	—	22	1979	675	42
1968	1000	22	1980	577	50
1969	1150	17	1981	570	30
1970	966	18	1982	590	13
1971	674	20	1983	811	23

(a) Integre as equações de Lotka-Volterra de 1960 até 2020. Determine os valores dos coeficientes que fornecem um ajuste ótimo. Compare sua simulação com os dados usando uma abordagem por série temporal e comente seus resultados.

(b) Trace a simulação da parte (**a**), mas use uma abordagem no espaço de estados.

(c) Depois de 1993, suponha que os administradores da vida selvagem consigam capturar um lobo por ano e transporte-o para fora da ilha. Faça uma previsão de como a população de lobos e alces iria evoluir até 2020. Apresente seus resultados tanto em gráfico de séries temporais quanto no espaço de estados. Para esse caso, bem como para (**d**), use os seguintes coeficientes: $a = 0{,}3$, $b = 0{,}01111$, $c = 0{,}2106$, $d = 0{,}0002632$.

(d) Suponha que em 1993, algum caçador conseguiu entrar na ilha e matou 50% dos alces. Faça uma previsão de como a população tanto dos lobos quanto dos alces iria evoluir até o ano de 2020. Apresente seus resultados tanto em gráfico de séries temporais quanto no espaço de estados.

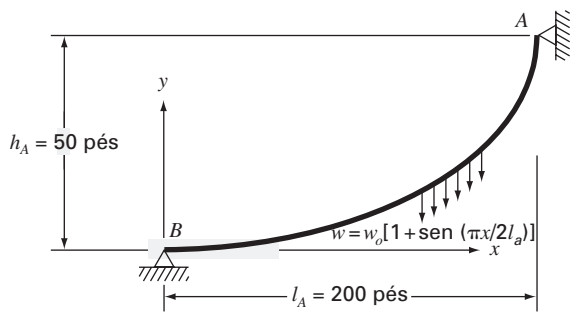
28.21 Um cabo está pendurado por dois suportes em A e B (Figura P28.21). O cabo está com carga, com uma distribuição de carga cujo módulo varia com x como

$$w = w_o\left[1 + \text{sen}\left(\frac{\pi x}{2l_A}\right)\right]$$

onde $w_o = 1000$ lbs/pés. A inclinação do cabo $(dy/dx) = 0$ em $x = 0$, que é o ponto mais baixo do cabo. Ele também é o ponto no qual a tensão no cabo está em um mínimo de T_o. A equação diferencial que governa o cabo é

$$\frac{d^2y}{dx^2} = \frac{w_o}{T_o}\left[1 + \text{sen}\left(\frac{\pi x}{2l_A}\right)\right]$$

Figura P28.21



Resolva essa equação usando um método numérico e trace a forma do cabo (y em função de x). Para a solução numérica, o valor de T_o é desconhecido, de modo que a solução deva usar uma técnica iterativa, análoga ao método *shooting*, para convergir para um valor correto de h_A para diversos valores de T_o.

28.22 A equação diferencial básica de uma curva elástica para uma viga em balanço (Figura P28.22) é dada por

$$EI\frac{d^2y}{dx^2} = -P(L-x)$$

onde E é o módulo de elasticidade e I é o momento de inércia. Determine a deflexão da barra usando um método numérico. Os seguintes valores dos parâmetros se aplicam: $E = 30.000$ ksi, $I = 800$ in^4, $P = 1$ kip, $L = 10$ pés. Compare seus resultados numéricos com a solução analítica,

$$y = -\frac{PLx^2}{2EI} + \frac{Px^3}{6EI}$$

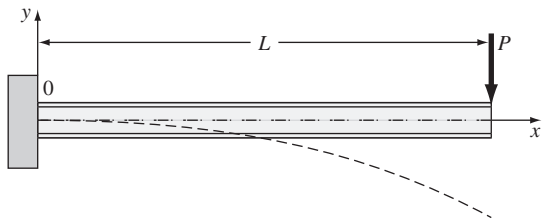


Figura P28.22

28.23 A equação diferencial básica de uma curva elástica para uma viga em balanço (Figura P28.23) é dada por

$$EI\frac{d^2y}{dx^2} = \frac{wLx}{2} - \frac{wx^2}{2}$$

onde E é o módulo de elasticidade e I é o momento de inércia. Determine a deflexão da barra usando (**a**) a abordagem por diferenças finitas ($\Delta x = 2$ pés) e (**b**) o método *shooting*. Os seguintes valores dos parâmetros se aplicam: $E = 30.000$ ksi, $I = 800$ in^4, $w = 1$ kip/pés, $L = 10$ pés. Compare seus resultados numéricos com a solução analítica,

$$y = \frac{wLx^3}{12EI} - \frac{wx^4}{24EI} - \frac{wL^3x}{24EI}$$

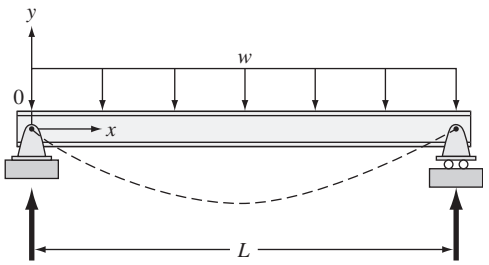


Figura P28.23

28.24 Um reservatório de água é drenado por um tubo, como mostrado na Figura P28.24. Sob diversas hipóteses simplificadoras, a seguinte equação diferencial descreve como a profundidade varia com o tempo:

$$\frac{dh}{dt} = \frac{\pi d^2}{4A(h)}\sqrt{2g(h+e)}$$

onde h é a profundidade (m), t é o tempo (s), d é o diâmetro do tubo (m), $A(h)$ é a área da superfície do reservatório como uma função da profundidade (m^2), g é a constante gravitacional ($= 9{,}81$ m/s^2), e e é a profundidade da saída do tubo abaixo do fundo do reservatório (m). Com base na seguinte tabela de área-profundidade, resolva essa equação diferencial para determinar quanto tempo leva para o reservatório se esvaziar, dado que $h(0) = 6$ m, $d = 0{,}25$ m, $e = 1$ m.

h, m	6	5	4	3	2	1	0
$A(h)$, 10^4 m^2	1,17	0,97	0,67	0,45	0,32	0,18	0

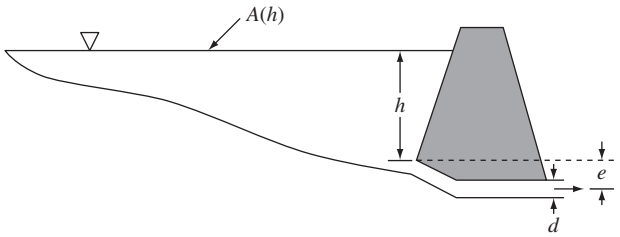


Figura P28.24

28.25 Engenheiros e cientistas usam modelos massa-mola para desenvolver uma percepção da dinâmica de estruturas sob a influência de distúrbios, tais como terremotos. A Figura P28.25 mostra tal representação para um edifício de três andares. Para esse caso, a análise é limitada ao movimento horizontal das estruturas. Os balanços de força podem ser escritos para esse modelo como

$$\left(\frac{k_1+k_2}{m_1} - \omega^2\right)X_1 \quad -\frac{k_2}{m_1} \quad X_2 \quad = 0$$

$$-\frac{k_2}{m_2} \quad X_1 + \left(\frac{k_2+k_3}{m_2} - \omega^2\right)X_2 - \frac{k_3}{m_2}X_3 = 0$$

$$-\frac{k_3}{m_3} \quad X_2 + \left(\frac{k_3}{m_3} - \omega^2\right)X_3 = 0$$

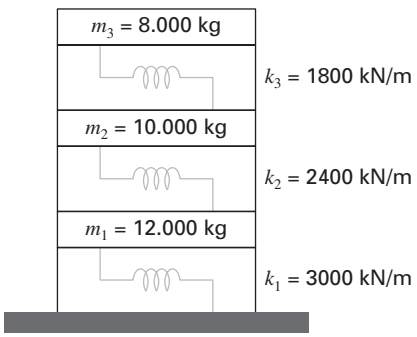


Figura P28.25

Determine os autovalores e os autovetores e a representação gráfica dos modos de vibração para a estrutura mostrando as amplitudes em função da altura para cada um dos autovetores. Normalize as amplitudes de modo que o deslocamento no terceiro piso seja um.

Engenharia Elétrica

28.26 Faça os mesmos cálculos realizados na primeira parte da Seção 28.3, mas com $R = 0{,}025\ \Omega$.

28.27 Resolva a EDO da primeira parte da Seção 8.3 de $t = 0$ a 0,5, usando técnicas numéricas se $q = 0{,}1$ e $i = -3{,}281515$ em $t = 0$. Use $R = 50$ com os outros parâmetros da Seção 8.3.

28.28 Para um circuito RL simples, a lei da voltagem de Kirchhoff exige que (se a lei de Ohm for válida)

$$L\frac{di}{dt} + Ri = 0$$

onde i é a corrente, L é a indutância e R é a resistência. Resolva para determinar i, se $L = 1$, $R = 1{,}5$, e $i(0) = 0{,}5$. Resolva esse problema analiticamente e com um método numérico. Apresente seus resultados graficamente.

28.29 Em contraste com o Problema 28.28, os resistores reais nem sempre obedecem à lei de Ohm. Por exemplo, a queda de voltagem pode não ser linear e a dinâmica do circuito poderia ser descrita por uma relação tal como

$$L\frac{di}{dt} + R\left[\frac{i}{I} - \left(\frac{i}{I}\right)^3\right] = 0$$

onde todos os outros parâmetros são como definidos no Problema 28.28 e I é uma corrente de referência conhecida igual a 1. Resolva para determinar i como uma função do tempo sob as mesmas condições especificadas no Problema 28.28.

28.30 Desenvolva um problema de autovalor para uma rede LC parecida com a da Figura 28.14, mas com apenas dois laços. Isto é, omita o laço i_3. Desenhe a rede, ilustrando como a corrente oscila em seus modos primários.

Engenharia Mecânica/Aeroespacial

28.31 Faça os mesmos cálculos realizados na Seção 28.4, mas para um pêndulo de 1 m de comprimento.

28.32 A Seção 8.4 apresenta uma equação diferencial de segunda ordem que pode ser usada para analisar as oscilações não forçadas do amortecedor de um automóvel. Dados $m = 1{,}2 \times 10^6$ g, $c = 1 \times 10^7$ g/s e $k = 1{,}25 \times 10^9$ g/s^2, use um método numérico para resolver no caso em que $x(0) = 0{,}4$ e $dx(0)/dt = 0{,}0$. Determine ambos o deslocamento e a velocidade de $t = 0$ a 0,5 s.

28.33 A taxa de resfriamento de um corpo pode ser expressa como

$$\frac{dT}{dt} = -k(T - T_a)$$

onde T é a temperatura do corpo (°C), T_a é a temperatura do meio circundante (°C) e k é a constante de proporcionalidade (min^{-1}). Logo, essa equação especifica que a taxa de resfriamento é proporcional à diferença de temperatura entre o corpo e o meio circundante. Se uma bola de metal aquecida a 90° C for jogada em água mantida a um valor constante de $T_a = 20°$ C, use um método numérico para calcular quanto tempo levará para que a bola seja resfriada para 40° C se $k = 0{,}25$ min^{-1}.

28.34 A taxa de fluxo de calor (condução) entre dois pontos em um cilindro aquecido em uma extremidade é dada por

$$\frac{dQ}{dt} = \lambda A\frac{dT}{dx}$$

onde λ é uma constante, A é a área da secção transversal do cilindro, Q é o fluxo de calor, T é a temperatura, t é o tempo e x é a distância da extremidade aquecida. Como a equação envolve duas derivadas, vamos simplificar essa equação, tomando

$$\frac{dT}{dx} = \frac{100(L - x)(20 - t)}{100 - xt}$$

onde L é o comprimento do cilindro. Combine essas duas equações e calcule o fluxo de calor de $t = 0$ a 25 s. A condição inicial é $Q(0)$ e os parâmetros são $\lambda = 0{,}5$ cal · cm/s, $A = 12$ cm^2, $L = 20$ cm e $x = 2{,}5$ cm. Faça o gráfico de seus resultados.

28.35 Repita o problema do pára-quedista em queda livre (Exemplo 1.2), porém com a força para cima em virtude do arrasto como de segunda ordem:

$$F_u = -cv^2$$

onde $c = 0{,}225$ kg/m. Resolva de $t = 0$ a 30, faça um gráfico de seus resultados e compare com aqueles do Exemplo 1.2.

28.36 Suponha que, depois de cair por 13 s, o pára-quedista dos Exemplos 1.1 e 1.2 puxa a corda de abertura. Nesse ponto, suponha que o coeficiente de arrasto aumente instantaneamente para um valor constante de 55 kg/s. Calcule a velocidade do pára-quedista de $t = 0$ a 30 s com o método de Heun (sem iteração para o corretor), usando um tamanho de passo de 2 s. Trace v em função de t de $t = 0$ a 30s.

28.37 A seguinte equação diferencial ordinária descreve o movimento de um sistema massa-mola amortecido (Figura P8.37):

$$m\frac{d^2x}{dt^2} + a\left|\frac{dx}{dt}\right|\frac{dx}{dt} + bx^3 = 0$$

onde x é o deslocamento a partir da posição de equilíbrio, t é o tempo, $m = 1$ kg é a massa e $a = 5$ N/(m/s)^2. O termo de amortecimento é não-linear e representa o amortecimento do ar.

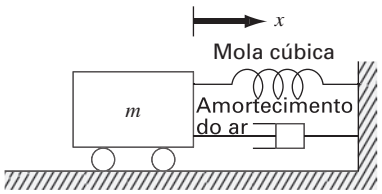


Figura P28.37

A mola é cúbica, também não-linear, com $b = 5$ N/m^3. As condições iniciais são

Velocidade inicial	$\frac{dx}{dt} = 0{,}5$ m/s
Deslocamento inicial	$x = 1$ m

Resolva essa equação usando um método numérico no período $0 \le t \le 8$ s. Faça o gráfico do deslocamento e da velocidade em função do tempo e trace o retrato no plano de fase (velocidade versus deslocamento) para todos os casos a seguir:

(a) Uma equação linear análoga

$$m\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + 5x = 0$$

(b) A equação não-linear apenas com o termo não-linear na mola

$$\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + bx^3 = 0$$

(c) A equação não-linear apenas com o termo de amortecimento não-linear

$$m\frac{d^2x}{dt^2} + a\left|\frac{dx}{dt}\right|\frac{dx}{dt} + 5x = 0$$

(d) A equação não-linear completa, em que tanto o termo de amortecimento quanto o termo da mola são não-lineares

$$m\frac{d^2x}{dt^2} + a\left|\frac{dx}{dt}\right|\frac{dx}{dt} + bx^3 = 0$$

28.38 Um sistema massa-mola amortecido forçado (Figura P28.38) tem a seguinte equação diferencial ordinária de movimento:

$$m\frac{d^2x}{dt^2} + a\left|\frac{dx}{dt}\right|\frac{dx}{dt} + kx = F_o\,\text{sen}(\omega t)$$

onde x é o deslocamento a partir da posição de equilíbrio, t é o tempo, $m = 2$ kg é a massa, $a = 5$ N/(m/s)^2, e $k = 6$ N/(m/s)^2. O termo de amortecimento é não-linear e representa o amortecimento do ar. O termo forçante $F_o\,\text{sen}(\omega t)$ tem valores de $F_o = 2{,}5$ N e $\omega = 0{,}5$ rad/s. As condições iniciais são

Velocidade inicial	$\dfrac{dx}{dt} = 0$ m/s
Deslocamento inicial	$x = 1$ m

Resolva essa equação utilizando um método numérico no período $0 \le t \le 15$ s. Trace o deslocamento e a velocidade em função do tempo, bem como o termo forçante no mesmo gráfico. Além disso, faça um gráfico separado da velocidade *versus* o deslocamento.

Figura P28.38

28.39 A distribuição de temperaturas em um apêndice de resfriamento cônico em forma de aleta (Figura P29.39) é descrito pela seguinte equação diferencial, que foi adimensionalizada

$$\frac{d^2u}{dx^2} + \left(\frac{2}{x}\right)\left(\frac{du}{dx}pu\right) = 0$$

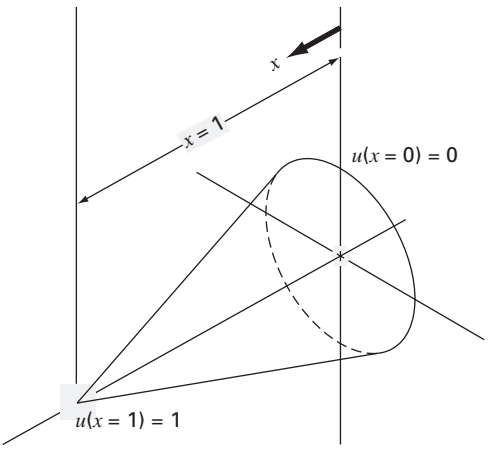


Figura P28.39

onde u é a temperatura ($0 \le u \le 1$), x é a distância radial ($0 \le x \le 1$) e p é um parâmetro adimensional que descreve a transferência de calor e a geometria

$$p = \frac{hL}{k}\sqrt{1 + \frac{4}{2m^2}}$$

onde h é o coeficiente de transferência de calor, k é a condutividade térmica, L é o comprimento ou a altura do cone e m é a inclinação da parede do cone. A equação tem as condições de contorno

$$u(x = 0) = 0 \qquad u(x = 1) = 1$$

Resolva essa equação para a distribuição de temperatura usando métodos de diferenças finitas. Use para derivadas diferenças finitas acuradas até segunda ordem. Escreva um programa computacional para obter a solução e trace a temperatura *versus* a distância axial para os diversos valores de $p = 10$, 20, 50 e 100.

28.40 A dinâmica de um sistema massa-mola amortecido forçado pode ser representada pela seguinte EDO de segunda ordem:

$$m\frac{d^2x}{dt^2} + c\frac{dx}{dt} + k_1x + k_3x^3 = P\cos(\omega t)$$

onde $m = 1$ kg, $c = 0{,}4$ N · s/m, $P = 0{,}5$ N, e $\omega = 0{,}5$/s. Use um método numérico para determinar o deslocamento (x) e a velocidade ($v = dx/dt$) como uma função do tempo com as condições iniciais $x = v = 0$. Expresse seus resultados graficamente como gráficos de séries temporais (x e v em função de t) e com um gráfico no plano de fase (v *versus* x). Faça simulações para ambas as molas (**a**) linear ($k_1 = 1$; $k_3 = 0$) e (**b**) não-linear ($k_1 = 1$; $k_3 = 0{,}5$).

28.41 A equação diferencial para a velocidade de um *bungee jumper* é diferente, dependendo se o saltador já caiu até uma distância na qual a corda está completamente estendida e começa a esticar. Logo, se a distância de queda for menor que o comprimento da corda, o saltador estará sujeito apenas às forças gravitacional e de arrasto. Uma vez que a corda começa a esticar, as forças elástica e de amortecimento da corda também devem ser incluídas. Essas duas condições podem ser expressas pelas seguintes equações:

$$\frac{dv}{dt} = g - \text{sign}(v)\frac{c_d}{m}v^2 \qquad x \le L$$

$$\frac{dv}{dt} = g - \text{sign}(v)\frac{c_d}{m}v^2 - \frac{k}{m}(x - L) - \frac{\gamma}{m}v \qquad x > L$$

onde v é a velocidade (m/s), t é o tempo (s), g é a constante gravitacional (= 9,81 m/s^2), sign(x) é a função que retorna -1, 0 e 1 para valores de x negativos, nulo e positivos, respectivamente, c_d é o coeficiente de arrasto de segunda ordem (kg/m), m é a massa (kg), k é a constante de mola da corda (N/m), γ é o coeficiente de amortecimento da corda (N · s/m) e L é o comprimento da corda (m). Determine a posição e a velocidade do saltador, dados os seguintes parâmetros: $L = 30$ m, $m = 68{,}1$ kg, $c_d = 0{,}25$ kg/m, $k = 40$ N/m e $\gamma = 8$ kg/s. Faça os cálculos de $t = 0$ a 50s e suponha que as condições iniciais sejam $x(0) = v(0) = 0$.



EPÍLOGO: PARTE SETE

PT7.4 PRÓS E CONTRAS

A Tabela PT7.3 contém os prós e contras associados aos métodos numéricos para a solução de equações diferenciais ordinárias com condições iniciais. Os fatores nessa tabela devem ser avaliados pelo engenheiro quando ele estiver escolhendo um método para cada problema aplicado particular.

Técnicas de início automático simples, tal como o método de Euler, podem ser usadas se o problema envolver um pequeno intervalo de integração. Nesse caso, a acurácia adequada pode ser obtida usando-se pequenos tamanhos de passos para evitar grandes erros de truncamento e os erros de arredondamento podem ser aceitáveis. O método de Euler também pode ser apropriado para os casos nos quais o modelo matemático tenha alto nível de incerteza inerente ou possua coeficientes ou termos forçantes com erros significativos, como poderia ocorrer durante um processo de medida. Portanto, a acurácia do próprio modelo simplesmente não justifica o esforço envolvido na utilização de métodos numéricos mais complicados. Finalmente, as técnicas mais simples podem ser melhores quando o problema ou a simulação precisarem ser feitos poucas vezes. Nessas aplicações, provavelmente é melhor usar um método simples que seja fácil de programar e entender apesar do fato de o método ser computacionalmente ineficiente e de consumir tempo de execução relativamente alto no computador.

TABELA PT7.3 Comparação das características dos métodos alternativos para a resolução numérica de EDOs. As comparações são baseadas em experiências gerais e não levam em conta o comportamento de funções especiais.

Método	Valores Iniciais	Necessidade de Iterações	Erro Global	Facilidade na Mudança do Tamanho do Passo	Esforço de Programação	Comentários
Passo único						
de Euler	1	Não	$O(h)$	Fácil	Fácil	Bom para estimativas rápidas
de Heun	1	Sim	$O(h^2)$	Fácil	Moderado	—
do ponto médio	1	Não	$O(h^2)$	Fácil	Moderado	—
de Ralston de segunda ordem	1	Não	$O(h^2)$	Fácil	Moderado	O método RK de segunda ordem que minimiza o erro de truncamento
RK de quarta ordem	1	Não	$O(h^4)$	Fácil	Moderado	Amplamente usado
RK adaptativo de quarta ordem ou RK Fehlberg	1	Não	$O(h^5)$*	Fácil	Moderado a difícil	A estimativa de erro permite o ajuste no tamanho do passo
Passo múltiplo						
Heun de início não-automático	2	Sim	$O(h^3)$*	Difícil	Moderado a difícil†	Método de passo múltiplo simples
de Milne	4	Sim	$O(h^5)$*	Difícil	Moderado a difícil†	Algumas vezes é instável
Adams de quarta ordem	4	Sim	$O(h^5)$*	Difícil	Moderado a difícil†	

*Desde que a estimativa de erro seja usada para modificar a solução.

†Com tamanho de passo variável.

Se o intervalo de integração do problema for suficientemente longo para envolver um grande número de passos, então pode ser necessário e apropriado usar uma técnica mais acurada que o método de Euler. O método RK de quarta ordem é popular e confiável para muitos problemas de engenharia. Nesses casos, também pode ser aconselhável fazer uma estimativa do erro de truncamento para cada passo como um guia na escolha do melhor tamanho do passo. Pode-se conseguir isso com as abordagens RK adaptativas ou Adams de quarta ordem. Se os erros de truncamento forem extremamente pequenos, pode ser prudente aumentar o tamanho do passo para economizar tempo de computador. Por outro lado, se o erro de truncamento for grande, o tamanho do passo deveria ser diminuído para evitar o acúmulo de erros. O método de Milne deve ser evitado se forem esperados problemas significativos de estabilidade. O método de Runge-Kutta é simples de programar e conveniente para usar, mas pode ser menos eficiente que métodos de passo múltiplo.

Um grande número de problemas de engenharia pode cair em uma região intermediária de intervalo de integração e de exigências de acurácia. Assim, os métodos RK de segunda ordem e Heun de início não-automático são simples de usar e são relativamente eficientes e acurados.

Os *sistemas rígidos* envolvem equações com componentes variando rapidamente e outras, lentamente. Técnicas especiais em geral são necessárias para uma resolução adequada de equações rígidas. Por exemplo, as abordagens implícitas são usadas freqüentemente. Você pode consultar Enright et al. (1975), Gear (1971) e Shampine e Gear (1979) para informações adicionais relativas a essas técnicas.

Diversas técnicas estão disponíveis para a resolução de problemas de autovalor. Para pequenos sistemas ou quando apenas uns poucos dos menores ou dos maiores autovalores forem necessários, abordagens simples como o método polinomial ou de potências estão disponíveis. Para sistemas simétricos, os métodos de Jacobi, Given ou Householder podem ser empregados. Finalmente, o método QR representa uma abordagem geral para determinar os autovalores de matrizes simétricas ou não simétricas.

PT7.5 RELAÇÕES E FÓRMULAS IMPORTANTES

A Tabela PT7.4 resume informações importantes que foram apresentadas na Parte Sete. Essa tabela pode ser consultada para acesso rápido às relações e às fórmulas importantes.

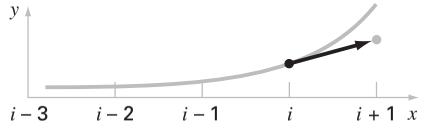
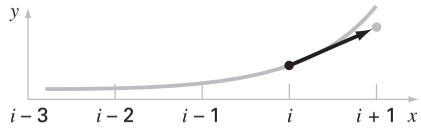
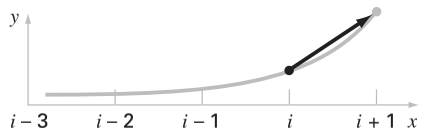
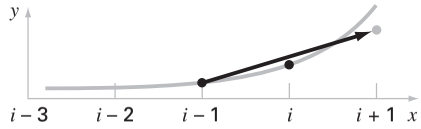
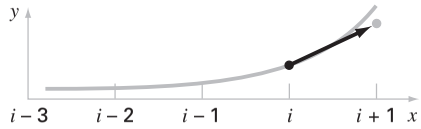
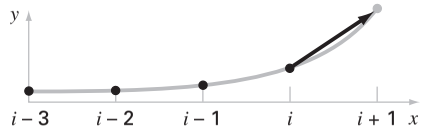
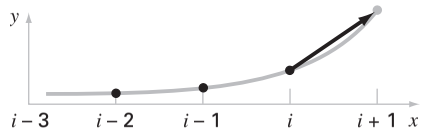
PT7.6 MÉTODOS AVANÇADOS E REFERÊNCIAS ADICIONAIS

Embora tenhamos revisado diversas técnicas para a solução de equações diferenciais ordinárias, existe informação adicional que é importante na prática da engenharia. A questão da *estabilidade* foi introduzida na Seção 26.2.4. Esse tópico tem relevância geral em todos os métodos para resolver EDOs. Mais discussão sobre esse tópico pode ser encontrada em Carnahan, Luther e Wilkes (1969), Gear (1971) e Hildebrand (1974).

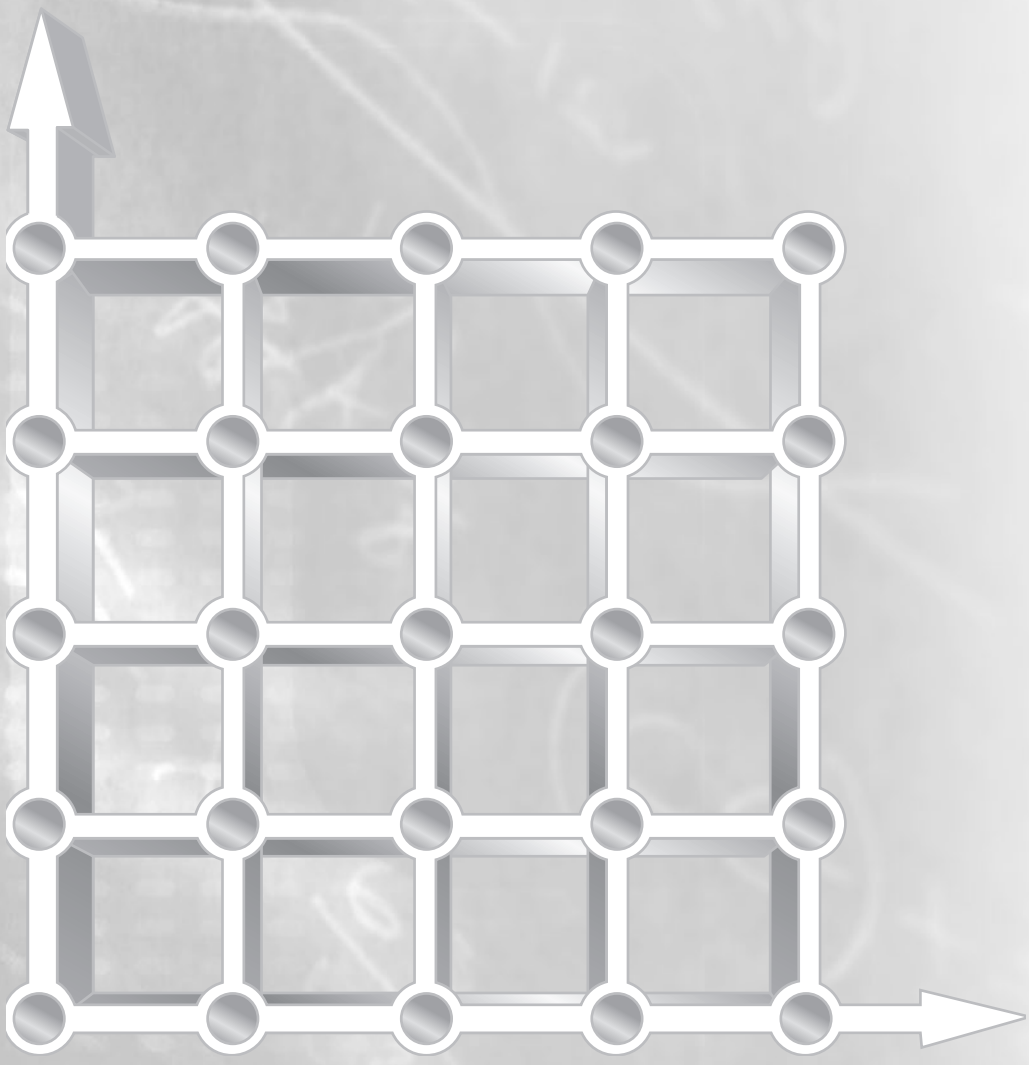
No Capítulo 27, introduzimos métodos para a resolução de *problemas de contorno*. Isaacson e Keller (1966), Keller (1968), Na (1979) e Scott e Watts (1976) podem ser consultados para informação adicional sobre problemas de contorno padrão. Material adicional sobre autovalores pode ser encontrado em Ralston e Rabinowitz (1978), Wilkinson (1965), Fadeev e Fadeeva (1963) e Householder (1953, 1964).

Em síntese, o propósito do material anterior é lhe fornecer avenidas para uma exploração mais profunda do assunto. Além disso, todas as referências anteriores fornecem descrições das técnicas básicas cobertas na Parte Sete. Recomendamos que você consulte essas fontes alternativas para ampliar seu entendimento dos métodos numéricos para a resolução de equações diferenciais.

TABELA PT7.4 Resumo de informações importantes apresentada na Parte Sete.

Método	Formulação	Interpretação Gráfica	Erros
Euler (RK de primeira ordem)	$y_{i+1} = y_i + hk_1$ $k_1 = f(x_i, y_i)$		Erro local $\simeq O(h^2)$ Erro global $\simeq O(h)$
RK de segunda ordem de Ralston	$y_{i+1} = y_i + h(\frac{1}{3}k_1 + \frac{2}{3}k_2)$ $k_1 = f(x_i, y_i)$ $k_2 = f(x_i + \frac{3}{4}h, y_i + \frac{3}{4}hk_1)$		Erro local $\simeq O(h^3)$ Erro global $\simeq O(h^2)$
RK clássico de quarta ordem	$y_{i+1} = y_i + h(\frac{1}{6}k_1 + \frac{1}{3}k_2 + \frac{1}{3}k_3 + \frac{1}{6}k_4)$ $k_1 = f(x_i, y_i)$ $k_2 = f(x_i + \frac{1}{2}h, y_i + \frac{1}{2}hk_1)$ $k_3 = f(x_i + \frac{1}{2}h, y_i + \frac{1}{2}hk_2)$ $k_4 = f(x_i + h, y_i + hk_3)$		Erro local $\simeq O(h^5)$ Erro global $\simeq O(h^4)$
Heun de início não-automático	Preditor: (método do ponto médio) $y^0_{i+1} = y^m_{i-1} + 2hf(x_i, y^m_i)$		Modificador do preditor: $E_p \simeq \frac{4}{5}(y^m_{i,u} + y^0_{i,u})$
	Corretor: (regra do trapézio) $y^j_{i+1} = y^m_i + h\ \dfrac{f(x_i, y^m_i) + f(x_{i+1}, y^{j-1}_{i+1})}{2}$		Modificador do corretor: $E_c \simeq -\dfrac{y^m_{i+1,u} - y^0_{i+1,u}}{5}$
Adams de quarta ordem	Preditor: (Adams-Bashforth de quarta ordem) $y^0_{i+1} = y^m_i + h(\frac{55}{24}f^m_i - \frac{59}{24}f^m_{i-1} + \frac{37}{24}f^m_{i-2} - \frac{9}{24}f^m_{i-3})$		Modificador do preditor: $E_p \simeq \frac{251}{270}(y^m_{i,u} - y^0_{i,u})$
	Corretor: (Adams-Moulton de quarta ordem) $y^j_{i+1} = y^m_i + h(\frac{9}{24}f^{j-1}_{i+1} + \frac{19}{24}f^m_i - \frac{5}{24}f^m_{i-1} + \frac{1}{24}f^m_{i-2})$		Modificador do corretor: $E_c \simeq -\frac{19}{270}(y^m_{i+1,u} - y^0_{i+1,u})$

PARTE OITO



EQUAÇÕES DIFERENCIAIS PARCIAIS

PT 8.1 MOTIVAÇÃO

Dada uma função u que dependa tanto de x quanto de y, a derivada parcial de u com relação a x em um ponto arbitrário (x, y) é definida por

$$\frac{\partial u}{\partial x} = \lim_{\Delta x \to 0} \frac{u(x + \Delta x, y) - u(x, y)}{\Delta x} \tag{PT8.1}$$

Analogamente, a derivada parcial com relação a y é definida por

$$\frac{\partial u}{\partial y} = \lim_{\Delta y \to 0} \frac{u(x, y + \Delta y) - u(x, y)}{\Delta y} \tag{PT8.2}$$

Uma equação envolvendo as derivadas parciais de uma função desconhecida de duas ou mais variáveis independentes é chamada *equação diferencial parcial* ou *EDP*. Por exemplo,

$$\frac{\partial^2 u}{\partial x^2} + 2xy\frac{\partial^2 u}{\partial y^2} + u = 1 \tag{PT8.3}$$

$$\frac{\partial^3 u}{\partial x^2 \partial y} + x\frac{\partial^2 u}{\partial y^2} + 8u = 5y \tag{PT8.4}$$

$$\left(\frac{\partial^2 u}{\partial x^2}\right)^3 + 6\frac{\partial^3 u}{\partial x \partial y^2} = x \tag{PT8.5}$$

$$\frac{\partial^2 u}{\partial x^2} + xu\frac{\partial u}{\partial y} = x \tag{PT8.6}$$

A *ordem* de uma EDP é aquela da derivada parcial de maior ordem que aparece na equação. Por exemplo, as Equações (PT8.3) e (PT8.4) são de segunda e terceira ordens, respectivamente.

Uma equação diferencial parcial é dita linear se ela for linear na função desconhecida e em todas as suas derivadas, com coeficientes dependendo apenas das variáveis independentes. Por exemplo, as Equações (PT8.3) e (PT8.4) são lineares, enquanto as Equações (PT8.5) e (PT8.6) não são.

Por causa de sua ampla gama de aplicações na engenharia, nosso tratamento das EDPs se concentrará nas equações diferenciais de segunda ordem. Para duas variáveis independentes, tais equações podem ser expressas na seguinte forma geral:

$$A\frac{\partial^2 u}{\partial x^2} + B\frac{\partial^2 u}{\partial x \partial y} + C\frac{\partial^2 u}{\partial y^2} + D = 0 \tag{PT8.7}$$

onde A, B e C são funções de x e y e D é uma função de x, y, u, $\partial u/\partial x$, e $\partial u/\partial y$. Dependendo dos valores dos coeficientes dos termos de segunda ordem — A, B , C —, a Equação PT8.7 pode ser classificada em uma de três categorias (Tabela PT8.1). Essa classificação, que é baseada no método das características (veja, por exemplo, Vichnevetsky, 1981, ou Lapidus e Pinder, 1981), é útil porque cada categoria está relacionada com contexto de problemas de engenharia específicos e distintos que exigem técnicas de solução

TABELA PT8.1 Categorias nas quais as equações diferenciais parciais lineares de segunda ordem em duas variáveis podem ser classificadas.

$B^2 - 4AC$	Categoria	Exemplo
< 0	Elíptica	Equação de Laplace (estado estacionário com duas dimensões espaciais) $\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0$
$= 0$	Parabólica	Equação de condução de calor (variável no tempo com uma dimensão espacial) $\frac{\partial T}{\partial t} = k' \frac{\partial^2 T}{\partial x^2}$
> 0	Hiperbólica	Equação da onda (variável no tempo com uma dimensão espacial) $\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 y}{\partial t^2}$

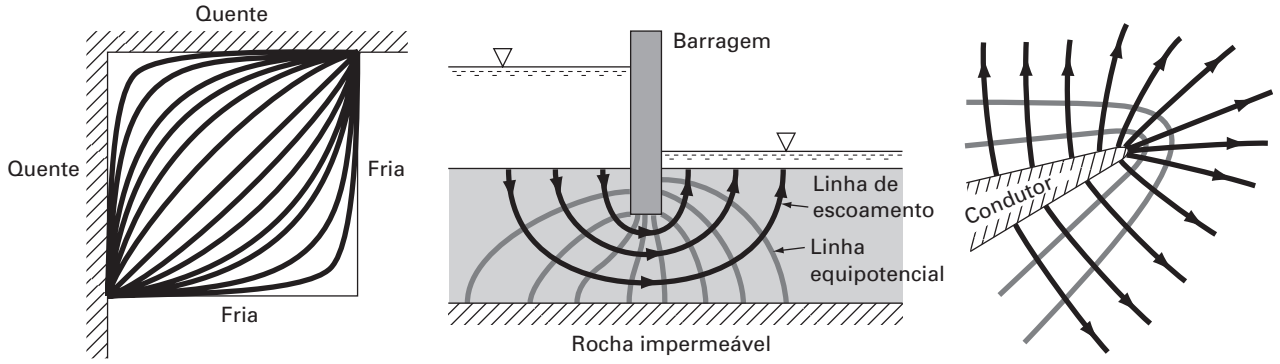
especiais. Deve ser observado que, para os casos nos quais A, B e C dependam de x e y, na realidade, a equação pode cair em categorias diferentes, dependendo da posição do domínio para o qual a equação é válida. Por simplicidade, limitaremos a discussão seguinte às EDPs que permanecem exclusivamente em uma das categorias.

PT 8.1.1 EDPs e a Prática da Engenharia

Cada uma das categorias de equações diferenciais parciais na Tabela PT8.1 se adequa a tipos específicos de problemas de engenharia. As sessões iniciais dos próximos capítulos serão dedicadas à dedução de cada tipo de equação para um contexto de problema de engenharia particular. Por enquanto, discutiremos suas propriedades e aplicações gerais e mostraremos como elas podem ser empregadas em diferentes contextos físicos.

As *equações elípticas* são tipicamente usadas para caracterizar sistemas *estacionários*. Como na *equação de Laplace* na Tabela PT 8.1, isso é indicado pela ausência de uma derivada no tempo. Assim, essas equações são usadas tipicamente para determinar distribuições estacionárias de uma incógnita em duas dimensões espaciais.

FIGURA PT 8.1
Três problemas de distribuições estacionários que podem ser caracterizados por EDPs elípticas. (*a*) Distribuição de temperatura em uma placa aquecida, (*b*) infiltração de água sob uma barragem e (*c*) o campo elétrico perto da ponta de um condutor.



(*a*) (*b*) (*c*)

Um exemplo simples é a chapa aquecida na Figura PT8.1*a*. Nesse caso, as bordas da placa são mantidas a temperaturas diferentes. Como o calor flui de regiões de alta para baixa temperatura, as condições de contorno provocam um potencial que leva a um fluxo de calor das bordas quentes para as frias. Se passar tempo suficiente, tal sistema eventualmente atingirá a distribuição estável ou estacionária de temperatura descrita na Figura PT8.1*a*. A equação de Laplace, com as condições de contorno apropriadas, fornecem meios para determinar essa distribuição. Por analogia, a mesma abordagem pode ser empregada para atacar outros problemas envolvendo potenciais, tal como a infiltração de água sob uma barragem (Figura PT8.1*b*) ou a distribuição de um campo elétrico (Figura PT8.1*c*).

Em contraste à categoria elíptica, as *equações parabólicas* determinam como uma incógnita varia tanto no espaço quanto no tempo. Isso se manifesta pela presença tanto de derivadas espaciais quanto temporais na *equação de condução de calor* na Tabela PT 8.1. Tais casos são conhecidos como *problemas de propagação* porque a solução se "propaga" ou varia no tempo.

Um exemplo simples é uma haste longa e fina que seja isolada em toda a parte, exceto em suas extremidades (Figura PT8.2*a*). O isolamento é usado para evitar complicações em razão da perda de calor ao longo do comprimento da barra. Como no caso para a barra aquecida na Figura PT8.1*a*, as extremidades da barra são mantidas a temperaturas fixas. Entretanto, em contraste à Figura PT8.1*a*, o fato de a barra ser fina nos permite supor que o calor esteja distribuído uniformemente sobre sua seção transversal — isto é, lateralmente. Como conseqüência, fluxo lateral de calor não é uma questão e o problema se reduz ao estudo da condução de calor ao longo do eixo longitudinal da barra. Em vez de nos concentrarmos na distribuição estacionária em duas dimensões, o problema é determinar como a distribuição espacial unidimensional varia como uma função do tempo (Figura PT8.2*b*). Assim, a solução consiste em uma série de distribuições espaciais correspondentes aos estados da barra em vários instantes. Fazendo uma analogia da fotografia, o caso elíptico fornece um retrato de um estado estável de um sistema, enquanto o caso parabólico fornece um filme em movimento de como ele varia de um estado para outro. Como com outros tipos de EDPs descritas aqui, as equações parabólicas podem ser usadas para caracterizar outros contextos de engenharia por analogia.

A classe final de EDPs, a categoria *hiperbólica*, também trata de *problemas de propagação*. Entretanto, uma diferença importante manifestada pela equação de onda na Tabela PT 8.1 é que a incógnita é caracterizada pela segunda derivada com relação ao tempo. Como conseqüência, a solução oscila.

FIGURA PT 8.2
(*a*) Uma haste longa e fina é isolada em toda a parte, exceto nas extremidades. A dinâmica da distribuição unidimensional da temperatura ao longo do comprimento da haste pode ser descrita por uma EDP parabólica. (*b*) A solução, que consiste nas distribuições correspondentes ao estado da haste em vários instantes.

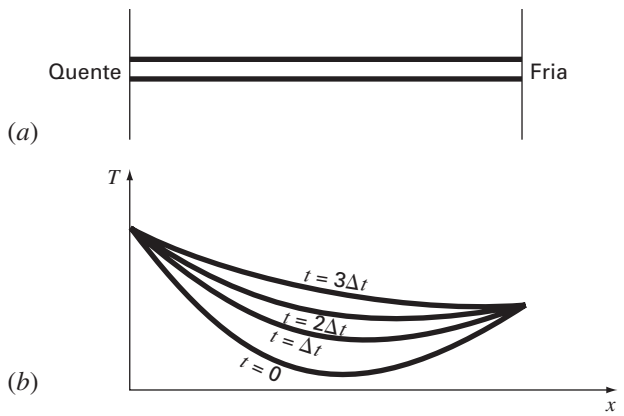




FIGURA PT 8.3
Uma corda tensionada vibrando com pequena amplitude é um sistema físico simples que pode ser caracterizado por uma EDP hiperbólica.

A corda vibrante na Figura PT 8.3 é um modelo físico simples que pode ser descrito com uma equação de onda. A solução consiste em diversos estados característicos nos quais a corda oscila. Diversos sistemas em engenharia, tais como as vibrações de hastes e vigas, movimentos de ondas de fluido e transmissão do som e de sinais elétricos podem ser caracterizados por esse modelo.

PT 8.1.2 Métodos Pré-computacionais para Resolver EDPs

Antes do aparecimento dos computadores digitais, os engenheiros se baseavam em soluções analíticas ou exatas de equações diferenciais parciais. Além dos casos mais simples, essas soluções freqüentemente exigiam um grande esforço e sofisticação matemática. Além disso, muitos sistemas físicos não podiam ser resolvidos diretamente, mas tinham de ser simplificados usando-se linearizações, representações geométricas simples e outras idealizações. Embora essas soluções sejam elegantes e forneçam uma solução geral, elas são limitadas em relação a quão fielmente elas representam os sistemas reais — especificamente aqueles que são altamente não-lineares e de forma irregular.

PT 8.2 ORIENTAÇÃO

Antes de prosseguir para os métodos numéricos para resolver equações diferenciais parciais, alguma orientação pode ser útil. O propósito de o material a seguir é fornecer a você uma visão geral da abordagem que será discutida na Parte Oito. Além do mais, formulamos objetivos para concentrar seus estudos no assunto.

PT 8.2.1 Escopo e Visão Geral

A Figura PT 8.4 fornece uma visão geral da Parte Oito. Duas amplas categorias de métodos numéricos serão discutidas nesta parte do livro. As abordagens por diferenças finitas, as quais são cobertas nos Capítulos 29 e 30, são baseadas na aproximação da solução em um número finito de pontos. Em contraste, os métodos de elementos finitos, os quais são cobertos no Capítulo 31, aproximam a solução em partes ou "elementos". Diversos parâmetros são ajustados até essas aproximações ficarem de acordo com a equação diferencial subjacente em um sentido ótimo.

O *Capítulo 29* é dedicado às soluções por *diferenças finitas* de *equações elípticas*. Antes de nos lançarmos aos métodos, deduziremos a equação de Laplace para o contexto de problema físico da distribuição de temperatura em uma placa aquecida. A seguir, uma abordagem-padrão de solução, o *método de Liebmann*, é descrita. Ilustraremos como essa abordagem é usada para calcular a distribuição de uma variável escalar primária, a temperatura, bem como de uma variável vetorial secundária, o fluxo de calor. A seção final deste capítulo trata das *condições de contorno*. Esse material inclui procedimentos para manipular diferentes tipos de condições, assim como fronteiras irregulares.

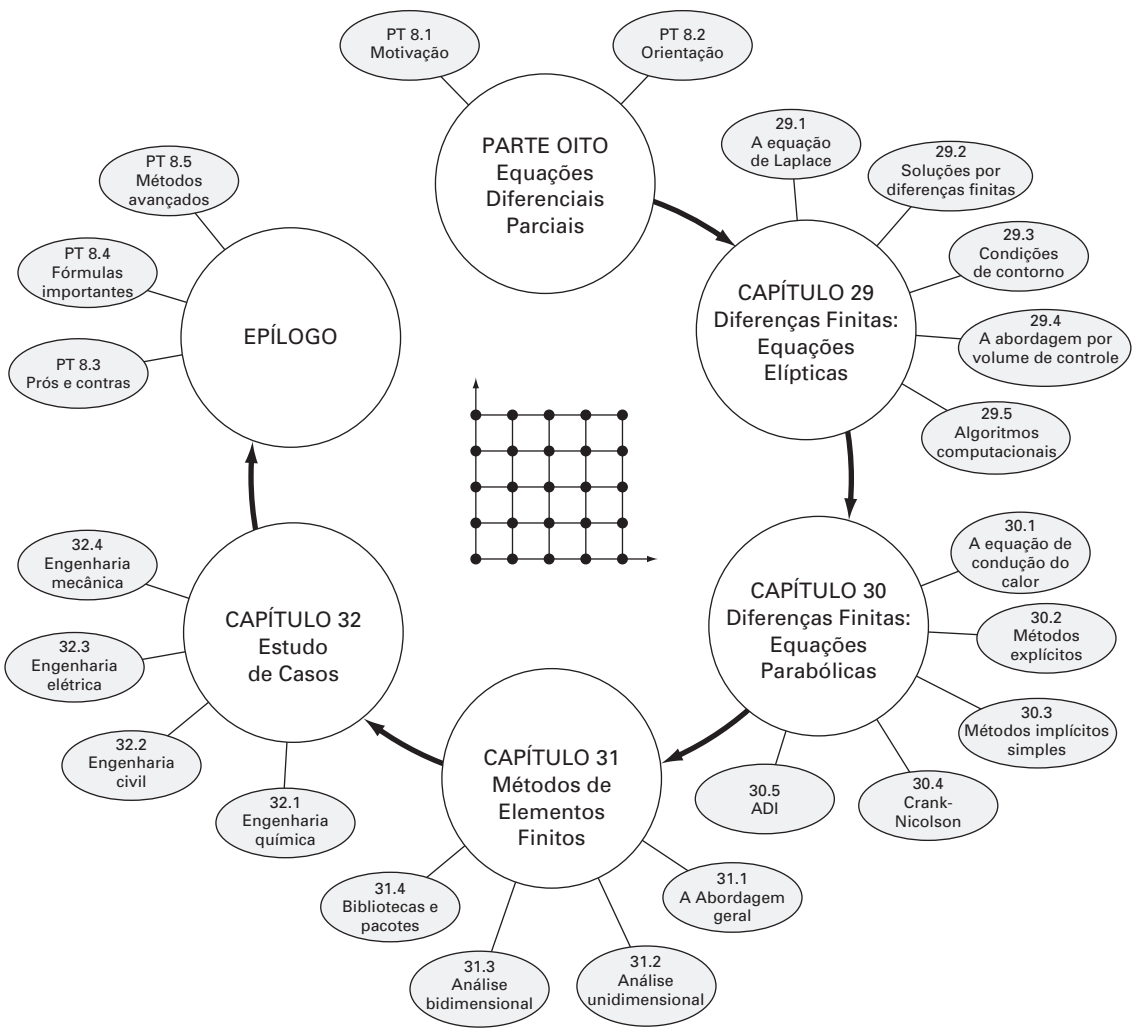
No *Capítulo 30*, nos voltaremos para as soluções por *diferenças finitas* de *equações parabólicas*. Como na discussão para as equações elípticas, forneceremos, inicialmente, uma introdução a um contexto de problema físico, a equação de condução de calor para uma barra unidimensional. A seguir, introduziremos algoritmos tanto explícitos quanto implícitos para resolver essa equação. Isso é seguido por um método implícito eficiente e confiável — a *técnica de Crank-Nicholson*. Finalmente, descreveremos uma abordagem efetiva particular para resolver equações parabólicas bidimensionais — o *método implícito de direção alternada* ou *ADI*.

Observe que, como elas estão um pouco além do escopo deste livro, preferimos omitir as equações hiperbólicas. O epílogo desta parte contém referências relativas a esse tipo de EDPs.

No *Capítulo 31*, nos voltamos para a outra abordagem principal para a resolução de EDPs — o *método de elementos finitos*. Como ele é tão fundamentalmente diferente da abordagem por diferenças finitas, nos dedicaremos à seção inicial do capítulo a uma visão geral. Então, mostraremos como o método de elementos finitos é usado para calcular a distribuição de temperatura estacionária de uma barra aquecida. Por fim, forneceremos uma introdução a algumas das questões envolvidas na extensão de tal análise a contextos de problemas tridimensionais.

O *Capítulo 32* é dedicado a aplicações em todos os campos da engenharia. Finalmente, uma pequena seção de revisão será incluída no final da Parte Oito. Esse epílogo resume informações importantes relacionadas com as EDPs. Este material inclui uma discussão dos prós e contras que são relevantes em suas implementações na prática da engenharia. Esse epílogo também inclui referências para tópicos avançados.

FIGURA PT 8.4
Representação esquemática da organização do material na Parte Oito: Equações Diferenciais Parciais.



PT 8.2.2 Metas e Objetivos

Objetivos de estudo. Depois de completar a Parte Oito, você deve ter aumentado muito seus recursos para confrontar e resolver equações diferenciais parciais. As metas de estudo gerais devem incluir o domínio das técnicas tendo a capacidade de avaliar a confiabilidade das respostas e sendo capaz de escolher o "melhor" método (ou métodos) para qualquer problema particular. Além desses objetivos gerais, os objetivos de estudo específicos na Tabela PT 8.2 devem ser dominados.

Objetivos computacionais. Podem ser desenvolvidos algoritmos computacionais para muitos dos métodos na Parte Oito. Por exemplo, você pode achar instrutivo desenvolver um programa geral para simular a distribuição estacionária de temperatura em uma placa aquecida. Além disso, você pode querer desenvolver programas para implementar ambos os métodos explícitos simples e de Crank-Nicholson para resolver EDPs parabólicas em uma dimensão espacial.

TABELA PT8.2 Objetivos de estudo específicos para a Parte Oito.

1. Reconhecer a diferença entre EDPs elípticas, parabólicas e hiperbólicas.
2. Entender a diferença fundamental entre as abordagens por diferenças finitas e por elementos finitos.
3. Reconhecer que o método de Liebmann é equivalente à abordagem de Gauss-Seidel para a solução de equações algébricas lineares simultâneas.
4. Saber como determinar variáveis secundárias para problemas de campo bidimensionais.
5. Reconhecer a distinção entre as condições de Dirichlet e as condições de contorno.
6. Entender como usar fatores de peso para incorporar fronteiras irregulares nos esquemas por diferenças finitas para EDPs.
7. Entender como implementar a abordagem por volume de controle para implementar soluções numéricas de EDPs.
8. Saber a diferença entre convergência e estabilidade de uma EDP parabólica.
9. Compreender a diferença entre os esquemas explícitos e implícitos para resolver EDPs parabólicas.
10. Reconhecer como o critério de estabilidade para métodos explícitos diminuem sua utilidade para resolver EDPs parabólicas.
11. Saber como interpretar moléculas computacionais.
12. Reconhecer como a abordagem ADI atinge alta eficiência na resolução de equações parabólicas em duas dimensões espaciais.
13. Entender a diferença entre o método direto e o método de resíduos com peso na dedução das equações de elementos.
14. Saber como implementar o método de Galerkin.
15. Entender os benefícios da integração por partes durante a dedução das equações de elementos; em particular, reconhecer as implicações de abaixar a derivada mais alta de uma segunda para uma primeira derivada.

Por fim, uma de nossas metas mais importantes deveria ser dominar diversos dos pacotes de software de propósito geral que estão amplamente disponíveis. Em particular, você deveria se tornar adepto do uso de tais ferramentas para implementar os métodos numéricos na resolução de problemas de engenharia.

CAPÍTULO 29

Diferenças Finitas: Equações Elípticas

As equações elípticas na engenharia são usadas tipicamente para caracterizar problemas de contorno estacionários. Antes de mostrar como eles podem ser resolvidos, ilustraremos como um caso simples — a equação de Laplace — é deduzido de um contexto de problema físico.

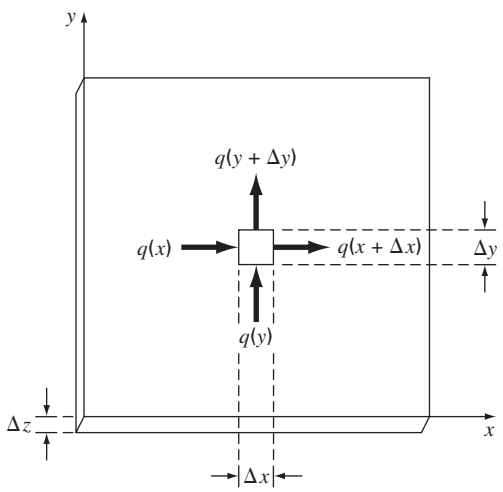
29.1 A EQUAÇÃO DE LAPLACE

Como mencionado na introdução desta parte do livro, a equação de Laplace pode ser usada para modelar diversos problemas envolvendo o potencial de uma variável desconhecida. Por causa de sua simplicidade e da relevância geral para a maioria das áreas da engenharia, vamos usar uma placa aquecida como nosso contexto fundamental na dedução e resolução dessa EDP elíptica. Lições de casa e aplicações na engenharia (Capítulo 32) serão utilizadas para ilustrar a aplicabilidade do modelo a outros contextos de problemas de engenharia.

A Figura 29.1 mostra um elemento na face de uma placa retangular fina de espessura Δz. A placa é isolada em toda parte, exceto nas bordas, onde a temperatura pode ser fixada em um nível prescrito. O isolamento e o fato de a placa ser fina significam que a transferência de calor é limitada às direções x e y. No estado estacionário, o fluxo de calor para dentro do elemento em um período unitário Δt deve ser igual ao fluxo para fora, como em

$$q(x)\,\Delta y\,\Delta z\,\Delta t + q(y)\,\Delta x\,\Delta z\,\Delta t = q(x+\Delta x)\,\Delta y\,\Delta z\,\Delta t + q(y+\Delta y)\Delta x\,\Delta z\,\Delta t \tag{29.1}$$

FIGURA 29.1
Uma placa fina de espessura Δz. É mostrado um elemento no qual é feito o balanço de calor.



onde $q(x)$ e $q(y)$ são os fluxos de calor nas direções x e y, respectivamente [cal/(cm$^2 \cdot$ s)]. Dividindo por Δz e Δt e agrupando os termos, obtemos

$$[q(x) - q(x + \Delta x)]\Delta y + [q(y) - q(y + \Delta y)]\Delta x = 0$$

Multiplicando o primeiro termo por $\Delta x/\Delta x$ e o segundo por $\Delta y/\Delta y$, obtemos

$$\frac{q(x) - q(x + \Delta x)}{\Delta x} \Delta x \, \Delta y + \frac{q(y) - q(y + \Delta y)}{\Delta y} \Delta y \, \Delta x = 0 \tag{29.2}$$

Dividindo por $\Delta x \, \Delta y$ e tomando o limite, resulta em

$$-\frac{\partial q}{\partial x} - \frac{\partial q}{\partial y} = 0 \tag{29.3}$$

onde as derivadas parciais vêm da definição nas Equações (PT7.1) e (PT7.2).

A Equação (29.3) é uma equação diferencial parcial que é uma expressão da conservação da energia na placa. Entretanto, a menos que os fluxos de calor sejam especificados nas bordas da placa, ela não pode ser resolvida. Como as condições de contorno para a temperatura são dadas, a Equação (29.3) deve ser reformulada em termos da temperatura. A ligação entre o fluxo e a temperatura é fornecida pela *lei de condução de calor de Fourier*, que pode ser representada por:

$$q_i = -k\rho C \frac{\partial T}{\partial i} \tag{29.4}$$

onde q_i é o fluxo de calor na direção da dimensão i [cal/(cm$^2 \cdot$ s)], k é o *coeficiente de difusividade térmica* (cm^2/s), ρ é a densidade do material (g/cm^3), C é a capacidade calorífica do material [cal/(g $\cdot$ °C)] e T é a temperatura, que é definida por

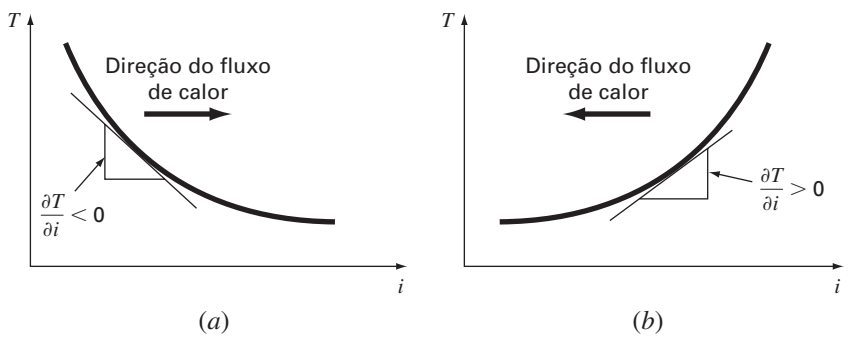
$$T = \frac{H}{\rho C V}$$

onde H é o calor (cal) e V é o volume (cm^3). Às vezes, o termo na frente da derivada na Equação (29.3) é tratado como um único termo,

$$k' = k\rho C \tag{29.5}$$

onde k' é conhecido como *coeficiente de condutividade térmica* [cal/(s $\cdot$ cm $\cdot$ °C)]. Em ambos os casos, k e k' são parâmetros que refletem quão bem o material conduz o calor.

FIGURA 29.2
Descrição gráfica de um gradiente de temperatura. Como o calor se move "para baixo" de altas a baixas temperaturas, o fluxo em (*a*) é da esquerda para a direita na direção *i* positiva. Entretanto, em virtude da orientação das coordenadas cartesianas, a inclinação é negativa nesse caso. Logo, um gradiente negativo leva a um fluxo positivo. Essa é a origem do sinal de menos na lei de condução de calor de Fourier. O caso inverso é descrito em (*b*), no qual um gradiente positivo leva a um fluxo de calor negativo da direita para a esquerda.



(*a*) (*b*)

A lei de Fourier é chamada algumas vezes *equação constitutiva*. Ela recebe essa designação porque fornece um mecanismo que define as interações internas do sistema. A inspeção da Equação (29.4) indica que a lei de Fourier especifica que o fluxo de calor perpendicular ao eixo i é proporcional ao gradiente ou inclinação da temperatura na direção i. O sinal negativo garante que um fluxo positivo na direção i resulta de uma inclinação negativa de alta para baixa temperatura (Figura 29.2). A substituição da Equação (29.4) na Equação (29.3) resulta em

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0 \tag{29.6}$$

que é a *equação de Laplace*. Observe que para o caso no qual existem fontes ou sumidouros de calor no domínio bidimensional, a equação pode ser representada por

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = f(x, y) \tag{29.7}$$

onde $f(x, y)$ é a função descrevendo as fontes ou sumidouros de calor. A Equação (29.7) é conhecida como *equação de Poisson*.

29.2 TÉCNICAS DE SOLUÇÕES

A solução numérica de EDPs elípticas, tais como a equação de Laplace, ocorre de maneira inversa à dedução da Equação (29.6) da seção anterior. Lembre-se de que a dedução da Equação (29.6) usou um balanço em um elemento discreto para produzir uma equação de diferença algébrica que caracterizasse o fluxo de calor em uma placa. Tomar limites transformou essa equação de diferença em uma equação diferencial [Equação (29.3)].

Para a solução numérica, representações por diferenças finitas, baseadas no tratamento da placa como uma grade de pontos discretos (Figura 29.3), substituem as derivadas parciais na Equação (29.6). Como descrito a seguir, a EDP é transformada em uma equação de diferença algébrica.

29.2.1 A Equação de Diferença de Laplace

As diferenças centradas baseadas no esquema de grade da Figura 29.3 são (lembre-se da Figura 23.3)

$$\frac{\partial^2 T}{\partial x^2} = \frac{T_{i+1,j} - 2T_{i,j} + T_{i-1,j}}{\Delta x^2}$$

e

$$\frac{\partial^2 T}{\partial y^2} = \frac{T_{i,j+1} - 2T_{i,j} + T_{i,j-1}}{\Delta y^2}$$

que têm erros de $O[\Delta(x)^2]$ e $O[\Delta(y)^2]$, respectivamente. Substituindo essas expressões na Equação (29.6), obtemos

$$\frac{T_{i+1,j} - 2T_{i,j} + T_{i-1,j}}{\Delta x^2} + \frac{T_{i,j+1} - 2T_{i,j} + T_{i,j-1}}{\Delta y^2} = 0$$

Para a grade quadrada da Figura 29.3, $\Delta x = \Delta y$, e, agrupando os termos, a equação se torna

$$T_{i+1,j} + T_{i-1,j} + T_{i,j+1} + T_{i,j-1} - 4T_{i,j} = 0 \tag{29.8}$$

Essa relação, que é válida para todos os pontos interiores da placa, é conhecida como *equação de diferença de Laplace*.

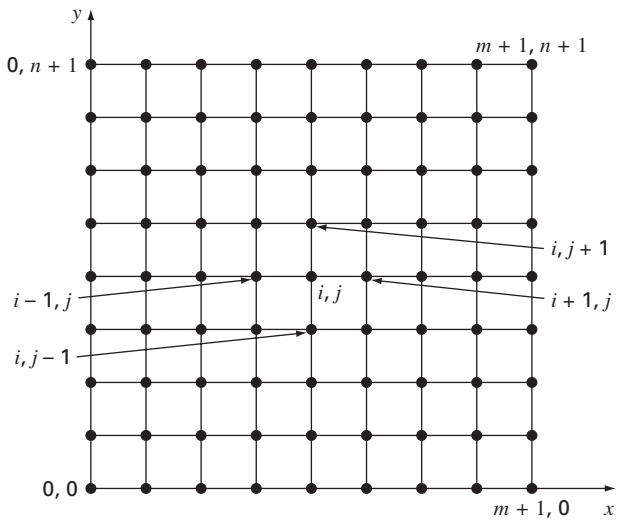


FIGURA 29.3
Uma grade usada para a solução por diferenças finitas de EDPs elípticas em duas variáveis independentes, como a equação de Laplace.

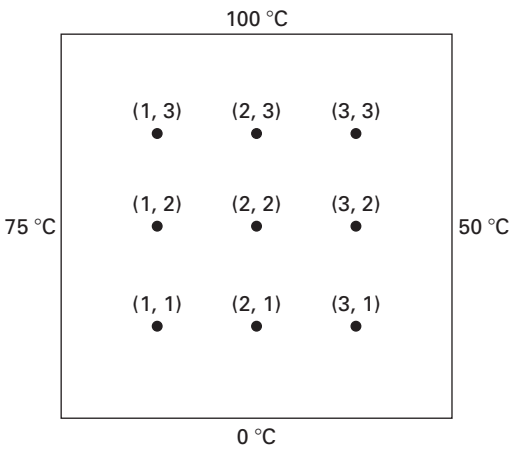
Além disso, devem ser especificadas condições de contorno ao longo das bordas da placa para se obter uma única solução. O caso mais simples é quando a temperatura nas bordas é mantida em valores fixos. Isso é chamado *condição de contorno de Dirichlet*. Esse é o caso na Figura 29.4, na qual as bordas são mantidas a temperaturas constantes. Para o caso ilustrado na Figura 29.4, um balanço para o nó (1, 1) é, de acordo com a Equação (29.8),

$$T_{21} + T_{01} + T_{12} + T_{10} - 4T_{11} = 0 \tag{29.9}$$

Entretanto, $T_{01} = 75$ e $T_{10} = 0$, e, portanto, a Equação (29.9) pode ser expressa como

$$-4T_{11} + T_{12} + T_{21} = -75$$

FIGURA 29.4
Uma placa aquecida na qual as temperaturas nas bordas são mantidas em níveis constantes. Esse caso é chamado condição de contorno de Dirichlet.



Equações análogas podem ser deduzidas para os outros pontos interiores. O resultado é o seguinte conjunto de nove equações simultâneas com nove incógnitas:

$$\begin{array}{rrrrrrrrrr}
4T_{11} & -T_{21} & & -T_{12} & & & & & & = 75 \\
-T_{11} & +4T_{21} & -T_{31} & & -T_{22} & & & & & = 0 \\
 & -T_{21} & +4T_{31} & & & -T_{32} & & & & = 50 \\
-T_{11} & & & +4T_{12} & -T_{22} & & -T_{13} & & & = 75 \\
 & -T_{21} & & -T_{12} & +4T_{22} & -T_{32} & & -T_{23} & & = 0 \\
 & & -T_{31} & & -T_{22} & +4T_{32} & & & -T_{33} & = 50 \\
 & & & -T_{12} & & & +4T_{13} & -T_{23} & & = 175 \\
 & & & & -T_{22} & & -T_{13} & +4T_{23} & -T_{33} & = 100 \\
 & & & & & -T_{32} & & -T_{23} & +4T_{33} & = 150
\end{array} \tag{29.10}$$

29.2.2 O método de Liebmann

A maioria das soluções numéricas da equação de Laplace envolve sistemas que são muito maiores do que o da Equação (29.10). Por exemplo, uma grade 10 por 10 envolve 100 equações algébricas lineares. Técnicas de solução para esse tipo de equações foram discutidas na Parte Três.

Observe que existe um máximo de cinco termos desconhecidos por linha na Equação (29.10). Para grades de tamanho maior, isso quer dizer que um número significativo de termos será zero. Quando aplicados a tais sistemas esparsos, métodos de eliminação para a matriz toda desperdiçam uma grande quantidade de memória computacional armazenando esses zeros. Por essa razão, os métodos aproximados fornecem uma abordagem viável para obter soluções de equações elípticas. A abordagem mais comumente usada é a do método de *Gauss-Seidel*, que, quando aplicado a EDPs, também é conhecido como *método de Liebmann*. Nessa técnica, a Equação (29.8) é expressa por

$$T_{i,j} = \frac{T_{i+1,j} + T_{i-1,j} + T_{i,j+1} + T_{i,j-1}}{4} \tag{29.11}$$

e resolvida iterativamente de $j = 1$ a n e de $i = 1$ a m. Como a Equação (29.8) é diagonal dominante, esse procedimento eventualmente convergirá para uma solução estável (lembre-se da Seção 11.2.1). A sobre-relaxação algumas vezes é usada para acelerar a taxa de convergência por meio da aplicação da seguinte fórmula, depois de cada iteração:

$$T_{i,j}^{\text{novo}} = \lambda T_{i,j}^{\text{novo}} + (1 - \lambda) T_{i,j}^{\text{velho}} \tag{29.12}$$

onde $T_{i,j}^{\text{novo}}$ e $T_{i,j}^{\text{velho}}$ são os valores de $T_{i,j}$ da iteração atual e da anterior, respectivamente e λ é um fator de peso que é fixado entre 1 e 2.

Como no método de Gauss-Seidel convencional, as iterações são repetidas até que os valores absolutos de todos os erros relativos porcentuais $(\varepsilon_a)_{i,j}$ caiam abaixo de um critério de parada pré-especificado ε_s. São obtidas estimativas para esses erros relativos porcentuais por

$$|(\varepsilon_a)_{i,j}| = \left| \frac{T_{i,j}^{\text{novo}} - T_{i,j}^{\text{velho}}}{T_{i,j}^{\text{novo}}} \right| 100\% \tag{29.13}$$

EXEMPLO 29.1 Temperatura de uma Placa Aquecida com Condição de Contorno Fixa

Enunciado do Problema. Use o método de Liebmann (Gauss-Seidel) para determinar a temperatura da placa aquecida na Figura 29.4. Use sobre-relaxação com um valor de 1,5 para o fator de peso e itere até $\varepsilon_s = 1\%$.

Solução. A Equação (29.11) em $i = 1, j = 1$ é

$$T_{11} = \frac{0 + 75 + 0 + 0}{4} = 18{,}75$$

e aplicando sobre-relaxação, obtemos

$$T_{11} = 1{,}5(18{,}75) + (1 - 1{,}5)0 = 28{,}125$$

Para $i = 2, j = 1$,

$$T_{21} = \frac{0 + 28{,}125 + 0 + 0}{4} = 7{,}03125$$

$$T_{21} = 1{,}5(7{,}03125) + (1 - 1{,}5)0 = 10{,}54688$$

Para $i = 3, j = 1$,

$$T_{31} = \frac{50 + 10{,}54688 + 0 + 0}{4} = 15{,}13672$$

$$T_{31} = 1{,}5(15{,}13672) + (1 - 1{,}5)0 = 22{,}70508$$

Os cálculos são repetidos para as outras linhas, o que resulta em

$$T_{12} = 38{,}67188 \qquad T_{22} = 18{,}45703 \qquad T_{32} = 34{,}18579$$
$$T_{13} = 80{,}12696 \qquad T_{23} = 74{,}46900 \qquad T_{33} = 96{,}99554$$

Como todos os $T_{i,j}$'s são inicialmente nulos, todos os ε_a's para a primeira iteração serão 100%.

Para a segunda iteração os resultados são

$$T_{11} = 32{,}51953 \qquad T_{21} = 22{,}35718 \qquad T_{31} = 28{,}60108$$
$$T_{12} = 57{,}95288 \qquad T_{22} = 61{,}63333 \qquad T_{32} = 71{,}86833$$
$$T_{13} = 75{,}21973 \qquad T_{23} = 87{,}95872 \qquad T_{33} = 67{,}68736$$

A estimativa do erro para $T_{1,1}$ pode ser obtida de [Equação (29.13)]

$$|(\varepsilon_a)_{1,1}| = \left|\frac{32{,}51953 - 28{,}12500}{32{,}51953}\right| 100\% = 13{,}5\%$$

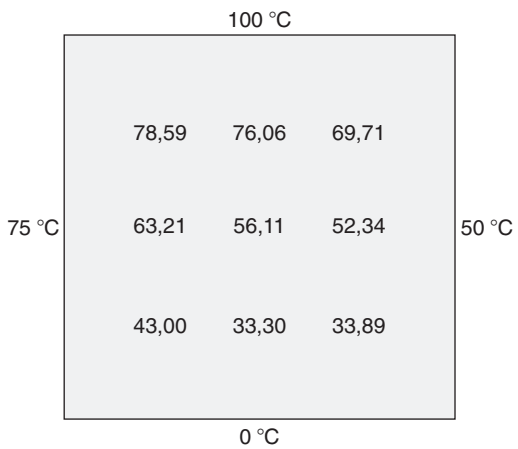
Como esse valor está acima do critério de parada de 1%, continua-se com os cálculos. A nona iteração dá os resultados

$$T_{11} = 43{,}00061 \qquad T_{21} = 33{,}29755 \qquad T_{31} = 33{,}88506$$
$$T_{12} = 63{,}21152 \qquad T_{22} = 56{,}11238 \qquad T_{32} = 52{,}33999$$
$$T_{13} = 78{,}58718 \qquad T_{23} = 76{,}06402 \qquad T_{33} = 69{,}71050$$

nos quais o erro máximo é 0,71%.

A Figura 29.5 mostra os resultados. Como esperado, é estabelecido um gradiente de calor das temperaturas altas para as baixas.

FIGURA 29.5
Distribuição de temperatura em uma placa aquecida sujeita a condições de contorno fixas.



29.2.3 Variáveis Secundárias

Como sua distribuição é descrita pela equação de Laplace, a temperatura é considerada a variável primária no problema da placa aquecida. Nesse caso, bem como para outros problemas envolvendo EDPs, variáveis secundárias também podem ter interesse. Na realidade, em certos contextos de engenharia, as variáveis secundárias podem de fato ser mais importantes.

Para a placa aquecida, uma variável secundária é a taxa de fluxo de calor através da superfície da placa. Essa quantidade pode ser calculada a partir da lei de Fourier. Aproximações por diferenças finitas centrais para as primeiras derivadas (lembre-se da Figura 23.3) podem ser substituídas na Equação (29.4) para fornecer os seguintes valores para o fluxo de calor nas dimensões x e y:

$$q_x = -k' \frac{T_{i+1,j} - T_{i-1,j}}{2\,\Delta x} \tag{29.14}$$

e

$$q_y = -k' \frac{T_{i,j+1} - T_{i,j-1}}{2\,\Delta y} \tag{29.15}$$

O fluxo de calor resultante pode ser calculado a partir dessas duas quantidades por

$$q_n = \sqrt{q_x^2 + q_y^2} \tag{29.16}$$

onde a direção de q_n é dada por

$$\theta = \text{tg}^{-1}\left(\frac{q_y}{q_x}\right) \tag{29.17}$$

para $q_x > 0$ e

$$\theta = \text{tg}^{-1}\left(\frac{q_y}{q_x}\right) + \pi \tag{29.18}$$

para $q_x < 0$. Lembre-se de que o ângulo pode ser expresso em graus multiplicando-o por $180^\circ/\pi$. Se $q_x = 0$, θ é $\pi/2$ (90°) ou $3\pi/2$ (270°), dependendo de q_y ser positivo ou negativo, respectivamente.

EXEMPLO 29.2 Distribuição do Fluxo para uma Placa Aquecida

Enunciado do Problema. Use os resultados do Exemplo 29.1 para determinar a distribuição de fluxo de calor na placa aquecida da Figura 29.4. Suponha que a placa tenha 40×40 cm e seja feita de alumínio [$k' = 0{,}49$ cal/(s · cm · °C)].

Solução. Para $i = j = 1$, a Equação (29.14) pode ser usada para calcular

$$q_x = -0{,}49 \frac{\text{cal}}{\text{s} \cdot \text{cm} \cdot {}^\circ\text{C}} \frac{(33{,}29755 - 75)^\circ\text{C}}{2(10\text{ cm})} = 1{,}022 \text{ cal/(cm}^2 \cdot \text{s)}$$

e [Eq. (29.15)]

$$q_y = -0{,}49 \frac{\text{cal}}{\text{s} \cdot \text{cm} \cdot {}^\circ\text{C}} \frac{(63{,}21152 - 0)^\circ\text{C}}{2(10\text{ cm})} = -1{,}549 \text{ cal/(cm}^2 \cdot \text{s)}$$

O fluxo resultante pode ser calculado com a Equação (29.16):

$$q_n = \sqrt{(1{,}022)^2 + (-1{,}549)^2} = 1{,}856 \text{ cal/(cm}^2 \cdot \text{s)}$$

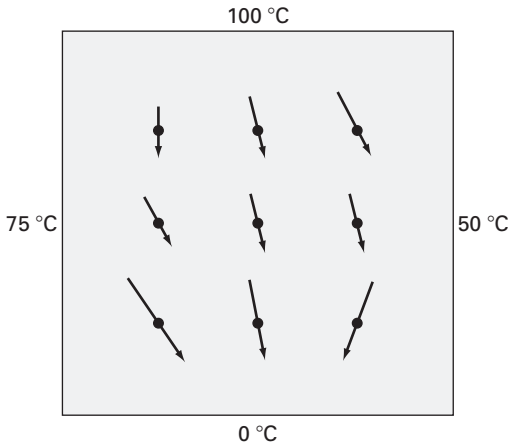


FIGURA 29.6
Fluxo de calor de uma placa sujeita a temperaturas fixas na borda. Observe que os comprimentos das setas são proporcionais ao módulo do campo.

e o ângulo de sua trajetória com a Equação (29.17)

$$\theta = \text{tg}^{-1}\left(\frac{-1{,}549}{1{,}022}\right) = -0{,}98758 \times \frac{180^\circ}{\pi} = -56{,}584^\circ$$

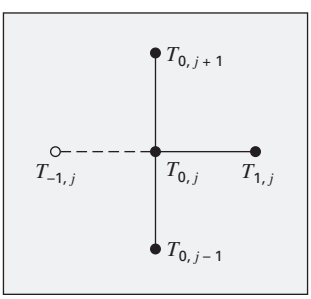
Assim, nesse ponto, o fluxo de calor aponta para baixo e para a direita. Os valores nos outros pontos da grade podem ser calculados; os resultados estão mostrados na Figura 29.6.

29.3 CONDIÇÕES DE CONTORNO

Como os fatores complicadores estão ausentes, a placa retangular com condições de contorno fixas foi um contexto ideal para mostrar como as EDPs elípticas podem ser resolvidas numericamente. Agora, consideraremos outras questões que vão expandir nossos recursos para tratar problemas mais realísticos. Estes envolvem bordas nas quais as derivadas são especificadas e bordas que têm forma irregular.

29.3.1 Condições de Contorno na Derivada

FIGURA 29.7
Um nó $(0, j)$ na borda, na aresta esquerda da placa aquecida. Para obter uma aproximação da derivada normal à aresta (isto é, a derivada na direção x), um ponto imaginário $(-1, j)$ é posicionado a uma distância Δx além da aresta.



A condição de contorno de Dirichlet ou fixa, discutida até este ponto, é apenas uma dentre os diversos tipos que são utilizados em equações diferenciais parciais. Uma alternativa comum é o caso em que a derivada é dada. Isso em geral é conhecido como *condição de contorno de Neumann*. Para o problema da placa aquecida, isso quer dizer especificar o fluxo de calor em vez da temperatura na borda. Um exemplo é a situação na qual a borda é isolada. Nesse caso, o qual é conhecido como uma *condição de contorno natural*, a derivada é nula. Essa conclusão é tirada diretamente da Equação (29.4), pois isolar a fronteira significa que o fluxo de calor (e, conseqüentemente, o gradiente) deve ser nulo. Outro exemplo seria quando se perde calor pela aresta por mecanismos previsíveis tais como radiação e condução.

A Figura 29.7 descreve um nó $(0, j)$ na aresta esquerda da placa aquecida. Aplicando-se a Equação (29.8) a esse ponto, obtemos

$$T_{1,j} + T_{-1,j} + T_{0,j+1} + T_{0,j-1} - 4T_{0,j} = 0 \tag{29.19}$$

Observe que um ponto imaginário $(-1, j)$, localizado fora da placa, é necessário nessa equação. Embora esse ponto exterior fictício possa parecer representar um problema, na realidade, serve como veículo para incorporar a condição de contorno na derivada ao problema. Isso é feito representando-se a primeira derivada na dimensão x em $(0, j)$ por uma diferença dividida finita

$$\frac{\partial T}{\partial x} \cong \frac{T_{1,j} - T_{-1,j}}{2\,\Delta x}$$

que pode ser reescrita como

$$T_{-1,j} = T_{1,j} - 2\,\Delta x \frac{\partial T}{\partial x}$$

Agora, temos uma relação para $T_{-1,j}$ que realmente inclui a derivada. Ela pode ser substituída na Equação (29.19) para fornecer

$$2T_{1,j} - 2\,\Delta x \frac{\partial T}{\partial x} + T_{0,j+1} + T_{0,j-1} - 4T_{0,j} = 0 \tag{29.20}$$

Assim, incorporamos a derivada no balanço.

Relações análogas podem ser desenvolvidas para condições de contorno na derivada nas outras arestas. O seguinte exemplo mostra como isso é feito para a placa aquecida.

EXEMPLO 29.3 Placa Aquecida com uma Aresta Isolada

Enunciado do Problema. Repita o mesmo problema que no Exemplo 29.1, mas com a aresta inferior isolada.

Solução. A equação geral para caracterizar a derivada na aresta inferior (isto é, em $j = 0$) da placa aquecida é

$$T_{i+1,0} + T_{i-1,0} + 2T_{i,1} - 2\,\Delta y \frac{\partial T}{\partial y} - 4T_{i,0} = 0$$

Para uma aresta isolada, a derivada é nula e a equação se torna

$$T_{i+1,0} + T_{i-1,0} + 2T_{i,1} - 4T_{i,0} = 0$$

As equações simultâneas para a distribuição de temperaturas na placa da Figura 29.4 com a aresta inferior isolada podem ser escritas na forma matricial como

$$\begin{bmatrix}
4 & -1 & & -2 & & & & & & & & \\
-1 & 4 & -1 & & -2 & & & & & & & \\
 & -1 & 4 & & & -2 & & & & & & \\
-1 & & & 4 & -1 & & -1 & & & & & \\
 & -1 & & -1 & 4 & -1 & & -1 & & & & \\
 & & -1 & & -1 & 4 & & & -1 & & & \\
 & & & -1 & & & 4 & -1 & & -1 & & \\
 & & & & -1 & & -1 & 4 & -1 & & -1 & \\
 & & & & & -1 & & -1 & 4 & & & -1 \\
 & & & & & & -1 & & & 4 & -1 & \\
 & & & & & & & -1 & & -1 & 4 & -1 \\
 & & & & & & & & -1 & & -1 & 4
\end{bmatrix}
\begin{Bmatrix} T_{10} \\ T_{20} \\ T_{30} \\ T_{11} \\ T_{21} \\ T_{31} \\ T_{12} \\ T_{22} \\ T_{32} \\ T_{13} \\ T_{23} \\ T_{33} \end{Bmatrix}
=
\begin{Bmatrix} 75 \\ 0 \\ 50 \\ 75 \\ 0 \\ 50 \\ 75 \\ 0 \\ 50 \\ 175 \\ 100 \\ 150 \end{Bmatrix}$$

Observe que, por causa da condição de contorno na derivada, a matriz foi aumentada para 12×12 em contraste com o sistema 9×9 na Equação (29.10) para levar em conta as três temperaturas desconhecidas ao longo da aresta inferior da placa. Essas equações podem ser resolvidas por

$$\begin{aligned}
T_{10} &= 71{,}91 \qquad T_{20} = 67{,}01 \qquad T_{30} = 59{,}54 \\
T_{11} &= 72{,}81 \qquad T_{21} = 68{,}31 \qquad T_{31} = 60{,}57 \\
T_{12} &= 76{,}01 \qquad T_{22} = 72{,}84 \qquad T_{32} = 64{,}42 \\
T_{13} &= 83{,}41 \qquad T_{23} = 82{,}63 \qquad T_{33} = 74{,}26
\end{aligned}$$

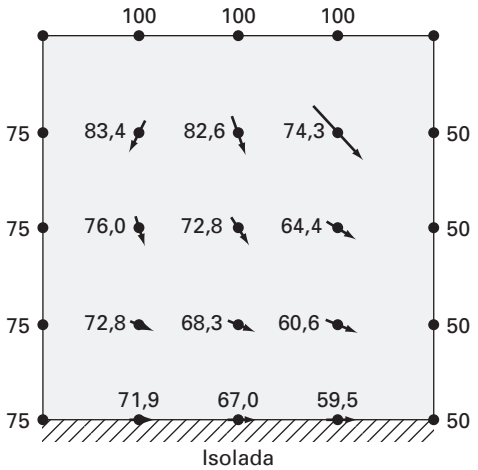


FIGURA 29.8
Distribuição de temperatura e de fluxo para uma placa aquecida sujeita a condições de contorno fixas, exceto por uma aresta inferior isolada.

Esses resultados e o fluxo calculado (para os mesmos parâmetros que no Exemplo 29.2) estão mostrados na Figura 29.8. Observe que, como a aresta inferior está isolada, a temperatura da placa é maior que na Figura 29.5, na qual a temperatura da aresta de baixo é fixada em zero. Além disso, o fluxo de calor (em contraste com a Figura 29.6) é agora dirigido para a direita e se move em paralelo à parede isolada.

29.3.2 Fronteiras Irregulares

Embora a placa retangular da Figura 29.4 tenha servido bem para ilustrar os aspectos fundamentais na resolução de EDPs elípticas, muitos problemas de engenharia não exibem tal geometria idealizada. Por exemplo, um grande número de sistemas tem fronteiras irregulares (Figura 29.9).

A Figura 29.9 é um sistema que serve para ilustrar como bordas não-retangulares podem ser tratadas. Como descrito, a borda esquerda inferior da placa é circular. Observe que associamos parâmetros — α_1, α_2, β_1, β_2 — a cada um dos comprimentos em torno do nó. É claro que, para a placa descrita na Figura 29.9, $\alpha_2 = \beta_2 = 1$. Entretanto, vamos manter esses parâmetros por toda a dedução a seguir, de modo que a equação resultante seja aplicável em geral, para qualquer borda irregular — e não apenas para o canto esquerdo inferior de uma placa aquecida. As primeiras derivadas na dimensão x podem ser aproximadas por

$$\left(\frac{\partial T}{\partial x}\right)_{i-1,i} \cong \frac{T_{i,j} - T_{i-1,j}}{\alpha_1 \, \Delta x} \tag{29.21}$$

e

$$\left(\frac{\partial T}{\partial x}\right)_{i,i+1} \cong \frac{T_{i+1,j} - T_{i,j}}{\alpha_2 \, \Delta x} \tag{29.22}$$

As segundas derivadas podem ser desenvolvidas a partir dessas primeiras derivadas. Para a dimensão x, a segunda derivada é

$$\frac{\partial^2 T}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial T}{\partial x}\right) = \frac{\left(\dfrac{\partial T}{\partial x}\right)_{i,i+1} - \left(\dfrac{\partial T}{\partial x}\right)_{i-1,i}}{\dfrac{\alpha_1 \, \Delta x + \alpha_2 \, \Delta x}{2}} \tag{29.23}$$

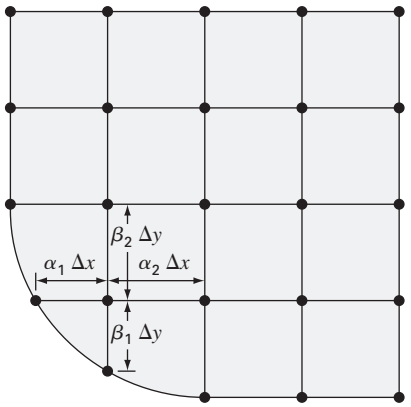


FIGURA 29.9
Uma grade para uma placa aquecida com a borda com formato irregular. Observe como os coeficientes de peso são usados para levar em conta o espaçamento não-uniforme na vizinhança da borda não-retangular.

Substituindo as Equações (29.21) e (29.22) na Equação (29.23), obtemos

$$\frac{\partial^2 T}{\partial x^2} = 2\frac{\dfrac{T_{i-1,j} - T_{i,j}}{\alpha_1\ \Delta x} - \dfrac{T_{i+1,j} - T_{i,j}}{\alpha_2\ \Delta x}}{\alpha_1\ \Delta x + \alpha_2\ \Delta x}$$

Agrupando os termos vem

$$\frac{\partial^2 T}{\partial x^2} = \frac{2}{\Delta x^2}\left[\frac{T_{i-1,j} - T_{i,j}}{\alpha_1(\alpha_1 + \alpha_2)} + \frac{T_{i+1,j} - T_{i,j}}{\alpha_2(\alpha_1 + \alpha_2)}\right]$$

Uma equação análoga pode ser desenvolvida para a dimensão y:

$$\frac{\partial^2 T}{\partial y^2} = \frac{2}{\Delta y^2}\left[\frac{T_{i,j-1} - T_{i,j}}{\beta_1(\beta_1 + \beta_2)} + \frac{T_{i,j+1} - T_{i,j}}{\beta_2(\beta_1 + \beta_2)}\right]$$

Substituindo essas equações na Equação (29.6), temos

$$\frac{2}{\Delta x^2}\left[\frac{T_{i-1,j} - T_{i,j}}{\alpha_1(\alpha_1 + \alpha_2)} + \frac{T_{i+1,j} - T_{i,j}}{\alpha_2(\alpha_1 + \alpha_2)}\right] + \frac{2}{\Delta y^2}\left[\frac{T_{i,j-1} - T_{i,j}}{\beta_1(\beta_1 + \beta_2)} + \frac{T_{i,j+1} - T_{i,j}}{\beta_2(\beta_1 + \beta_2)}\right] = 0 \tag{29.24}$$

Como ilustrado no exemplo a seguir, a Equação (29.24) pode ser aplicada a qualquer nó que seja adjacente a uma fronteira irregular, do tipo Dirichlet.

EXEMPLO 29.4

Placa Aquecida com Fronteira Irregular

Enunciado do Problema. Repita o mesmo problema que no Exemplo 29.1, mas com a borda inferior como descrito na Figura 29.9.

Solução. Para o caso na Figura 29.9, $\Delta x = \Delta y$, $\alpha_1 = \beta_1 = 0{,}732$ e $\alpha_2 = \beta_2 = 1$. Substituindo-se esses valores na Equação (29.24), obtemos o seguinte balanço para o nó (1, 1):

$$0{,}788675(T_{01} - T_{11}) + 0{,}57735(T_{21} - T_{11}) + 0{,}788675(T_{10} - T_{11}) + 0{,}57735(T_{12} - T_{11}) = 0$$

Agrupando os termos, podemos expressar essa equação como

$$-4T_{11} + 0{,}8453T_{21} + 0{,}8453T_{12} = -1{,}1547T_{01} - 1{,}1547T_{10}$$

As equações simultâneas para a distribuição de temperaturas na placa na Figura 29.9 com a temperatura na borda inferior de 75 podem ser escritas na forma matricial como

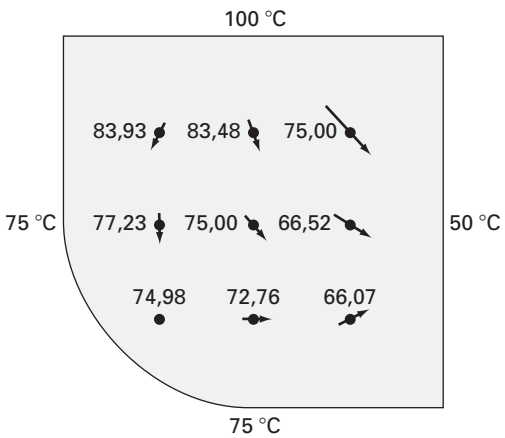
$$\begin{bmatrix} 4 & -0{,}845 & & -0{,}845 & & & & & \\ -1 & 4 & -1 & & -1 & & & & \\ & -1 & 4 & & & -1 & & & \\ -1 & & & 4 & -1 & & -1 & & \\ & -1 & & -1 & 4 & -1 & & -1 & \\ & & -1 & & -1 & 4 & & & -1 \\ & & & -1 & & & 4 & -1 & \\ & & & & -1 & & -1 & 4 & -1 \\ & & & & & -1 & & -1 & -4 \end{bmatrix} \begin{Bmatrix} T_{11} \\ T_{21} \\ T_{31} \\ T_{12} \\ T_{22} \\ T_{32} \\ T_{13} \\ T_{23} \\ T_{33} \end{Bmatrix} = \begin{Bmatrix} 173{,}2 \\ 75 \\ 125 \\ 75 \\ 0 \\ 50 \\ 175 \\ 100 \\ 150 \end{Bmatrix}$$

As soluções dessas equações são

$$\begin{aligned} T_{11} &= 74{,}98 & T_{21} &= 72{,}76 & T_{31} &= 66{,}07 \\ T_{12} &= 77{,}23 & T_{22} &= 75{,}00 & T_{32} &= 66{,}52 \\ T_{13} &= 83{,}93 & T_{23} &= 83{,}48 & T_{33} &= 75{,}00 \end{aligned}$$

Esses resultados, juntamente com os fluxos calculados, estão mostrados na Figura 29.10. Observe que os fluxos foram calculados da mesma forma que na Seção 29.2.3, com a exceção de que $(\alpha_1 + \alpha_2)$ e $(\beta_1 + \beta_2)$ substituíram os 2's no denominador das Equações (29.14) e (29.15), respectivamente. A Seção 32.3 ilustra como isso é feito.

FIGURA 29.10
Distribuição de temperatura e fluxo em uma placa aquecida com uma borda circular.



Condições nas derivadas em bordas com formato irregular são mais difíceis de formular. A Figura 29.11 mostra um ponto próximo a uma borda irregular em que a derivada normal é especificada.

A derivada normal no nó 3 pode ser aproximada pelo gradiente entre os nós 1 e 7,

$$\left.\frac{\partial T}{\partial \eta}\right|_3 = \frac{T_1 - T_7}{L_{17}} \tag{29.25}$$

Quando θ for menor que 45°, como mostrado, a distância do nó 7 ao 8 é $\Delta x \operatorname{tg} \theta$, e podemos obter uma estimativa usando interpolação linear

$$T_7 = T_8 + (T_6 - T_8)\frac{\Delta x \operatorname{tg} \theta}{\Delta y}$$

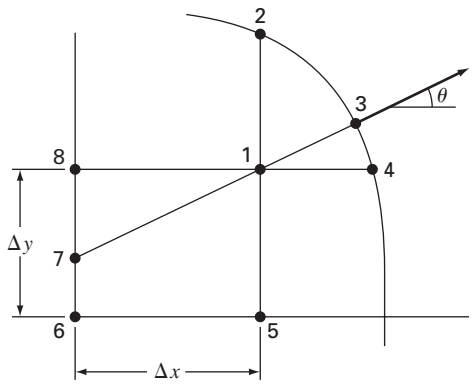


FIGURA 29.11
Uma borda curva na qual o gradiente normal é especificado.

O comprimento L_{17} é igual a $\Delta x/\cos\theta$. Esse comprimento, junto com a aproximação para T_7, pode ser substituído na Equação (29.25) para fornecer

$$T_1 = \left(\frac{\Delta x}{\cos\theta}\right)\left.\frac{\partial T}{\partial \eta}\right|_3 + T_6\frac{\Delta x \text{ tg } \theta}{\Delta y} + T_8\left(1 - \frac{\Delta x \text{ tg } \theta}{\Delta y}\right) \tag{29.26}$$

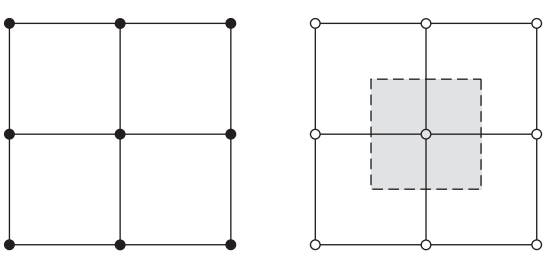
Tal equação fornece uma forma de incorporar o gradiente normal na abordagem por diferença finita. Para casos em que θ for maior que 45° uma equação diferente deveria ser usada. A determinação dessa fórmula será deixada como exercício.

29.4 A ABORDAGEM POR VOLUME DE CONTROLE

Para resumir, a abordagem por diferenças finitas ou por série de Taylor divide um meio contínuo em nós (Figura 29.12*a*). A equação diferencial parcial subjacente é escrita para cada um desses nós. Aproximações por diferenças finitas substituem, então, as derivadas para converter as equações para uma forma algébrica.

Tal abordagem é bastante simples e direta para grades ortogonais (isto é, retangulares) e coeficientes constantes. Entretanto, a abordagem se torna uma tarefa mais difícil para condições nas derivadas em contornos com formato irregular.

FIGURA 29.12
Duas perspectivas diferentes para desenvolver soluções aproximadas de EDPs: (*a*) diferenças finitas e (*b*) volume de controle.



(*a*) Abordagem por diferenças finitas, ponto a ponto

(*b*) Abordagem por volume de controle

A Figura 29.13 é um exemplo de um sistema no qual aparecem dificuldades adicionais. Essa placa é feita de dois materiais diferentes e tem o espaçamento da grade desigual. Além disso, a metade de sua aresta superior está sujeita a transferência de calor por convecção, enquanto metade é isolada. Desenvolver equações para o nó (4, 2) exigiria algumas deduções adicionais além das abordagens desenvolvidas até este ponto.

A *abordagem por volume de controle* (também chamada de *abordagem por integral no volume*) oferece uma maneira alternativa de obter aproximações numéricas para EDPs que é especialmente útil nos casos como o da Figura 29.13. Como na Figura 29.12*b*, a abordagem se parece com a abordagem ponto a ponto pelo fato de que são determinados pontos em todo o domínio. No entanto, em vez de obter uma aproximação da EDP em um ponto, a aproximação é aplicada a um volume que circunda o ponto. Para uma grade ortogonal, o volume é formado por retas perpendiculares passando pelos pontos médios de cada segmento ligando dois nós adjacentes. Um balanço de calor pode então ser desenvolvido para cada volume de forma análoga à Equação (29.1).

Como um exemplo, aplicaremos a abordagem de controle de volume ao nó (4, 2). Primeiro, o volume é definido bissectando os segmentos ligando os nós. Como na Figura 29.14, o volume tem transferência de calor por condução por suas bordas esquerda, direita e inferior e transferência de calor por convecção através de metade de sua borda superior. Observe que a transferência pela borda inferior envolve ambos os materiais.

FIGURA 29.13
Uma placa aquecida com espaçamento da rede desigual, dois materiais e condições de contorno mistas.

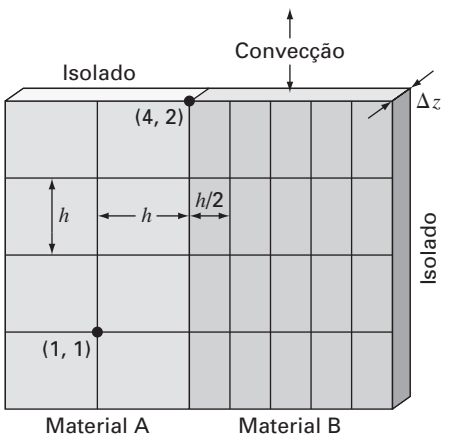
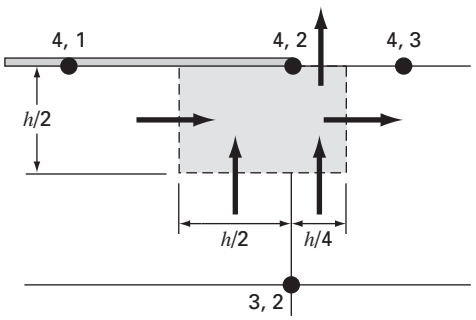


FIGURA 29.14
Um volume de controle para o nó (4, 2) com setas indicando a transferência de calor nas bordas.



Um balanço de calor estacionário para o volume pode ser escrito em termos qualitativos como

$$0 = \begin{pmatrix}\text{condução no}\\ \text{lado esquerdo}\end{pmatrix} - \begin{pmatrix}\text{condução no}\\ \text{lado direito}\end{pmatrix} + \begin{pmatrix}\text{condução na parte}\\ \text{inferior do material “a”}\end{pmatrix} + \begin{pmatrix}\text{condução na parte}\\ \text{inferior do material “b”}\end{pmatrix} - \begin{pmatrix}\text{convecção na}\\ \text{parte superior}\end{pmatrix} \tag{29.27}$$

Agora, a taxa do fluxo por condução pode ser representada pela versão em diferença finita ou pela lei de Fourier. Por exemplo, para o ganho por condução no lado esquerdo, ela seria

$$q = -k'_a \frac{T_{42} - T_{41}}{h}$$

onde q tem unidades de cal/cm^2/s. Essa taxa de fluxo deve então ser multiplicada pela área pela qual ela entra ($\Delta z \times h/2$) para dar a taxa de calor entrando no volume por unidade de tempo:

$$Q = -k'_a \frac{T_{42} - T_{41}}{h} \frac{h}{2} \Delta z$$

onde Q tem unidades de cal/s.

O fluxo de calor devido à convecção pode ser formulado como

$$q = h_c(T_a - T_{42})$$

onde h_c é uma constante de convecção de calor [cal/(s $\cdot$ cm^2 $\cdot$ °C)] e T_a é a temperatura do ar (ºC). Novamente, multiplicando pela área adequada, obtemos a taxa de fluxo de calor por unidade de tempo,

$$Q = h_c(T_a - T_{42}) \frac{h}{4} \Delta z$$

As outras transferências podem ser desenvolvidas de forma análoga e substituídas na Equação (29.27), resultando em

$$0 = \underbrace{-k'_a \frac{T_{42} - T_{41}}{h} \frac{h}{2} \Delta z}_{\text{(condução no lado esquerdo)}} + \underbrace{k'_b \frac{T_{43} - T_{42}}{h} \frac{h}{2} \Delta z}_{\text{(condução no lado direito)}}$$

$$\underbrace{-k'_a \frac{T_{42} - T_{32}}{h} \frac{h}{2} \Delta z}_{\begin{pmatrix}\text{condução na parte}\\ \text{inferior do material “a”}\end{pmatrix}} - \underbrace{k'_b \frac{T_{42} - T_{32}}{h} \frac{h}{4} \Delta z}_{\begin{pmatrix}\text{condução na parte}\\ \text{inferior do material “b”}\end{pmatrix}} + \underbrace{h_c(T_a - T_{42}) \frac{h}{4} \Delta z}_{\text{(convecção acima)}}$$

Os valores dos parâmetros podem ser substituídos para produzir a equação de balanço de calor final. Por exemplo, se $\Delta z = 0{,}5$ cm, $h = 10$ cm, $k'_a = 0{,}3$ cal/(s $\cdot$ cm $\cdot$ °C), $k'_b = 0{,}5$ cal/(s $\cdot$ cm $\cdot$ °C) e $h_c = 0{,}1$ cal/(s $\cdot$ cm^2 $\cdot$ °C), a equação se torna

$$0{,}5875T_{42} - 0{,}075T_{41} - 0{,}25T_{43} - 0{,}1375T_{32} = 2{,}5$$

Para tornar essa equação comparável com o laplaciano padrão, a equação pode ser multiplicada por $4/0{,}5875$, de modo que o coeficiente do nó-base seja 4,

$$4T_{42} - 0{,}510638T_{41} - 1{,}702128T_{43} - 0{,}93617T_{32} = 17{,}02128$$

Nos casos padrão cobertos até este ponto, as abordagens por volume de controle e ponto a ponto por diferenças finitas fornecem resultados idênticos. Por exemplo, para o nó (1, 1) na Figura 29.13, o balanço seria

$$0 = -k'_a \frac{T_{11} - T_{01}}{h} h\ \Delta z + k'_a \frac{T_{21} - T_{11}}{h} h\ \Delta z - k'_a \frac{T_{11} - T_{10}}{h} h\ \Delta z + k'_a \frac{T_{12} - T_{11}}{h} h\ \Delta z$$

o que se simplifica para o laplaciano padrão,

$$0 = 4T_{11} - T_{01} - T_{21} - T_{12} - T_{10}$$

Olharemos para outros casos padrão (por exemplo, condição de contorno na derivada) e exploraremos a abordagem por volume de controle com detalhes adicionais nos problemas no final deste capítulo.

29.5 SOFTWARE PARA RESOLVER EQUAÇÕES ELÍPTICAS

Modificar um programa computacional para incluir condição de contorno na derivada para sistemas retangulares é uma tarefa relativamente simples. Envolve simplesmente garantir que sejam geradas equações adicionais para caracterizar os nós na borda, nos quais as derivadas são especificadas. Além disso, o código precisa ser modificado de modo que essas equações incorporem a derivada como foi visto na Equação (29.20).

Desenvolver um software geral para caracterizar sistemas com bordas irregulares é uma tarefa muito mais difícil. Por exemplo, um algoritmo bastante complicado seria necessário para modelar a junta de vedação simples descrita na Figura 29.15. Isso envolveria duas grandes modificações. Primeira, teria de ser desenvolvido um esquema para entrar de forma conveniente a configuração dos nós e identificar quais estão na borda. Segunda, seria necessário um algoritmo para gerar equações simultâneas adequadas com base na informação de entrada. O resultado final é que um software geral para resolver EDPs elípticas é relativamente complicado.

Um método usado para simplificar tais esforços é exigir uma grade muito fina. Em tais casos, é freqüentemente suposto que um nó mais próximo sirva como ponto de borda. Dessa forma, a análise não tem de considerar os parâmetros de peso da Seção 29.3.2. Embora isso introduza algum erro, o uso de uma grade suficientemente fina pode tornar a discrepância final desprezível. Entretanto, isso envolve prós e contras em virtude da carga computacional introduzida pelo aumento no número de equações simultâneas.

Como conseqüência dessas considerações, os analistas numéricos desenvolveram abordagens alternativas que diferem radicalmente dos métodos por diferenças finitas. Ainda que esses métodos por elementos finitos sejam conceitualmente difíceis, podem acomodar de forma muito mais fácil fronteiras irregulares. Voltaremo-nos para esses métodos no Capítulo 31. Antes de fazer isso, entretanto, primeiro vamos descrever as abordagens por diferenças finitas para outra categoria de EDPs — as equações parabólicas.

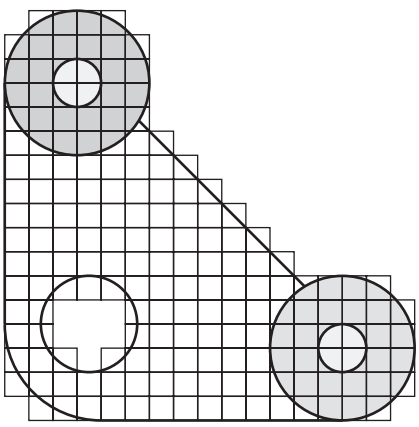


FIGURA 29.15
Uma grade de diferenças finitas superposta a uma junta de vedação com formato irregular.

PROBLEMAS

29.1 Use o método de Liebmann para determinar a temperatura da placa aquecida na Figura 29.4, mas com a condição de contorno superior aumentada para 120 ºC e a esquerda diminuída para 60 ºC. Use um fator de relaxação de 1,2 e itere até $\varepsilon_s = 1\%$.

29.2 Calcule os fluxos para o Problema 29.1 usando os parâmetros do Exemplo 29.3.

29.3 Repita o Exemplo 29.1, mas use 49 nós interiores (isto é, $\Delta x = \Delta y = 5$ cm).

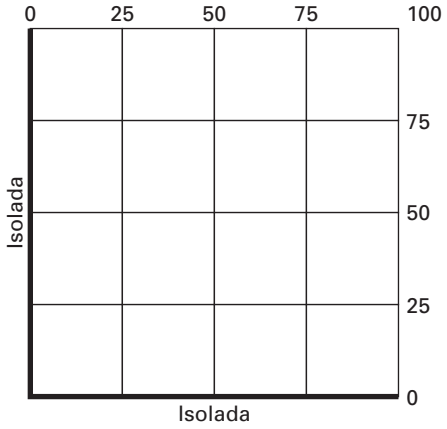
29.4 Repita o Problema 29.3, mas para o caso em que a aresta inferior seja isolada.

29.5 Repita os Exemplos 29.1 e 29.3, mas para o caso em que o fluxo na aresta inferior está sendo dirigido para baixo, com um valor de 1 cal/cm^2 · s.

29.6 Repita o Exemplo 29.4 para o caso em que os cantos inferior esquerdo e superior direito são arredondados da mesma forma que o canto inferior esquerdo na Figura 29.9. Observe que todas as temperaturas de contorno nos lados superior e direito estão fixas em 100 ºC e todas nos lados inferior e esquerdo estão fixadas em 50 ºC.

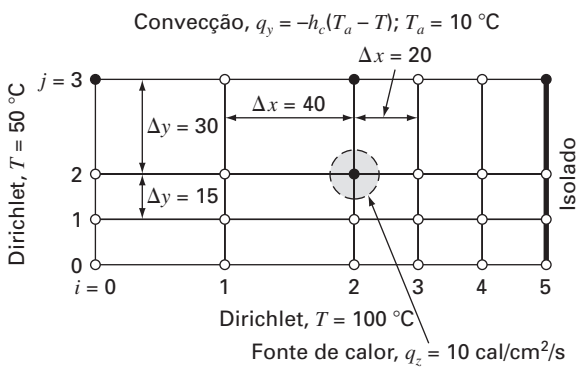
29.7 Com exceção das condições de contorno, a placa na Figura P29.7 tem exatamente as mesmas características que a placa usada nos Exemplos 29.1 a 29.4. Simule tanto as temperaturas quanto os fluxos para a placa.

Figura P29.7



29.8 Escreva equações para os nós escurecidos na grade na Figura P29.8. Observe que todas as unidades são do sistema cgs. O coeficiente de condutividade térmica para a placa é 0,75 cal/(s · cm · °C), o coeficiente de convecção é $h_c = 0{,}015$ cal/(cm^2 · C · s) e a espessura da placa é 0,5 cm.

Figura P29.8



29.9 Escreva equações para os nós escurecidos na grade na Figura P29.9. Observe que todas as unidades são do sistema cgs. O coeficiente de convecção é $h_c = 0{,}015$ cal/(cm^2 · C · s) e a espessura da placa é 1,5 cm.

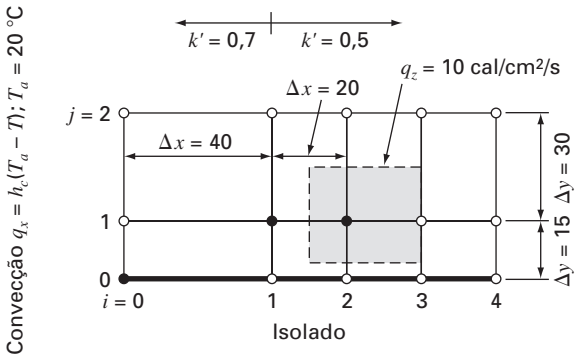


Figura P29.9

29.10 Aplique a abordagem por volume de controle para desenvolver a equação para o nó $(0, j)$ na Figura 29.7.

29.11 Deduza uma equação como a Equação (29.26) para o caso em que θ seja maior que 45° na Figura 29.11.

29.12 Desenvolva um programa computacional amigável ao usuário para implementar o método de Liebmann para uma placa retangular com condições de contorno de Dirichlet. Projete o programa de modo que ele possa calcular tanto a temperatura quanto o fluxo. Teste o programa duplicando os resultados dos Exemplos 29.1 e 29.2.

29.13 Use o programa do Problema 29.12 para resolver os Problemas 29.1 e 29.2.

29.14 Use o programa do Problema 29.12 para resolver o Problemas 29.3.

29.15 Use a abordagem por volume de controle e deduza a equação de nó para o nó (2, 2) na Figura 29.13 e inclua uma fonte de calor nesse ponto. Utilize os seguintes valores para as constantes: $\Delta z = 0{,}25$ cm, $h = 10$ cm, $k_A = 0{,}25$ W/cm · C e $k_B = 0{,}45$ W/cm · C. A fonte de calor vem apenas do material A a uma taxa de 6 W/cm^3.

29.16 Calcule o fluxo de calor (W/cm^2) para o nó (2, 2) na Figura 29.13, usando aproximações por diferenças finitas para os gradientes de temperatura nesse nó. Calcule o fluxo na direção horizontal nos materiais A e B e determine se esses dois fluxos deveriam ser iguais. Calcule também o fluxo vertical nos materiais A e B. Esses dois fluxos devem ser iguais? Use os seguintes valores para as constantes: $\Delta z = 0{,}5$ cm, $h = 10$ cm, $k_A = 0{,}25$ W/cm · C, $k_B = 0{,}45$ W/cm · C e as temperaturas nodais: $T_{22} = 51{,}6$ °C, $T_{21} = 74{,}2$ °C, $T_{23} = 45{,}3$ °C, $T_{32} = 38{,}6$ °C e $T_{12} = 87{,}4$ °C.

CAPÍTULO 30

Diferenças Finitas: Equações Parabólicas

O capítulo anterior tratava de EDPs estacionárias. Agora, voltamos-nos para as equações parabólicas, que são empregadas para caracterizar problemas que variam no tempo. Na parte final deste capítulo, ilustraremos como isso é feito em duas dimensões espaciais para a placa aquecida. Antes de fazer isso, mostraremos, primeiro, como o caso unidimensional mais simples é abordado.

30.1 A EQUAÇÃO DE CONDUÇÃO DE CALOR

De maneira parecida com a dedução da equação de Laplace [Equação (29.6)], a conservação de calor pode ser usada para desenvolver um balanço de calor para um elemento infinitesimal na barra (ou haste) isolada longa e fina mostrada na Figura 30.1. Entretanto, em vez de examinar o caso estacionário, o balanço presente também considerará as quantidades de calor armazenadas no elemento em um período Δt. Logo, o balanço é da seguinte forma, entradas − saídas = armazenado, ou

$$q(x)\ \Delta y\ \Delta z\ \Delta t - q(x + \Delta x)\ \Delta y\ \Delta z\ \Delta t = \Delta x\ \Delta y\ \Delta z \rho C\ \Delta T$$

Dividindo pelo volume do elemento ($= \Delta x\ \Delta y\ \Delta z$) e Δt, obtemos

$$\frac{q(x) - q(x + \Delta x)}{\Delta x} = \rho C \frac{\Delta T}{\Delta t}$$

Tomando o limite, temos

$$-\frac{\partial q}{\partial x} = \rho C \frac{\partial T}{\partial t}$$

Substituindo a lei de condução de calor de Fourier [Equação (29.4)], obtemos

$$k \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t} \tag{30.1}$$

que é a *equação de condução de calor*.

Do mesmo modo que com as EDPs elípticas, as equações parabólicas podem ser resolvidas substituindo-se as derivadas parciais por diferenças divididas finitas. Entretanto, em contraste com as EDPs elípticas, devemos agora considerar variações no tempo, bem como no espaço. Enquanto as equações elípticas eram limitadas em todas as dimensões

FIGURA 30.1
Uma barra fina, isolada em todos os pontos, exceto nas extremidades.



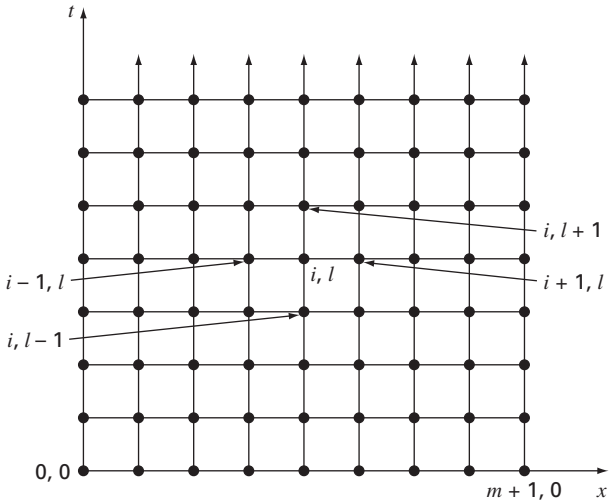


FIGURA 30.2
Uma grade usada para a solução por diferenças finitas de EDPs parabólicas em duas variáveis independentes, tal como a equação de condução de calor. Observe como, em contraste com a Figura 29.3, essa grade tem uma extremidade aberta na dimensão temporal.

relevantes, as EDPs parabólicas têm uma extremidade aberta no tempo (Figura 30.2). Por causa de sua natureza variável no tempo, a solução dessas equações envolve diversas questões novas, em particular, a estabilidade. Isso, assim como outros aspectos das EDPs parabólicas, será examinado nas seções seguintes, enquanto apresentamos duas abordagens de solução fundamentais — esquemas explícito e implícito.

30.2 MÉTODOS EXPLÍCITOS

A equação de condução de calor exige aproximações para a segunda derivada no espaço e para a primeira derivada no tempo. A primeira é representada da mesma maneira que na equação de Laplace, por uma diferença dividida finita centrada:

$$\frac{\partial^2 T}{\partial x^2} = \frac{T_{i+1}^l - 2T_i^l + T_{i-1}^l}{\Delta x^2} \tag{30.2}$$

que tem um erro (lembre-se da Figura 23.3) de $O[(\Delta x)^2]$. Observe a pequena variação na notação dos sobrescritos, que é usada para denotar o tempo. Isso é feito para que um segundo subscrito possa ser usado para designar uma segunda dimensão espacial quando a abordagem for expandida para duas dimensões espaciais.

Uma diferença finita dividida progressiva é usada para obter uma aproximação da derivada no tempo

$$\frac{\partial T}{\partial t} = \frac{T_i^{l+1} - T_i^l}{\Delta t} \tag{30.3}$$

que tem um erro (lembre-se da Figura 23.1) de $O(\Delta t)$.

Substituindo as Equações (30.2) e (30.3) na Equação (30.1), temos

$$k\frac{T_{i+1}^l - 2T_i^l + T_{i-1}^l}{(\Delta x)^2} = \frac{T_i^{l+1} - T_i^l}{\Delta t} \tag{30.4}$$

que pode ser reescrita como

$$T_i^{l+1} = T_i^l + \lambda\left(T_{i+1}^l - 2T_i^l + T_{i-1}^l\right) \tag{30.5}$$

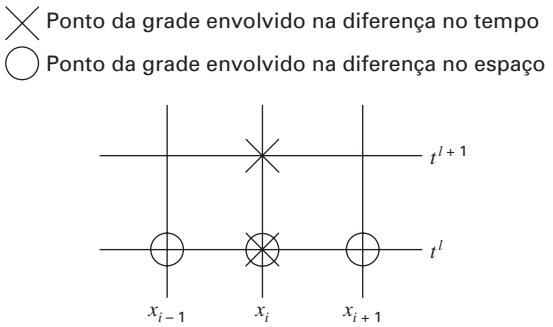


FIGURA 30.3
Uma molécula computacional para a forma explícita.

onde $\lambda = k\ \Delta t/(\Delta x)^2$.

Essa equação pode ser escrita para todos os nós interiores da barra. Ela então fornece um meio explícito de calcular valores em cada nó para um instante futuro baseado nos valores atuais no nó e em seus vizinhos. Observe que essa abordagem é, na realidade, uma manifestação do método de Euler para resolver um sistema de EDOs. Isto é, se soubermos a distribuição de temperatura como uma função da posição em um instante inicial, podemos calcular a distribuição em um instante futuro com base na Equação (30.5).

Uma molécula computacional para o método explícito é descrita na Figura 30.3, mostrando os nós que constituem as aproximações espacial e temporal. Essa molécula pode ser contrastada com outras neste capítulo, para ilustrar as diferenças entre as abordagens.

EXEMPLO 30.1 Solução Explícita da Condução de Calor Unidimensional

Enunciado do Problema. Use o método explícito para determinar a distribuição de temperatura de uma barra longa e fina, com um comprimento de 10 cm e com os seguintes valores: $k' = 0{,}49$ cal/(s · cm · °C), $\Delta x = 2$ cm e $\Delta t = 0{,}1$ s. Em $t = 0$, a temperatura da barra é zero e as condições de contorno são fixas em todos os tempos em $T(0) = 100$ °C e $T(10) = 50$ °C. Observe que a barra é de alumínio com $C = 0{,}2174$ cal/(g · °C) e $\rho = 2{,}7$ g/cm^3. Portanto, $k = 0{,}49/(2{,}7 \cdot 0{,}2174) = 0{,}835$ cm^2/s e $\lambda = 0{,}835(0{,}1)/(2)^2 = 0{,}020875$.

Solução. Aplicando a Equação (30.5), obtemos o seguinte valor em $t = 0{,}1$ s para o nó em $x = 2$ cm:

$$T_1^1 = 0 + 0{,}020875[0 - 2(0) + 100] = 2{,}0875$$

Nos outros pontos interiores, $x = 4$, 6 e 8 cm, os resultados são

$$T_2^1 = 0 + 0{,}020875[0 - 2(0) + 0] = 0$$
$$T_3^1 = 0 + 0{,}020875[0 - 2(0) + 0] = 0$$
$$T_4^1 = 0 + 0{,}020875[50 - 2(0) + 0] = 1{,}0438$$

Em $t = 0{,}2$ s, os valores nos quatro nós interiores são calculados como

$$T_1^2 = 2{,}0875 + 0{,}020875[0 - 2(2{,}0875) + 100] = 4{,}0878$$
$$T_2^2 = 0 + 0{,}020875[0 - 2(0) + 2{,}0875] = 0{,}043577$$
$$T_3^2 = 0 + 0{,}020875[1{,}0438 - 2(0) + 0] = 0{,}021788$$
$$T_4^2 = 1{,}0438 + 0{,}020875[50 - 2(1{,}0438) + 0] = 2{,}0439$$

Os cálculos continuam e os resultados com intervalos de 3 s são descritos na Figura 30.4. O aumento geral na temperatura com o tempo indica que o cálculo captura a difusão de calor das fronteiras para o interior da barra.

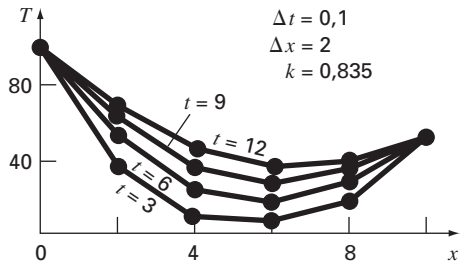


FIGURA 30.4
Distribuição de temperatura em uma barra longa e fina calculada com o método explícito descrito na Seção 30.2.

30.2.1 Convergência e Estabilidade

A *convergência* significa que, quando Δx e Δt tendem a zero, os resultados da técnica por diferenças finitas se aproximam da solução verdadeira. A *estabilidade* significa que erros em qualquer estágio do cálculo não são amplificados, mas, sim, atenuados conforme os cálculos progridem. Pode ser mostrado (Carnahan et al., 1969) que o método explícito é tanto convergente quanto estável se $\lambda \leq 1/2$, ou

$$\Delta t \leq \frac{1}{2}\frac{\Delta x^2}{k} \tag{30.6}$$

Além disso, deve ser observado que tomar $\lambda \leq 1/2$ poderia resultar em uma solução na qual os erros não crescem, mas oscilam. Tomar $\lambda \leq 1/4$ garante que a solução não oscilará. Também é conhecido que tomar $\lambda = 1/6$ tende a minimizar o erro de truncamento (Carnahan et al., 1969).

A Figura 30.5 é um exemplo da instabilidade causada pela violação da Equação (30.6). Esse gráfico é para o mesmo caso que no Exemplo 30.1, mas com $\lambda = 0{,}735$, o que é consideravelmente maior que 0,5. Como na Figura 30.5, a solução sofre oscilações progressivamente crescentes. Essa situação continuará a se deteriorar conforme os cálculos continuam.

Embora o fato de a Equação (30.6) ser satisfeita aliviar as instabilidades do tipo manifestado na Figura 30.5, ela também coloca uma forte limitação no método explícito. Por exemplo, suponha que Δx seja dividido na metade para melhorar a aproximação na segunda derivada espacial. De acordo com a Equação (30.6), o passo no tempo deve ser dividido por 4 para manter a convergência e a estabilidade. Logo, para fazer cálculos comparáveis, o número de passos no tempo deve ser aumentado por um fator 4. Além disso, o cálculo para cada um desses passos no tempo levará duas vezes mais tempo, já que dividir Δx por 2 dobra o número total de nós para os quais as equações precisam ser escritas. Conseqüentemente, para o caso unidimensional, dividir Δx por 2 resulta em um aumento de oito vezes no número de cálculos. Assim, a carga computacional pode ser grande para atingir uma acucária aceitável. Como será descrito em breve, estão disponíveis outras técnicas que não sofrem limitações tão severas.

30.2.2 Condições de Contorno na Derivada

Como foi o caso para as EDPs elípticas (lembre-se da Seção 29.3.1), as condições de contorno na derivada podem ser prontamente incorporadas nas equações parabólicas. Para uma barra unidimensional, isso exige a adição de duas equações para caracterizar o balanço de calor nos nós nas extremidades. Por exemplo, o nó na extremidade esquerda ($i = 0$) seria representado por

$$T_0^{l+1} = T_0^l + \lambda\left(T_1^l - 2T_0^l + T_{-1}^l\right)$$

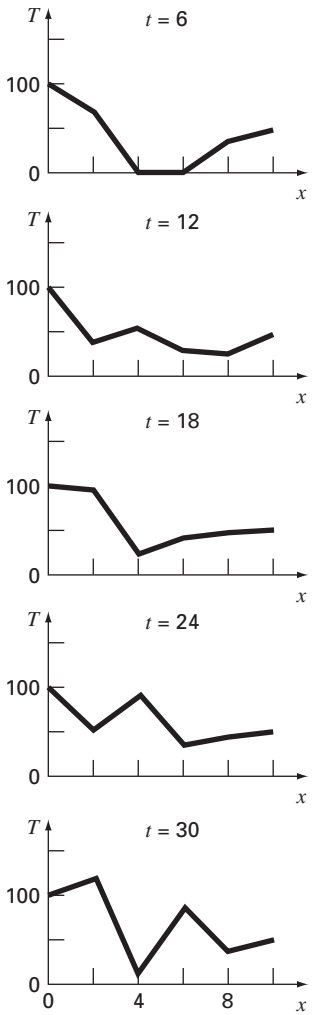


FIGURA 30.5
Uma ilustração da instabilidade. Solução do Exemplo 30.1, mas com $\lambda = 0{,}735$.

Portanto, um ponto imaginário é introduzido em $i = -1$ (lembre-se da Figura 29.7). Entretanto, como no caso elíptico, esse ponto fornece um veículo para incorporar a condição de contorno da derivada na análise. O Problema 30.2 no final do capítulo trata desse exercício.

30.2.3 Aproximações Temporais de Ordem Superior

A idéia geral de reexpressar a EDP como um sistema de EDOs algumas vezes é denominada *método das linhas*. Obviamente, uma maneira de melhorar a abordagem usada anteriormente seria empregar um esquema de integração mais acurado para resolver as EDOs. Por exemplo, o método de Heun pode ser empregado para obter acurácia temporal de segunda ordem. Um exemplo dessa abordagem é chamado *método de MacCormack*. Este e outros métodos explícitos melhorados são discutidos em outras referências (por exemplo, Hoffman, 1992).

30.3 UM MÉTODO IMPLÍCITO SIMPLES

Como observado anteriormente, formulações por diferenças finitas explícitas têm problemas relacionados com a estabilidade. Além disso, como descrito na Figura 30.6, eles excluem informações que têm influência na solução. Os métodos implícitos superam ambas as dificuldades ao custo de algoritmos um pouco mais complicados.

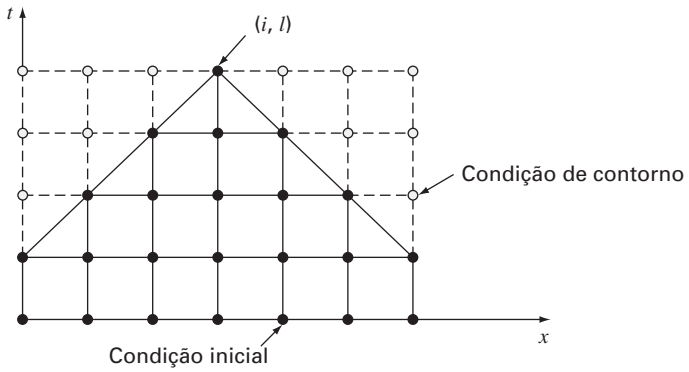


FIGURA 30.6
Representação do efeito de outros nós na aproximação por diferença finita no nó (*i*, *l*), usando um esquema explícito por diferença finita. Os nós escuros têm uma influência no nó (*i*, *l*), enquanto os não escurecidos, os quais, na realidade, afetam (*i*, *l*), são excluídos.

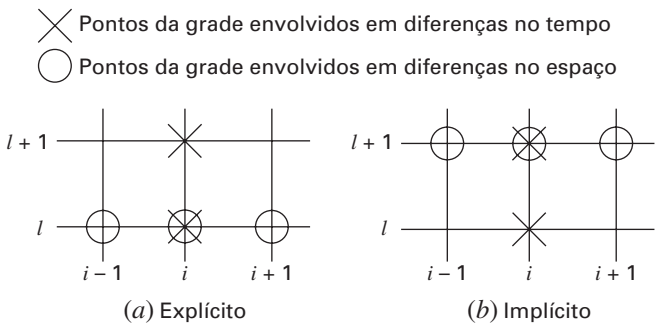


FIGURA 30.7
Molécula fundamental ilustrando as diferenças entre os métodos: (*a*) explícitos e (*b*) implícitos.

A diferença fundamental entre as aproximações explícita e implícita está descrita na Figura 30.7. Para a forma explícita, aproximamos a derivada espacial no nível l de tempo (Figura 30.7*a*). Lembre-se de que, quando substituímos essa aproximação na equação diferencial parcial, obtemos uma equação de diferenças (30.4) com uma única incógnita T_i^{l+1}. Logo, podemos escrever "explicitamente" a solução para essa incógnita como na Equação (30.5).

Nos métodos implícitos, a derivada espacial é aproximada em um nível de tempo avançado, $l + 1$. Por exemplo, a segunda derivada seria aproximada por (Figura 30.7*b*)

$$\frac{\partial^2 T}{\partial x^2} \cong \frac{T_{i+1}^{l+1} - 2T_i^{l+1} + T_{i-1}^{l+1}}{(\Delta x)^2} \tag{30.7}$$

tem acurácia de segunda ordem. Quando essa relação é substituída na EDP original, a equação de diferenças resultante contém diversas incógnitas. Portanto, ela não pode ser resolvida explicitamente por uma reorganização algébrica simples como foi feito na passagem da Equação (30.4) para a Equação (30.5). Em vez disso, o sistema de equações todo deve ser resolvido simultaneamente. Isso é possível porque, juntamente com as condições de contorno, as formulações implícitas resultam em um conjunto de equações lineares algébricas com o mesmo número de incógnitas. Logo, o método se reduz à solução de um conjunto de equações simultâneas em cada ponto no tempo.

Para ilustrar como isso é feito, substitua as Equações (30.3) e (30.7) na Equação (30.1) para obter

$$k\frac{T_{i+1}^{l+1} - 2T_i^{l+1} + T_{i-1}^{l+1}}{(\Delta x)^2} = \frac{T_i^{l+1} - T_i^l}{\Delta t}$$

que pode ser expresso como

$$-\lambda T_{i-1}^{l+1} + (1+2\lambda)T_i^{l+1} - \lambda T_{i+1}^{l+1} = T_i^l \tag{30.8}$$

onde $\lambda = k\,\Delta t/(\Delta x)^2$. Essa equação se aplica a todos, exceto o primeiro e o último nós interiores, nos quais precisa ser modificada para refletir as condições de contorno. Para o caso no qual as temperaturas nas extremidades da barra são dadas, a condição de contorno na extremidade esquerda da barra ($i = 0$) pode ser expressa como

$$T_0^{l+1} = f_0(t^{l+1}) \tag{30.9}$$

onde $f_0(t^{l+1})$ é uma função descrevendo como a temperatura de contorno varia com o tempo. Substituindo a Equação (30.9) na Equação (30.8), temos a equação de diferença para o primeiro nó interior ($i = 1$):

$$(1+2\lambda)T_1^{l+1} - \lambda T_2^{l+1} = T_1^l + \lambda f_0(t^{l+1}) \tag{30.10}$$

De forma análoga, para o último nó interior ($i = m$),

$$-\lambda T_{m-1}^{l+1} + (1+2\lambda)T_m^{l+1} = T_m^l + \lambda f_{m+1}(t^{l+1}) \tag{30.11}$$

onde $f_{m+1}(t^{l+1})$ descreve a variação de temperatura especificada na extremidade direita da barra ($i = m+1$).

Quando as Equações (30.8), (30.10) e (30.11) são escritas para todos os nós interiores, o conjunto resultante de m equações algébricas lineares tem m incógnitas. Além disso, o método tem a vantagem adicional do sistema ser tridiagonal. Assim, podemos utilizar os algoritmos de solução extremamente eficientes (lembre-se da Seção 11.1) que estão disponíveis para sistemas tridiagonais.

EXEMPLO 30.2 Solução Implícita Simples da Equação de Condução de Calor

Enunciado do Problema. Use a aproximação implícita simples por diferenças finitas para resolver o mesmo problema que no Exemplo 30.1.

Solução. Para a barra do Exemplo 30.1, $\lambda = 0{,}020875$. Portanto, em $t = 0$, a Equação (30.10) pode ser escrita para o primeiro nó interior como

$$1{,}04175T_1^1 - 0{,}020875T_2^1 = 0 + 0{,}020875(100)$$

ou

$$1{,}04175T_1^1 - 0{,}020875T_2^1 = 2{,}0875$$

De modo parecido, as equações (30.8) e (30.11) podem ser aplicadas aos outros nós interiores. Isso nos leva ao seguinte conjunto de equações simultâneas:

$$\begin{bmatrix} 1{,}04175 & -0{,}020875 & & \\ -0{,}020875 & 1{,}04175 & -0{,}020875 & \\ & -0{,}020875 & 1{,}04175 & -0{,}020875 \\ & & -0{,}020875 & 1{,}04175 \end{bmatrix} \begin{Bmatrix} T_1^1 \\ T_2^1 \\ T_3^1 \\ T_4^1 \end{Bmatrix} = \begin{Bmatrix} 2{,}0875 \\ 0 \\ 0 \\ 1{,}04375 \end{Bmatrix}$$

que pode ser resolvido para determinar a temperatura em $t = 0{,}1$ s:

$$T_1^1 = 2{,}0047$$
$$T_2^1 = 0{,}0406$$
$$T_3^1 = 0{,}0209$$
$$T_4^1 = 1{,}0023$$

Observe como, em contraste com o Exemplo 30.1, todos os pontos se tornaram diferentes da condição inicial durante o primeiro passo.

Para determinar as temperaturas em $t = 0{,}2$, o vetor à direita deve ser modificado para levar em conta os resultados do primeiro passo, como em

$$\begin{Bmatrix} 4{,}09215 \\ 0{,}04059 \\ 0{,}02090 \\ 2{,}04069 \end{Bmatrix}$$

As equações simultâneas podem, então, ser resolvidas para determinar as temperaturas em $t = 0{,}2$ s:

$$T_1^2 = 3{,}9305$$
$$T_2^2 = 0{,}1190$$
$$T_3^2 = 0{,}0618$$
$$T_4^2 = 1{,}9653$$

Embora o método implícito descrito seja estável e convergente, seu defeito é que a aproximação temporal por diferença tem acurácia de primeira ordem, ao passo que a aproximação espacial por diferença tem acurácia de segunda ordem (Figura 30.8). Na próxima seção, apresentaremos um método implícito alternativo que remedia essa situação.

Antes de prosseguir, deve ser mencionado que, embora o método implícito simples seja incondicionalmente estável, existe um limite na acurácia quando se usar passos grandes no tempo. Como conseqüência, ele não é muito mais eficiente que as abordagens explícitas para a maioria dos problemas que variam no tempo.

Onde ele realmente se sobressai é nos problemas estacionários. Lembre-se, do Capítulo 29, que pode ser usada uma forma do Gauss-Seidel (método de Liebmann) para obter soluções estacionárias para equações elípticas. Uma abordagem alternativa seria executar uma solução dependente do tempo até que ela atinja o estado estacionário. Nesse caso, como os resultados intermediários inexatos não causam problemas, os métodos implícitos permitem que você empregue passos maiores no tempo e, assim, possa gerar resultados estacionários de maneira eficiente.

FIGURA 30.8
Uma molécula computacional para o método implícito simples.

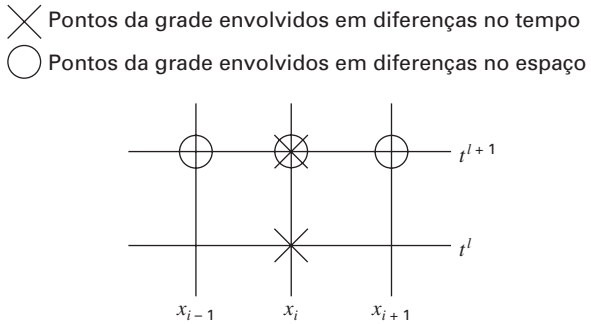
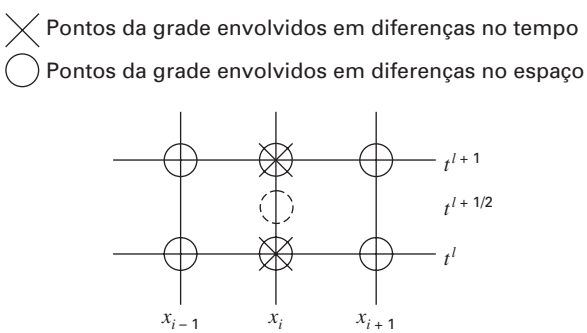


FIGURA 30.9
Uma molécula computacional para o método de Crank-Nicolson.



30.4 O MÉTODO DE CRANK-NICOLSON

O *método de Crank-Nicolson* fornece um esquema implícito alternativo que tem acurácia de segunda ordem, tanto no espaço quanto no tempo. Para fornecer essa acurácia, são desenvolvidas aproximações por diferença no ponto médio do incremento no tempo (Figura 30.9). Para fazer isso, a primeira derivada no tempo pode ser aproximada em $t^{l+1/2}$ por

$$\frac{\partial T}{\partial t} \cong \frac{T_i^{l+1} - T_i^l}{\Delta t} \tag{30.12}$$

A segunda derivada no espaço pode ser determinada no ponto médio tomando-se a média da aproximação por diferenças no início (t^l) e no final (t^{l+1}) do incremento de tempo:

$$\frac{\partial^2 T}{\partial x^2} \cong \frac{1}{2}\left[\frac{T_{i+1}^l - 2T_i^l + T_{i-1}^l}{(\Delta x)^2} + \frac{T_{i+1}^{l+1} - 2T_i^{l+1} + T_{i-1}^{l+1}}{(\Delta x)^2}\right] \tag{30.13}$$

Substituindo as Equações (30.12) e (30.13) na Equação (30.1) e agrupando os termos, obtemos

$$-\lambda T_{i-1}^{l+1} + 2(1+\lambda)T_i^{l+1} - \lambda T_{i+1}^{l+1} = \lambda T_{i-1}^l + 2(1-\lambda)T_i^l + \lambda T_{i+1}^l \tag{30.14}$$

onde $\lambda = k\,\Delta t/(\Delta x)^2$. Como acontecia com a abordagem implícita simples, as condições de contorno de $T_0^{l+1} = f_0(t^{l+1})$ e $T_{m+1}^{l+1} = f_{m+1}(t^{l+1})$ podem ser impostas para deduzir versões da Equação (30.14) para o primeiro e o último nó interior. Para o primeiro nó interior

$$2(1+\lambda)T_1^{l+1} - \lambda T_2^{l+1} = \lambda f_0(t^l) + 2(1-\lambda)T_1^l + \lambda T_2^l + \lambda f_0(t^{l+1}) \tag{30.15}$$

e para o último nó interior,

$$-\lambda T_{m-1}^{l+1} + 2(1+\lambda)T_m^{l+1} = \lambda f_{m+1}(t^l) + 2(1-\lambda)T_m^l + \lambda T_{m-1}^l + \lambda f_{m+1}(t^{l+1}) \tag{30.16}$$

Embora as Equações (30.14) a (30.16) sejam ligeiramente mais complicadas que as Equações (30.8), (30.10) e (30.11), elas também são tridiagonais e, portanto, podem ser resolvidas eficientemente.

EXEMPLO 30.3 A Solução de Crank-Nicolson para a Equação de Condução de Calor

Enunciado do Problema. Use o método de Crank-Nicolson para resolver o mesmo problema que nos Exemplos 30.1 e 30.2.

Solução. As Equações (30.14) a (30.16) podem ser usadas para gerar o seguinte conjunto de equações tridiagonais:

$$\begin{bmatrix} 2{,}04175 & -0{,}020875 & & \\ -0{,}020875 & 2{,}04175 & -0{,}020875 & \\ & -0{,}020875 & 2{,}04175 & -0{,}020875 \\ & & -0{,}020875 & 2{,}04175 \end{bmatrix} \begin{Bmatrix} T_1^1 \\ T_2^1 \\ T_3^1 \\ T_4^1 \end{Bmatrix} = \begin{Bmatrix} 4{,}175 \\ 0 \\ 0 \\ 2{,}0875 \end{Bmatrix}$$

o qual pode ser resolvido para determinar as temperaturas em $t = 0{,}1$ s:

$$T_1^1 = 2{,}0450$$
$$T_2^1 = 0{,}0210$$
$$T_3^1 = 0{,}0107$$
$$T_4^1 = 1{,}0225$$

Para estabelecer as temperaturas em $t = 0{,}2$ s, o vetor à direita deve ser modificado para

$$\begin{Bmatrix} 8{,}1801 \\ 0{,}0841 \\ 0{,}0427 \\ 4{,}0901 \end{Bmatrix}$$

As equações simultâneas podem então ser resolvidas por

$$T_1^2 = 4{,}0073$$
$$T_2^2 = 0{,}0826$$
$$T_3^2 = 0{,}0422$$
$$T_4^2 = 2{,}0036$$

30.4.1 Comparação dos Métodos Unidimensionais

A Equação (30.1) pode ser resolvida analiticamente. Por exemplo, está disponível uma solução para o caso no qual a temperatura da barra é, inicialmente, nula. Em $t = 0$, a condição de contorno em $x = L$ é instantaneamente aumentada para um nível constante $\overline{T}$, enquanto $T(0)$ é mantido no 0. Nesse caso, a temperatura pode ser calculada por

$$T = \overline{T}\left[\frac{x}{L} + \sum_{n=0}^{\infty} \frac{2}{n\pi}(-1)^n \operatorname{sen}\left(\frac{n\pi x}{L}\right) \exp\left(\frac{-n^2\pi^2 kt}{L^2}\right)\right] \tag{30.17}$$

onde L é o comprimento total da barra. Essa equação pode ser utilizada para calcular a evolução da distribuição da temperatura para cada condição de contorno. Então, a solução total pode ser determinada por superposição.

EXEMPLO 30.4 Comparação das Soluções Analítica e Numéricas

Enunciado do Problema. Compare a solução analítica da Equação (30.17) com os resultados numéricos obtidos com as técnicas explícitas, implícita simples e de Crank-Nicolson. Faça a comparação para a barra empregada nos Exemplos 30.1, 30.2 e 30.3.

Solução. Lembre-se, a partir dos exemplos anteriores, de que $k = 0{,}835$ cm^2/s, $L = 10$ cm e $\Delta x = 2$ cm. Nesse caso, a Equação (30.17) pode ser usada para prever que a temperatura em $x = 2$ cm e $t = 10$ s seria igual a 64,8018. A Tabela 30.1 apresenta as previsões numéricas de $T(2, 10)$. Observe que diversos passos no tempo são empregados. Esses resultados indicam várias propriedades dos métodos numéricos. Primeiro, podemos ver que o método explícito é instável para valores altos de λ. Essa instabilidade não é manifestada por nenhuma das abordagens implícitas. Segundo, o método de Crank-Nicolson converge mais rapidamente, conforme λ diminui, e fornece resultados moderadamente acurados mesmo quando λ é relativamente alto. Essas saídas são como se espera porque o método de Crank-

TABELA 30.1 Comparação de três métodos para resolver uma EDP parabólica: a barra aquecida. Os resultados são para temperaturas em $t = 10$ s em $x = 2$ cm, para a barra dos Exemplos 30.1 a 30.3. Observe que a solução analítica é $T(2, 10) = 64{,}8018$.

Δt	λ	Explícita	Implícita	Crank-Nicolson
10	2,0875	208,75	53,01	79,77
5	1,04375	−9,13	58,49	64,79
2	0,4175	67,12	62,22	64,87
1	0,20875	65,91	63,49	64,77
0,5	0,104375	65,33	64,12	64,74
0,2	0,04175	64,97	64,49	64,73

Nicolson tem acurácia de segunda ordem com relação a ambas as variáveis independentes. Finalmente, observe que, conforme λ diminui, os métodos parecem estar convergindo para um valor de 64,73, que é diferente do resultado analítico de 64,80. Isso não deveria surpreender, já que foi usado um valor fixo $\Delta x = 2$ para caracterizar a dimensão x. Se tanto Δx quanto Δt diminuírem, conforme λ diminui (isto é, se forem usados mais segmentos espaciais), a solução numérica estaria mais próxima do resultado analítico.

O método de Crank-Nicolson é freqüentemente usado para resolver EDPs parabólicas lineares em uma dimensão espacial. Sua vantagem se torna ainda mais pronunciada para aplicações mais complicadas, como aquelas envolvendo grades desigualmente espaçadas. Tal espaçamento não uniforme é, às vezes, vantajoso, quando temos conhecimento prévio de que a solução varia rapidamente em partes localizadas do sistema. Uma discussão mais profunda dessas aplicações e do método de Crank-Nicolson pode ser encontrada em outras referências (Ferziger, 1981; Lapidus e Pinder, 1981; Hoffman, 1992).

30.5 EQUAÇÕES PARABÓLICAS EM DUAS DIMENSÕES ESPACIAIS

A equação de condução de calor pode ser aplicada a mais de uma dimensão espacial. Para duas dimensões, ela tem a forma

$$\frac{\partial T}{\partial t} = k\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right) \tag{30.18}$$

Uma aplicação dessa equação é modelar a distribuição de temperatura em uma placa aquecida. Entretanto, em vez de caracterizar sua distribuição estacionária, como foi feito no Capítulo 29, a Equação (30.18) fornece um meio de calcular a distribuição de temperatura na placa, conforme ela varia com o tempo.

30.5.1 Esquemas padrão Explícito e Implícito

Uma solução explícita pode ser obtida substituindo aproximações por diferenças finitas da forma das Equações (30.2) e (30.3) na Equação (30.18). Entretanto, como no caso unidimensional, essa abordagem é limitada por um critério de estabilidade restritivo. Para o caso bidimensional, o critério é

$$\Delta t \leq \frac{1}{8}\frac{(\Delta x)^2 + (\Delta y)^2}{k}$$

Logo, para uma grade uniforme ($\Delta x = \Delta y$), $\lambda = k\,\Delta t/(\Delta x)^2$ deve ser menor ou igual a $1/4$. Conseqüentemente, dividir o tamanho do passo por dois resulta em um aumento de quatro vezes no número de nós e aumento de 16 vezes no esforço computacional.

Como no caso dos sistemas unidimensionais, as técnicas implícitas fornecem alternativas que garantem a estabilidade. Entretanto, a aplicação direta de métodos implícitos, tais como a técnica de Crank-Nicolson, leva à solução de $m \times n$ equações simultâneas. Além disso, quando escrita para duas ou três dimensões espaciais, essas equações perdem a propriedade valiosa de serem tridiagonais. Logo, o armazenamento da matriz e o esforço computacional podem ser exorbitantemente grandes. O método descrito na próxima seção oferece uma solução para esse dilema.

30.5.2 O Esquema ADI

O esquema implícito de direção alternada ou ADI fornece um meio para resolver equações parabólicas em duas dimensões espaciais, usando matrizes tridiagonais. Para fazer isso, cada incremento no tempo é executado em dois passos (Figura 30.10). Para o primeiro passo, a Equação (30.18) é aproximada por

$$\frac{T_{i,j}^{l+1/2} - T_{i,j}^{l}}{\Delta t/2} = k\left[\frac{T_{i+1,j}^{l} - 2T_{i,j}^{l} + T_{i-1,j}^{l}}{(\Delta x)^2} + \frac{T_{i,j+1}^{l+1/2} - 2T_{i,j}^{l+1/2} + T_{i,j-1}^{l+1/2}}{(\Delta y)^2}\right] \tag{30.19}$$

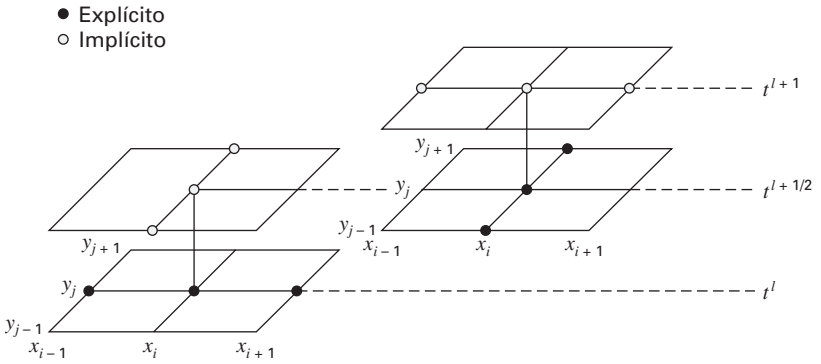


FIGURA 30.10
Os dois meios-passos usados na implementação do esquema implícito de direção alternada para resolver as equações parabólicas em duas dimensões espaciais.

Assim, a aproximação de $\partial^2 T/\partial x^2$ é escrita explicitamente — isto é, no ponto base t^l, no qual os valores da temperatura são conhecidos. Por conseguinte, apenas os três termos de temperatura na aproximação de $\partial^2 T/\partial y^2$ são desconhecidos. No caso de uma grade quadrada ($\Delta y = \Delta x$), essa equação pode ser expressa como

$$-\lambda T_{i,j-1}^{l+1/2} + 2(1+\lambda)T_{i,j}^{l+1/2} - \lambda T_{i,j+1}^{l+1/2} = \lambda T_{i-1,j}^{l} + 2(1-\lambda)T_{i,j}^{l} + \lambda T_{i+1,j}^{l} \tag{30.20}$$

a qual, quando escrita para o sistema, resulta em um conjunto tridiagonal de equações simultâneas.

Para o segundo passo de $t^{l+1/2}$ a t^{l+1}, a Equação (30.18) é aproximada por

$$\frac{T_{i,j}^{l+1} - T_{i,j}^{l+1/2}}{\Delta t/2} = k\left[\frac{T_{i+1,j}^{l+1} - 2T_{i,j}^{l+1} + T_{i-1,j}^{l+1}}{(\Delta x)^2} + \frac{T_{i,j+1}^{l+1/2} - 2T_{i,j}^{l+1/2} + T_{i,j-1}^{l+1/2}}{(\Delta y)^2}\right] \tag{30.21}$$

Em contraste com a Equação (30.19), a aproximação de $\partial^2 T/\partial x^2$ agora é implícita. Portanto, o viés introduzido pela Equação (30.19) será parcialmente corrigido. Para uma grade quadrada, a Equação (30.21) pode ser escrita como

$$-\lambda T_{i-1,j}^{l+1} + 2(1+\lambda)T_{i,j}^{l+1} - \lambda T_{i+1,j}^{l+1} = \lambda T_{i,j-1}^{l+1/2} + 2(1-\lambda)T_{i,j}^{l+1/2} + \lambda T_{i,j+1}^{l+1/2} \tag{30.22}$$

Novamente, quando escrita para uma grade bidimensional, a equação resulta em um sistema tridiagonal (Figura 30.11). Como no exemplo a seguir, isso leva a uma solução numérica eficiente.

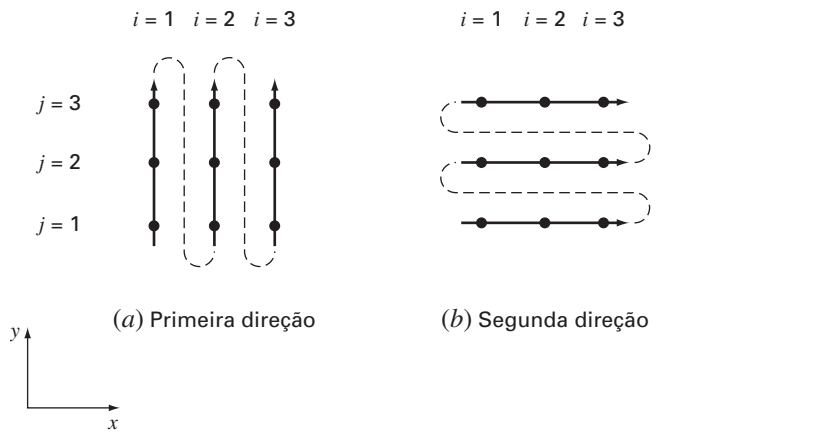


FIGURA 30.11
O método ADI resulta em equações tridimensionais apenas se for aplicado na dimensão em que é implícita. Logo, no primeiro passo (*a*), ele é aplicado na dimensão *y* e, no segundo passo (*b*), na dimensão *x*. Essas "direções alternadas" são a raiz do nome do método.

EXEMPLO 30.5 O método ADI

Enunciado do Problema. Use o método ADI para determinar a temperatura da placa nos Exemplos 29.1 e 29.2. Em $t = 0$, suponha que a temperatura da placa seja 0 e que as temperaturas de contorno sejam levadas instantaneamente para os níveis mostrados na Figura 29.4. Use um passo no tempo de 10 s. Lembre-se, do Exemplo 30.1, de que o coeficiente de difusividade térmica do alumínio é $k = 0{,}835\ \text{cm}^2/\text{s}$.

Solução. Um valor de $\Delta x = 10$ cm foi usado para caracterizar a placa 40×40 cm dos Exemplos 29.1 e 29.2. Portanto, $\lambda = 0{,}835(10)/(10)^2 = 0{,}0835$. Para o primeiro passo para $t = 5$ (Figura 30.11*a*), a Equação (30.20) é aplicada aos nós (1, 1), (1, 2), e (1, 3) para fornecer as seguintes equações tridiagonais:

$$\begin{bmatrix} 2{,}167 & -0{,}0835 & \\ -0{,}0835 & 2{,}167 & -0{,}0835 \\ & -0{,}0835 & 2{,}167 \end{bmatrix} \begin{Bmatrix} T_{1,1} \\ T_{1,2} \\ T_{1,3} \end{Bmatrix} = \begin{Bmatrix} 6{,}2625 \\ 6{,}2625 \\ 14{,}6125 \end{Bmatrix}$$

as quais podem ser resolvidas por

$$T_{1,1} = 3{,}01060 \qquad T_{1,2} = 3{,}2708 \qquad T_{1,3} = 6{,}8692$$

De maneira análoga, podem ser desenvolvidas e resolvidas equações tridiagonais para

$$T_{2,1} = 0{,}1274 \qquad T_{2,2} = 0{,}2900 \qquad T_{2,3} = 4{,}1291$$

e

$$T_{3,1} = 2{,}0181 \qquad T_{3,2} = 2{,}2477 \qquad T_{3,3} = 6{,}0256$$

Para o segundo passo para $t = 10$ (Figura 30.11*b*), a Equação (30.22) é aplicada aos nós (1, 1), (2, 1), e (3, 1) para fornecer

$$\begin{bmatrix} 2{,}167 & -0{,}0835 & \\ -0{,}0835 & 2{,}167 & -0{,}0835 \\ & -0{,}0835 & 2{,}167 \end{bmatrix} \begin{Bmatrix} T_{1,1} \\ T_{2,1} \\ T_{3,1} \end{Bmatrix} = \begin{Bmatrix} 13{,}0639 \\ 0{,}2577 \\ 8{,}0619 \end{Bmatrix}$$

a qual pode ser resolvida por

$$T_{1,1} = 5{,}5855 \qquad T_{2,1} = 0{,}4782 \qquad T_{3,1} = 3{,}7388$$

Equações tridiagonais para as outras linhas podem ser desenvolvidas e resolvidas por

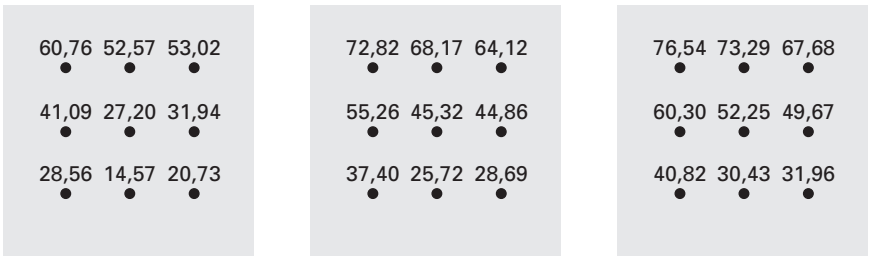
$$T_{1,2} = 6{,}1683 \qquad T_{2,2} = 0{,}8238 \qquad T_{3,2} = 4{,}2359$$

e

$$T_{1,3} = 13{,}1120 \qquad T_{2,3} = 8{,}3207 \qquad T_{3,3} = 11{,}3606$$

O cálculo pode ser repetido e resultados para $t = 100$, 200, e 300 s são descritos nas Figuras 30.12*a* a *c*, respectivamente. Como esperado, a temperatura da placa aumenta. Depois de passar tempo suficiente a temperatura se aproximará da distribuição estacionária na Figura 29.5.

FIGURA 30.12
Solução para placa aquecida do Exemplo 30.5 em (*a*) $t = 100$ s, (*b*) $t = 200$ s, e (*c*) $t = 300$ s.



(*a*) $t = 100$ s (*b*) $t = 200$ s (*c*) $t = 300$ s

O método ADI é apenas um de um grupo de técnicas chamadas métodos *splitting*. Alguns deles representam esforços para contornar defeitos do ADI. Discussão de outros métodos *splitting*, bem como mais informação sobre o ADI pode ser encontrada em outras referências (Ferziger, 1981; Lapidus e Pinder, 1981).

PROBLEMAS

30.1 Repita o Exemplo 30.1, mas use o método do ponto médio para gerar sua solução.

30.2 Repita o Exemplo 30.1, mas para o caso em que a barra está inicialmente a 25 °C e a derivada em $x = 0$ é igual a 1 e em $x = 10$ é igual a 0. Interprete seus resultados.

30.3 (**a**) Repita o Exemplo 30.1, mas para um passo no tempo de $\Delta t = 0{,}05$ s. Calcule resultados até $t = 0{,}2$. (**b**) Além disso, faça os mesmos cálculos com o método de Heun (sem iteração do corretor) com um tamanho de passo muito menor, de $\Delta t = 0{,}001$ s. Supondo que os resultados de (**b**) sejam aproximações válidas da solução verdadeira, determine erros relativos porcentuais para os resultados obtidos no Exemplo 30.1, bem como na parte (**a**).

30.4 Repita o Exemplo 30.2, mas para o caso em que a derivada em $x = 10$ seja igual a zero.

30.5 Repita o Exemplo 30.3, mas para $\Delta x = 1$ cm.

30.6 Repita o Exemplo 30.5, mas para a placa descrita no Problema 29.1.

30.7 A equação de advecção-difusão é usada para calcular a distribuição de concentração ao longo do comprimento de um reator químico retangular (ver a Seção 32.1),

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - U\frac{\partial c}{\partial x} - kc$$

onde c é a concentração (mg/m^3), t é o tempo (min), D é um coeficiente de difusão (m^2/min), x é a distância ao longo do eixo longitudinal do tanque (m) em que $x = 0$ na entrada do tanque, U é a velocidade na direção x (m/min) e k é uma taxa de reação (min^{-1}) pela qual a substância química decai para outra forma. Desenvolva um esquema explícito para resolver essa equação numericamente. Teste-o para $k = 0{,}15$, $D = 100$ e $U = 1$ para um tanque de comprimento 10 m. Use $\Delta x = 1$ e um tamanho de passo $\Delta t = 0{,}005$. Suponha que a concentração de entrada seja 100 e que a concentração inicial no tanque seja zero. Faça a simulação de $t = 0$ a 100 e trace as concentrações resultantes finais em função de x.

30.8 Desenvolva um programa computacional amigável ao usuário para o método explícito simples da Seção 30.2. Teste-o duplicando o Exemplo 30.1.

30.9 Modifique o programa no Problema 30.8 de modo que ele empregue tanto condições de contorno Dirichlet quanto condições de contorno na derivada. Teste-o resolvendo o Problema 30.2.

30.10 Desenvolva um programa computacional amigável ao usuário para implementar o esquema implícito simples da Seção 30.3. Teste-o duplicando o Exemplo 30.2.

30.11 Desenvolva um programa computacional amigável ao usuário para implementar o método de Crank-Nicolson da Seção 30.4. Teste-o duplicando o Exemplo 30.3.

30.12 Desenvolva um programa computacional amigável ao usuário para o método ADI descrito na Seção 30.5. Teste-o duplicando o Exemplo 30.5.

30.13 A forma adimensional para a condução de calor transiente em uma barra isolada (Equação 30.1) pode ser descrita como

$$\frac{\partial^2 u}{\partial \bar{x}^2} = \frac{\partial u}{\partial \bar{t}}$$

onde o espaço, tempo e temperatura adimensionais são definidos por

$$\bar{x} = \frac{x}{L} \qquad \bar{t} = \frac{T}{(\rho C L^2/k)} \qquad u = \frac{T - T_o}{T_L - T_o}$$

onde L é o comprimento da barra, k é a condutividade térmica do material da barra, ρ é a densidade, C é o calor específico, T_o é a temperatura em $x = 0$ e T_L é a temperatura em $x = L$. Isso se transforma nas seguintes condições de contorno e iniciais:

Condições de contorno	$u(0, \bar{t}) = 0$	$u(1, \bar{t}) = 0$
Condições iniciais	$u(\bar{x}, 0) = 0$	$0 \le \bar{x} \le 1$

Resolva essa equação adimensional para determinar a distribuição de temperatura, usando métodos de diferenças finitas e uma formulação de Crank-Nicolson com acurácia de segunda ordem para integrar no tempo. Escreva um programa computacional para obter a solução. Aumente o valor de $\Delta\bar{t}$ em 10% para cada passo no tempo para obter mais rapidamente a solução estacionária e escolha valores de $\Delta\bar{x}$ e $\Delta\bar{t}$ para uma boa acurácia. Trace a temperatura adimensional em função dos comprimentos adimensionais para diversos valores do tempo adimensional.

30.14 O problema com o fluxo de calor radial transiente em uma barra circular na forma adimensional é descrito por

$$\frac{\partial^2 u}{\partial \bar{r}^2} + \frac{1}{\bar{r}}\frac{\partial u}{\partial \bar{r}} = \frac{\partial u}{\partial \bar{t}}$$

Condições de contorno	$u(1, \bar{t}) = 1$	$\frac{\partial u}{\partial \bar{t}}(0, \bar{t}) = 0$
Condições iniciais	$u(\bar{x}, 0) = 0$	$0 \le \bar{x} \le 1$

Resolva a equação de condução de calor radial transiente adimensional em uma barra circular para a distribuição de temperatura em diversos instantes de tempo, conforme a temperatura da barra se aproxima do estado estacionário. Use análogos por diferenças finitas com acurácia de segunda ordem para as derivadas com uma formulação de Crank-Nicolson. Escreva um programa computacional para a solução. Escolha valores de $\Delta\bar{r}$ e $\Delta\bar{t}$ para uma boa acurácia. Trace a temperatura u em função do raio $\bar{r}$ para diversos instantes $\bar{t}$.

30.15 Resolva a seguinte EDP:

$$\frac{\partial^2 u}{\partial x^2} + b\frac{\partial u}{\partial x} = \frac{\partial u}{\partial t}$$

Condições de contorno	$u(0, t) = 0$	$u(1, t) = 0$
Condições iniciais	$u(x, 0) = 0$	$0 \le x \le 1$

Use análogos por diferenças finitas com acurácia de segunda ordem para as derivadas com uma formulação de Crank-Nicolson para integrar no tempo. Escreva um programa computacional para a solução. Aumente o valor de Δt em 10% para cada passo no tempo para obter mais rapidamente a solução estacionária e escolha valores

de Δx e Δt para uma boa acurácia. Trace u em função de x para diversos valores de t. Resolva para valores de $b = 4, 2, 0, -2, -4$.

30.16 Determine a temperatura ao longo de uma barra horizontal de 1 m descrita pela equação de condução de calor (Equação 30.1). Suponha que a extremidade direita seja isolada e que a condição de contorno na extremidade esquerda ($x = 0$) seja representada por

$$-k' \frac{\partial T}{\partial x}\bigg|_{x=0} = h\,(T_a - T_0)$$

onde k' é o coeficiente de condutividade térmica (W/m · °C), h é o coeficiente de transferência de calor convectivo ((W/m^2 · °C), T_a é a temperatura ambiente (ºC) e T_0 é a temperatura da barra em $x = 0$ (°C). Determine a temperatura como uma função do tempo, usando um passo espacial de $\Delta x = 1$ cm e os seguintes valores dos parâmetros: $k = 2 \times 10^{-5}$ m^2/s, $k' = 10$ W/m · °C, $h = 25$ W/m^2 · °C e $T_a = 50$ °C. Suponha que a temperatura inicial na barra seja zero.

CAPÍTULO 31

Método dos Elementos Finitos

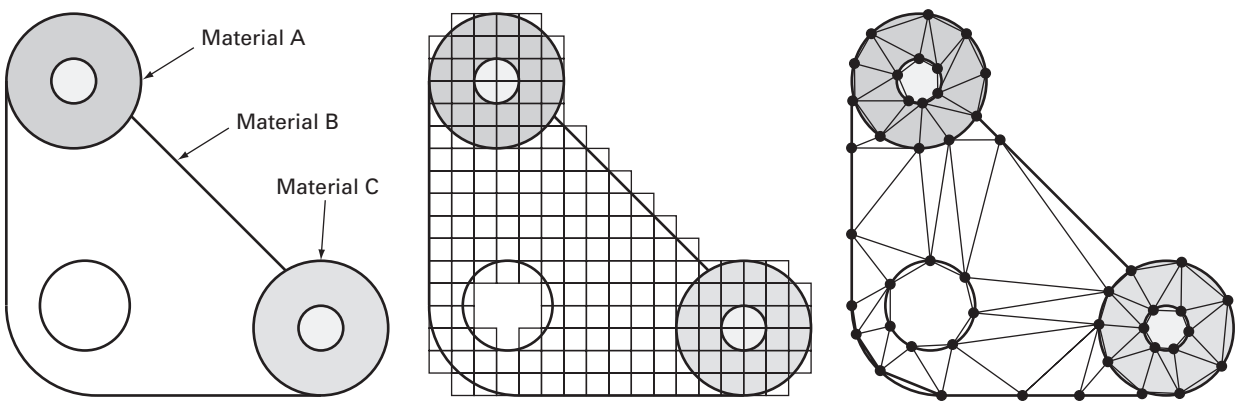
Até este momento, utilizamos métodos de *diferenças finitas* para resolver equações diferenciais parciais. Nesses métodos, o domínio da solução é dividido em uma grade de pontos discretos ou nós (Figura 31.1*b*). A EDP é então escrita para cada nó e suas derivadas substituídas por diferença divididas finitas. Embora tal aproximação "ponto a ponto" seja conceitualmente fácil de entender, ela tem várias desvantagens. Em particular, ela se torna mais difícil de aplicar para sistemas de geometria irregular, condições de contorno não usuais ou composição heterogênica.

O método dos *elementos finitos* fornece uma alternativa que é mais adequada para tais sistemas. Em contraste com as técnicas de diferenças finitas, o método dos elementos finitos divide o domínio da solução em regiões de forma simples ou "elementos" (Figura 31.1*c*). Uma solução aproximada para a EDP pode então ser desenvolvida para cada um desses elementos. A solução total é gerada juntando ou "montando" as soluções individuais, tomando cuidado para garantir a continuidade nas fronteiras entre os elementos. Logo, a EDP é satisfeita em cada parte.

Como na Figura 31.1*c*, o uso de elementos em vez de uma grade retangular fornece uma aproximação muito melhor para sistemas com formas irregulares. Além disso, os valores das incógnitas podem ser gerados continuamente por todo o domínio da solução em vez de em pontos isolados.

Como uma discrição abrangente está além do escopo deste livro, este capítulo fornece uma introdução geral do método dos elementos finitos. Nosso objetivo principal é fazer que você se sinta confortável com a abordagem e que conhece seus recursos. Nesse espírito, a seção a seguir é dedicada a uma visão geral dos passos envolvidos na solução típica por elementos finitos de um problema. A seguir, apresentamos um exemplo

FIGURA 31.1
(*a*) Uma junta de vedação com geometria irregular e composição não homogênica. (*b*) Tal sistema é muito difícil de modelar com a abordagem de diferenças finitas. Isso se deve ao fato de que são necessárias aproximações complicadas nas bordas do sistema e nas fronteiras entre regiões de composição diferente. (*c*) Uma discretização por elementos finitos é muito mais adequada.



(*a*) (*b*) (*c*)

simples: uma barra aquecida unidimensional estacionária. Embora esse exemplo não envolva EDPs, ele nos permite desenvolver e ilustrar os aspectos principais da abordagem por elementos finitos sem a presença de fatores complicadores. Podemos então discutir algumas questões envolvidas no uso do método de elementos finitos para EDPs.

31.1 A ABORDAGEM GERAL

Embora as particularidades variem, a implementação da abordagem por elementos finitos, em geral, segue um procedimento padrão passo a passo. O material a seguir fornece uma breve visão geral de cada um desses passos. A aplicação desses passos a contexto de problemas de engenharia será desenvolvida nas seções subseqüentes.

31.1.1 Discretização

Este passo envolve a divisão do domínio da solução em elementos finitos. A Figura 31.2 fornece exemplos dos elementos usados em uma, duas e três dimensões. Os pontos de interseção das retas que delimitam os lados dos elementos são conhecidos como nós e os lados propriamente ditos são chamados *retas* ou *planos nodais*.

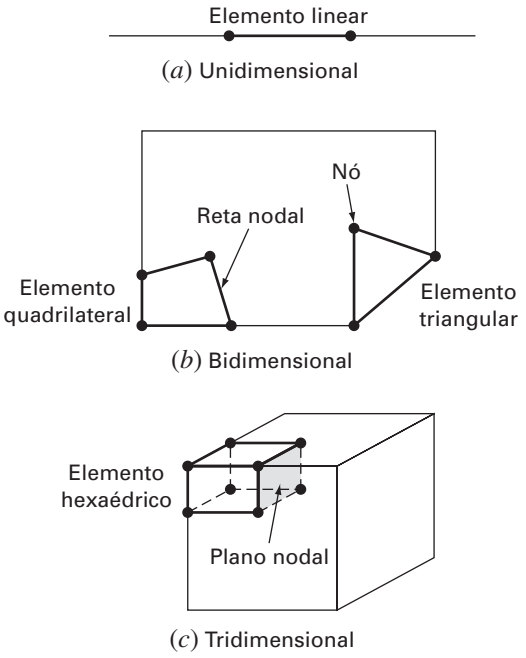
31.1.2 Equações dos Elementos

O próximo passo é desenvolver equações para aproximar as soluções para cada elemento. Isso envolve dois passos. Primeiro, devemos escolher uma função apropriada com coeficientes desconhecidos que será usada para aproximar a solução. Segundo, calculamos os coeficientes de modo que a função aproxime a solução de uma forma ótima.

Escolha das Funções de Aproximação. Como são fáceis de manipular matematicamente, os polinômios são empregados para esse propósito. Para o caso unidimensional, a alternativa mais simples é um polinômio do primeiro grau ou uma reta,

$$u(x) = a_0 + a_1 x \tag{31.1}$$

FIGURA 31.2
Exemplos dos elementos empregados em (*a*) uma, (*b*) duas e (*c*) três dimensões.



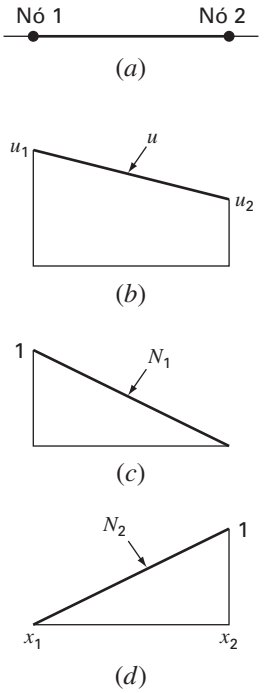


FIGURA 31.3
(*b*) Uma aproximação linear ou função de forma para (*a*) um elemento linear. As funções interpoladoras correspondentes são mostradas em (*c*) e (*d*).

onde $u(x)$ é a variável dependente a_0 e a_1 são constantes e x é a variável independente. Essa função deve passar pelos valores de $u(x)$ nas extremidades do elemento em x_1 e x_2. Portanto,

$$u_1 = a_0 + a_1 x_1$$
$$u_2 = a_0 + a_1 x_2$$

onde $u_1 = u(x_1)$ e $u_2 = u(x_2)$. Essas equações podem ser resolvidas usando a regra de Cramer por

$$a_0 = \frac{u_1 x_2 - u_2 x_1}{x_2 - x_1} \qquad a_1 = \frac{u_2 - u_1}{x_2 - x_1}$$

Esses resultados podem então ser substituídos na Equação (31.1), o que, depois de agrupar os termos, pode ser escrita como

$$u = N_1 u_1 + N_2 u_2 \tag{31.2}$$

onde

$$N_1 = \frac{x_2 - x}{x_2 - x_1} \tag{31.3}$$

e

$$N_2 = \frac{x - x_1}{x_2 - x_1} \tag{31.4}$$

A Equação (31.2) é chamada *função de aproximações* ou *de forma* e N_1 e N_2 são denominadas *funções interpoladoras*. Uma inspeção cuidadosa revela que a Equação (31.2) é, na realidade, o polinômio interpolador de Lagrange de primeiro grau. Ele fornece um meio de prever os valores intermediários (isto é, interpolar) entre os valores dados de u_1 e u_2 entre os nós.

A Figura 31.3 mostra a função de forma juntamente com as funções interpoladoras correspondentes. Observe que a soma das funções interpoladoras é igual a 1.

Além disso, o fato de estarmos lidando com equações lineares facilita operações, tais como diferenciação e integração. Tais manipulações serão importantes em seções posteriores. A derivada da Equação (31.2) é

$$\frac{du}{dx} = \frac{dN_1}{dx} u_1 + \frac{dN_2}{dx} u_2 \tag{31.5}$$

De acordo com as Equações (31.3) e (31.4), as derivadas dos *N's* podem ser calculadas como

$$\frac{dN_1}{dx} = -\frac{1}{x_2 - x_1} \qquad \frac{dN_2}{dx} = \frac{1}{x_2 - x_1} \tag{31.6}$$

e, portanto, a derivada de u é

$$\frac{du}{dx} = \frac{1}{x_2 - x_1}(-u_1 + u_2) \tag{31.7}$$

Em outras palavras, ela é uma diferença dividida representando a inclinação da reta ligando os nós.

A integral pode ser expressa como

$$\int_{x_1}^{x_2} u \, dx = \int_{x_1}^{x_2} N_1 u_1 + N_2 u_2 \, dx$$

Cada termo do lado direito é simplesmente a integral de um triângulo retângulo com base $x_2 - x_1$ e altura u. Isto é,

$$\int_{x_1}^{x_2} Nu\, dx = \frac{1}{2}(x_2 - x_1)u$$

Logo, a integral toda é

$$\int_{x_1}^{x_2} u\, dx = \frac{u_1 + u_2}{2}(x_2 - x_1) \tag{31.8}$$

Em outras palavras, ela é simplesmente a regra do trapézio.

Obtendo um Ajuste Ótimo da Função à Solução. Uma vez que a função interpoladora seja escolhida, a equação que controla o comportamento do elemento deve ser desenvolvida. Essa equação representa um ajuste da função à solução da equação diferencial subjacente. Diversos métodos estão disponíveis para esse propósito. Entre os mais comuns estão a abordagem direta, o método dos resíduos ponderados e a abordagem variacional. A saída de todos esses métodos é análoga ao ajuste de curvas. Entretanto, em vez de ajustar funções aos dados, esses métodos especificam relações entre as incógnitas na Equação (31.2) que satisfazem a EDP subjacente de uma forma ótima.

Matematicamente, as equações dos elementos resultantes freqüentemente consistirão em um conjunto de equações algébricas lineares que podem ser expressas na forma matricial,

$$[k]\{u\} = \{F\} \tag{31.9}$$

onde $[k]$ é uma *propriedade do elemento* ou a *matriz de rigidez*, $\{u\}$ é um vetor coluna de incógnitas no nó e $\{F\}$ é um vetor coluna que reflete o efeito de qualquer influência externa aplicada aos nós. Observe que, em alguns casos, as equações podem ser não-lineares. Entretanto, para os exemplos elementares descritos a seguir e para muitos problemas práticos, os sistemas são lineares.

31.1.3 Montagem

Depois que as equações dos elementos individuais são deduzidas, elas podem ser agrupadas ou montadas para caracterizar o comportamento unificado do sistema inteiro. O processo de montagem é governado pelo conceito de continuidade. Isto é, as soluções para elementos vizinhos são equiparadas, de modo que os valores das incógnitas (e, algumas vezes, das derivadas) nos seus nós comuns sejam iguais. Assim, a solução total será contínua.

Quando todas as versões individuais da Equação (31.9) forem finalmente montadas, o sistema inteiro será expresso na forma matricial como

$$[K]\{u'\} = \{F'\} \tag{31.10}$$

onde $[K]$ é a *matriz de propriedade de montagem* e $\{u'\}$ e $\{F'\}$ são vetores coluna para as incógnitas e as forças externa que são marcados com linhas para denotar que eles são uma montagem dos vetores $\{u\}$ e $\{F\}$ dos elementos individuais.

31.1.4 Condições de Contorno

Antes da Equação (31.10) poder ser resolvida, ela deve ser modificada para levar em conta as condições de contorno do sistema. Esses ajustes resultam em

$$[\bar{k}]\{u'\} = \{\bar{F}'\} \tag{31.11}$$

em que as barras significam que as condições de contorno foram incorporadas.

31.1.5 Solução

As soluções da Equação (31.11) podem ser obtidas com as técnicas descritas anteriormente na Parte Três, como a decomposição *LU*. Em muitos casos, os elementos podem ser configurados de modo que as equações resultantes sejam de banda. Assim, esquemas de solução altamente eficientes disponíveis para tais sistemas podem ser empregados.

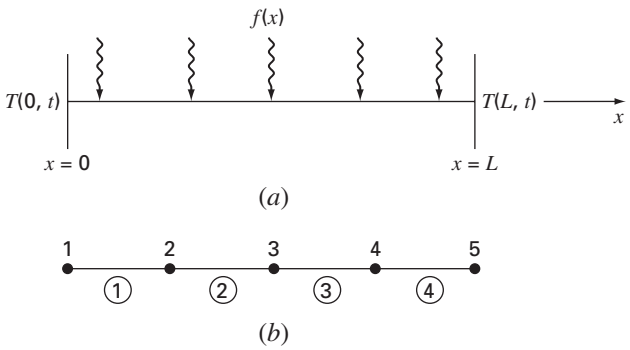


FIGURA 31.4
(*a*) Uma barra longa e fina sujeita a condições de contorno fixas e a uma fonte de calor contínua ao longo do seu eixo. (*b*) A representação em elementos finitos que consiste em quatro elementos de comprimentos iguais e a cinco nós.

31.1.6 Pós-processamento

Após obter-se a solução, ela pode ser exibida em uma forma de tabela ou graficamente. Além disso, variáveis secundárias podem ser determinadas e exibidas.

Embora os passo anteriores sejam muito gerais, são comuns à maioria das implementações da abordagem por elementos finitos. Na seção seguinte, ilustramos como podem ser aplicados para obter resultados numéricos para um sistema físico simples — uma barra aquecida.

31.2 APLICAÇÃO DE ELEMENTOS FINITOS EM UMA DIMENSÃO

A Figura 31.4 mostra um sistema que pode ser modelado por uma forma unidimensional da equação de Poisson

$$\frac{d^2T}{dx^2} = -f(x) \tag{31.12}$$

onde $f(x)$ é uma função que define uma fonte de calor ao longo da barra e em que as extremidades da barra estão mantidas a temperaturas fixas,

$$T(0, t) = T_1$$

e

$$T(L, t) = T_2$$

Observe que essa não é uma equação diferencial parcial, mas sim uma EDO com condição de contorno. Esse modelo simples é usado porque ele nos permitirá introduzir a abordagem por elementos finitos sem algumas das complicações envolvidas em, por exemplo, uma EDP bidimensional.

EXEMPLO 31.1 Solução Analítica para Barra Aquecida

Enunciado do Problema. Resolva a Equação (31.12) para uma barra de 10 cm com condições de contorno de $T(0, t) = 40$ e $T(10, t) = 200$ e uma fonte de calor uniforme $f(x) = 10$.

Solução. A equação a ser resolvida é

$$\frac{d^2T}{dx^2} = -10$$

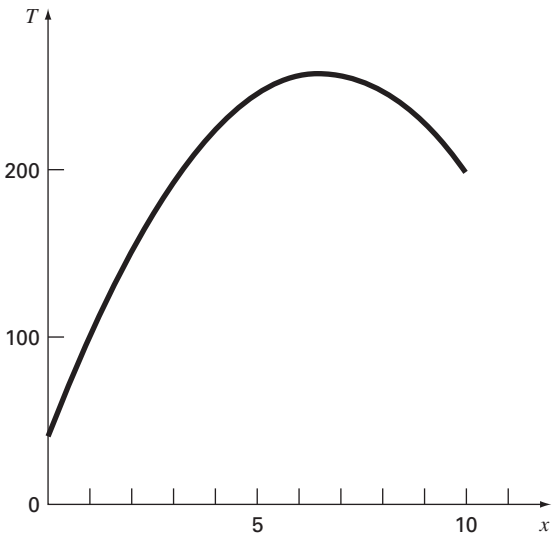


FIGURA 31.5
A distribuição de temperatura ao longo de uma barra aquecida sujeita a uma fonte de calor uniforme e mantida a temperaturas fixas nas extremidades.

Suponha que a solução tenha a forma

$$T = ax^2 + bx + c$$

a qual pode ser diferenciada duas vezes para fornecer $T'' = 2a$. Substituindo esse resultado na equação diferencial, obtemos $a = -5$. As condições de contorno podem ser usadas para calcular os coeficientes remanescentes. Para a primeira condição em $x = 0$,

$$40 = -5(0)^2 + b(0) + c$$

ou $c = 40$. Analogamente, para a segunda condição,

$$200 = -5(10)^2 + b(10) + 40$$

que pode ser usada para determinar que $b = 66$. Portanto, a solução final é

$$T = -5x^2 + 66x + 40$$

Os resultados estão traçados na Figura 31.5.

31.2.1 Discretização

Uma configuração simples para modelar o sistema é uma série de elementos de comprimentos iguais (Figura 31.4*b*). Assim o sistema é tratado como quatro elementos de comprimentos iguais e cinco nós.

31.2.2 Equações dos Elementos

Um elemento individual é mostrado na Figura 31.6*a*. A distribuição de temperatura para o elemento pode ser representado pela função de aproximação

$$\tilde{T} = N_1 T_1 + N_2 T_2 \tag{31.13}$$

onde N_1 e N_2 são funções interpoladoras lineares especificadas pelas Equações (31.3) e (31.4), respectivamente. Assim, como descrito na Figura 31.6*b*, a função de aproximação resulta em uma interpolação linear entre as temperaturas dos dois nós.

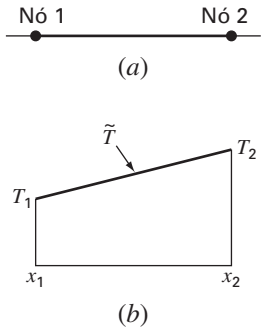


FIGURA 31.6
(*a*) Elemento individual. (*b*) A função de aproximação usada para caracterizar a distribuição de temperatura ao longo do elemento.

Como observado na Seção 31.1, existem diversas abordagens para o desenvolvimento das equações dos elementos. Nesta seção, empregaremos duas delas. Em primeiro lugar, a *abordagem direta* será usada para o caso simples no qual $f(x) = 0$. A seguir, por causa de sua aplicabilidade para o caso geral de engenharia, dedicaremos a maior parte da seção ao *método dos resíduos ponderados*.

Abordagem Direta. Para o caso no qual $f(x) = 0$, pode ser empregado o método direto para gerar as equações dos elementos. A relação entre fluxo de calor e gradiente de temperatura pode ser representada pela lei de Fourier:

$$q = -k' \frac{dT}{dx}$$

onde q é o fluxo [cal/(cm^2 · s)] e k' é o coeficiente de condutividade térmica [cal/(s · cm · °C)]. Se uma função de aproximação linear for utilizada para caracterizar a temperatura do elemento, o fluxo de calor entrando no elemento pelo nó *1* pode ser representado por

$$q_1 = k' \frac{T_1 - T_2}{x_2 - x_1}$$

onde q_1 é o fluxo de calor no nó 1. De forma análoga, para o nó 2,

$$q_2 = k' \frac{T_2 - T_1}{x_2 - x_1}$$

Essas duas equações expressam a relação da distribuição da temperatura interna do elemento (como refletido nas temperaturas nodais) com o fluxo de calor em suas extremidades. Dessa forma, elas constituem nossas equações dos elementos desejadas. E podem ser ainda mais simplificadas reconhecendo-se que a lei de Fourier pode ser usada para expressar os próprios fluxos nas extremidades em termos dos gradientes das temperaturas nas bordas. Isto é,

$$q_1 = -k' \frac{dT(x_1)}{dx} \qquad q_2 = k' \frac{dT(x_2)}{dx}$$

as quais podem ser substituídas na equações dos elementos para fornecer

$$\frac{1}{x_2 - x_1} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} = \begin{Bmatrix} -\dfrac{dT(x_1)}{dx} \\ \dfrac{dT(x_2)}{dx} \end{Bmatrix} \tag{31.14}$$

Observe que a Equação (31.14) foi colocada na forma da Equação (31.9). Portanto, conseguimos gerar uma equação matricial que descreve o comportamento de um elemento típico em nosso sistema.

A abordagem direta tem grande apelo intuitivo. Além disso, em áreas como a mecânica, ela pode ser empregada para resolver problemas significativos. Em outros contextos, em geral é difícil ou impossível deduzir equações dos elementos finitos diretamente. Conseqüentemente, como descreveremos a seguir, técnicas matemáticas mais gerais estão disponíveis.

Método dos Resíduos Ponderados. A equação diferencial (31.12) pode ser reescrita como

$$0 = \frac{d^2T}{dx^2} + f(x)$$

A solução aproximada [Equação (31.13)] pode ser substituída nessa equação. Como a Equação (31.13) não é a solução exata, o lado direito da equação resultante não será zero, mas sim igual a um resíduo,

$$R = \frac{d^2\tilde{T}}{dx^2} + f(x) \tag{31.15}$$

O *método dos resíduos ponderados* (MRP) consiste em encontrar o mínimo para o

resíduo de acordo com a fórmula geral

$$\int_D RW_i \, dD = 0 \qquad i = 1, 2, \ldots, m \tag{31.16}$$

onde D é domínio da solução e W_i são as funções peso linearmente independentes.

Neste ponto, existem diversas escolhas que poderiam ser feita para as funções peso (Quadro 31.1). A abordagem mais comum para o método dos elementos finitos é empregar as funções interpoladoras N_i como funções peso. Quando estas são substituídas na Equação (31.16), o resultado é conhecido com método de Galerkin,

$$\int_D RN_i \, dD = 0 \qquad i = 1, 2, \ldots, m$$

Para barra unidimensional, a Equação (31.15) pode ser substituída nessa equação para fornecer

$$\int_{x_1}^{x_2} \left[\frac{d^2\tilde{T}}{dx^2} + f(x)\right] N_i \, dx \qquad i = 1, 2$$

a qual pode ser reescrita como

$$\int_{x_1}^{x_2} \frac{d^2\tilde{T}}{dx^2} N_i(x) \, dx = -\int_{x_1}^{x_2} f(x) N_i(x) \, dx \qquad i = 1, 2 \tag{31.17}$$

Neste ponto, diversas manipulações matemáticas serão aplicadas para simplificar e calcular a Equação (31.17). Entre as mais importantes está a simplificação do lado esquerdo usando integração por partes. Lembre-se do cálculo que essa operação pode ser expressa, em geral, como

$$\int_a^b u \, dv = uv\Big|_a^b - \int_a^b v \, du$$

Se u e v forem escolhidas apropriadamente, a nova integral no lado direito será mais

Quadro 31.1 Esquemas de Resíduos Alternativos para o MRP

Diversas escolhas podem ser feitas para funções peso na Equação (31.16). Cada uma representa uma abordagem alternativa para o MRP.

Na *abordagem da colocação*, escolhemos tantas posições quantos coeficientes desconhecidos existentes. Então, os coeficientes são ajustados até que os resíduos se anulem em cada uma dessas posições. Por conseguinte, a função de aproximação produzirá resultados perfeitos nas posições escolhidas, mas terá resíduos não-nulos nos outros pontos. Assim, ela é parecida com os métodos de interpolação do Capítulo 18. Observe que a *colocação* se resume a usar as funções peso

$$W = \delta(x - x_i) \qquad \text{para } i = 1, 2, \ldots, n$$

em que n é o número de coeficientes desconhecidos e $\delta(x - x_i)$ é a *função delta de Dirac* que se anula em todos os pontos, exceto em $x = x_i$, em que é igual a infinito.

No *método do subdomínio*, o intervalo é subdividido em tantos segmentos, ou "subdomínios", quantos os coeficientes desconhecidos existentes. Então, os coeficientes são ajustados até que o valor médio do resíduo seja nulo em cada subdomínio. Logo, para cada subdomínio, a função peso é igual a 1 e a Equação (31.16) é

$$\int_{x_{i-1}}^{x_i} R \, dx = 0 \qquad \text{para } i = 1, 2, \ldots, n$$

onde x_{i-1} e x_i são as extremidades do subdomínio.

No caso dos *mínimos quadrados*, os coeficientes são ajustados de modo a minimizar a integral dos quadrados dos resíduos. Assim, as funções peso são

$$W_i = \frac{\partial R}{\partial a_i}$$

que podem ser substituídas na Equação (31.16) para fornecer

$$\int_D R \frac{\partial R}{\partial a_i} \, dD = 0 \qquad i = 1, 2, \ldots, n$$

ou

$$\frac{\partial}{\partial a_i} \int_D R^2 \, dD = 0 \qquad i = 1, 2, \ldots, n$$

A comparação dessa formulação com aquelas do Capítulo 17 mostra que isso é a forma contínua da regressão.

O método de Galerkin emprega as funções interpoladoras N_i como funções peso. Lembre-se de que a soma dessas funções é sempre 1 em qualquer posição de um elemento. Para muitos contextos de problemas, o método de Galerkin produz os mesmos resultados obtidos por métodos variacionais. Conseqüentemente, ele é a versão mais comumente usada do MRP na análise de elementos finitos.

fácil de calcular que a original no lado esquerdo. Isso pode ser feito para o termo no lado esquerdo da Equação (31.17) escolhendo $N_i(x)$ como u e $(d^2T/dx^2)\,dx$ como dv para produzir

$$\int_{x_1}^{x_2} N_i(x)\,\frac{d^2\tilde{T}}{dx^2}\,dx = N_i(x)\frac{d\tilde{T}}{dx}\bigg|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{d\tilde{T}}{dx}\,\frac{dN_i}{dx}\,dx \qquad i = 1, 2 \tag{31.18}$$

Logo, demos um passo significativo em baixar a ordem do termo de maior ordem na formulação de uma segunda para uma primeira derivada.

A seguir, podemos calcular os termos individuais que criamos na Equação (31.18). Para $i = 1$, o primeiro termo no lado direito da Equação (31.18) pode ser calculado como

$$N_1(x)\frac{d\tilde{T}}{dx}\bigg|_{x_1}^{x_2} = N_1(x_2)\,\frac{d\tilde{T}(x_2)}{dx} - N_1(x_1)\,\frac{d\tilde{T}(x_1)}{dx}$$

Entretanto, lembre-se, a partir da Figura 31.3, de que $N_1(x_2) = 0$ e $N_1(x_1) = 1$, e, portanto,

$$N_1(x)\frac{d\tilde{T}}{dx}\bigg|_{x_1}^{x_2} = -\frac{d\tilde{T}(x_1)}{dx} \tag{31.19}$$

De modo similar, para $i = 2$,

$$N_2(x)\,\frac{d\tilde{T}}{dx}\bigg|_{x_1}^{x_2} = \frac{d\tilde{T}(x_2)}{dx} \tag{31.20}$$

Portanto, o primeiro termo no lado direito da Equação (31.18) representa as condições de contorno naturais nas extremidades dos elementos.

Agora, antes de prosseguir, vamos substituir nossos resultados de volta na equação original. Substituindo-se as Equações (31.18) a (31.20) na Equação (31.17) e reorganizando, obtemos, para $i = 1$,

$$\int_{x_1}^{x_2} \frac{d\tilde{T}}{dx}\,\frac{dN_1}{dx}\,dx = -\frac{d\tilde{T}(x_1)}{dx} + \int_{x_1}^{x_2} f(x)N_1(x)\,dx \tag{31.21}$$

e, para $i = 2$,

$$\int_{x_1}^{x_2} \frac{d\tilde{T}}{dx}\,\frac{dN_2}{dx}\,dx = \frac{d\tilde{T}(x_2)}{dx} + \int_{x_1}^{x_2} f(x)N_2(x)\,dx \tag{31.22}$$

Observe que a integração por partes levou a dois resultados importantes. Primeiro, ela incorporou as condições de contorno diretamente nas equações dos elementos. Segundo, ela abaixou o cálculo de ordem mais alto de uma segunda para uma primeira derivada. Esse último resultado produz a conseqüência significativa que as funções de aproximação precisam preservar a continuidade do valor, mas não da inclinação nos nós.

Observe também que, agora, podemos começar a associar algum significado físico aos termos individuais que deduzimos. No lado direito de cada equação, o primeiro termo representa uma das condições de contorno do elemento do termo e o segundo é o efeito — forçante do sistema — no caso presente, a fonte de calor $f(x)$. Como se tornará evidente, o lado esquerdo engloba mecanismos internos que regem a distribuição de temperaturas no elemento. Isto é, em termos do método de elementos finitos, o lado esquerdo se tornará a matriz de propriedades do elemento.

Para ver isso, vamos nos concentrar nos termos no lado esquerdo. Para $i = 1$, o termo é

$$\int_{x_1}^{x_2} \frac{d\tilde{T}}{dx}\,\frac{dN_1}{dx}\,dx \tag{31.23}$$

Lembre-se, a partir da Seção 31.1.2, de que a natureza linear da função de forma torna a derivação e a integração simples. Substituindo-se as Equações (31.6) e (31.7) na Equação (31.23), temos

$$\int_{x_2}^{x_1} \frac{T_1 - T_2}{(x_2 - x_1)^2}\, dx = \frac{1}{x_1 - x_2}(T_1 - T_2) \tag{31.24}$$

Substituições análogas para $i = 2$ [Equação (31.22)] fornece

$$\int_{x_2}^{x_1} \frac{-T_1 + T_2}{(x_2 - x_1)^2}\, dx = \frac{1}{x_1 - x_2}(-T_1 + T_2) \tag{31.25}$$

A comparação com a Equação (31.14) mostra que essas relações são parecidas com as que foram desenvolvidas com o método direto, usando a lei de Fourier. Isso pode ser tornado ainda mais claro reescrevendo as Equações (31.24) e (31.25) na forma matricial como

$$\frac{1}{x_2 - x_1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix}$$

Substituindo esse resultado nas Equações (31.21) e (31.22) e expressando o resultado na forma matricial, obtemos a versão final das equações dos elementos

$$\underbrace{\frac{1}{x_2 - x_1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\{T\}}_{\text{Matriz de rigidez do elemento}} = \underbrace{\begin{Bmatrix} -\dfrac{dT(x_1)}{dx} \\ \dfrac{dT(x_2)}{dx} \end{Bmatrix}}_{\substack{\text{Condição} \\ \text{de contorno}}} + \underbrace{\begin{Bmatrix} \displaystyle\int_{x_1}^{x_2} f(x)N_1(x)\, dx \\ \displaystyle\int_{x_1}^{x_2} f(x)N_2(x)\, dx \end{Bmatrix}}_{\text{Efeitos externos}} \tag{31.26}$$

Observe que, além dos métodos direto e dos resíduos ponderados, as equações dos elementos também podem ser deduzidas usando-se o cálculo variacional (por exemplo, ver Allaire, 1985). Para o caso presente, essa abordagem produz equações que são idênticas às deduzidas acima.

EXEMPLO 31.2 Equação de Elemento para a Barra Aquecida

Enunciado do Problema. Empregue a Equação (31.26) para desenvolver as equações dos elementos para uma barra de 10 cm com condições de contorno $T(0, t) = 40$ e $T(10, t) = 200$ e uma fonte de calor uniforme $f(x) = 10$. Use quatro elementos de comprimentos iguais a 2,5 cm.

Solução. O termo de fonte de calor na primeira linha da Equação (31.26) pode ser calculado utilizando a Equação (31.3) e integrando para obter

$$\int_0^{2,5} 10\, \frac{2,5 - x}{2,5}\, dx = 12,5$$

Analogamente, a Equação (31.4) pode ser substituída no termo de fonte de calor da segunda linha da Equação (31.26), que também pode ser integrado para fornecer

$$\int_0^{2,5} 10\, \frac{x - 0}{2,5}\, dx = 12,5$$

Esses resultados juntamente com os outros valores dos parâmetros podem ser substituídos na Equação (31.26) para fornecer

$$0,4T_1 - 0,4T_2 = -\frac{dT}{dx}(x_1) + 12,5$$

e

$$-0,4T_1 + 0,4T_2 = \frac{dT}{dx}(x_2) + 12,5$$

31.2.3 Montagem

Antes das equações dos elementos serem montados, deve ser estabelecido um sistema de enumeração global para especificar a topologia ou a distribuição espacial do sistema. Na Tabela 31.1, isso define a conectividade dos elementos da grade. Como o caso presente é unidimensional, o esquema de enumeração pode parecer tão previsível que se torna trivial. Entretanto, para problemas bi e tridimensionais, ele oferece a única maneira de especificar quais nós pertencem a quais elementos.

Uma vez que a topologia seja especificada, a equação de elemento (31.26) pode ser escrita para cada elemento usando-se coordenadas globais. Então, elas podem ser adicionadas uma de cada vez para montar a matriz do sistema total (observe que esse processo é explorado mais na Seção 32.4). O processo é descrito na Figura 31.7.

TABELA 31.1 A topologia do sistema para o esquema de segmentação por elementos finitos da Figura 31.4*b*.

	Números de nós	
Elemento	**Local**	**Global**
1	1	1
	2	2
2	1	2
	2	3
3	1	3
	2	4
4	1	4
	2	5

FIGURA 31.7
A montagem das equações para o sistema total.

$$(a)\quad \begin{bmatrix} 0{,}4 & -0{,}4 & 0 & 0 & 0 \\ -0{,}4 & 0{,}4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ 0 \\ 0 \\ 0 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12{,}5 \\ dT(x_2)/dx + 12{,}5 \\ 0 \\ 0 \\ 0 \end{Bmatrix}$$

$$(b)\quad \begin{bmatrix} 0{,}4 & -0{,}4 & 0 & 0 & 0 \\ -0{,}4 & 0{,}4 & +0{,}4 & -0{,}4 & 0 \\ 0 & & -0{,}4 & 0{,}4 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ 0 \\ 0 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12{,}5 \\ 12.5 + 12{,}5 \\ dT(x_3)/dx + 12{,}5 \\ 0 \\ 0 \end{Bmatrix}$$

$$(c)\quad \begin{bmatrix} 0{,}4 & -0{,}4 & 0 & 0 & 0 \\ -0{,}4 & 0{,}8 & -0{,}4 & 0 & 0 \\ 0 & -0{,}4 & 0{,}4 & +0{,}4 & -0{,}4 \\ 0 & 0 & & -0{,}4 & 0{,}4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ 0 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12{,}5 \\ 25 \\ 12{,}5 + 12{,}5 \\ dT(x_4)/dx + 12{,}5 \\ 0 \end{Bmatrix}$$

$$(d)\quad \begin{bmatrix} 0{,}4 & -0{,}4 & 0 & 0 & 0 \\ -0{,}4 & 0{,}8 & -0{,}4 & 0 & 0 \\ 0 & -0{,}4 & 0{,}8 & -0{,}4 & 0 \\ 0 & 0 & -0{,}4 & 0{,}4 & +0{,}4 & -0{,}4 \\ 0 & 0 & 0 & & -0{,}4 & 0{,}4 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12{,}5 \\ 25 \\ 25 \\ 12{,}5 + 12{,}5 \\ dT(x_5)/dx + 12{,}5 \end{Bmatrix}$$

$$(e)\quad \begin{bmatrix} 0{,}4 & -0{,}4 & 0 & 0 & 0 \\ -0{,}4 & 0{,}8 & -0{,}4 & 0 & 0 \\ 0 & -0{,}4 & 0{,}8 & -0{,}4 & 0 \\ 0 & 0 & -0{,}4 & 0{,}8 & -0{,}4 \\ 0 & 0 & 0 & -0{,}4 & 0{,}4 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \end{Bmatrix} = \begin{Bmatrix} -dT(x_1)/dx + 12{,}5 \\ 25 \\ 25 \\ 25 \\ dT(x_5)/dx + 12{,}5 \end{Bmatrix}$$

31.2.4 Condições de Contorno

Observe que, conforme as equações são montadas, as condições de contorno internas se cancelam. Logo, o resultado final para $\{F\}$ na Figura 31.7*e* tem condições de contorno apenas para o primeiro e o último nó. Como T_1 e T_5 são dados, essas condições de contorno naturais nas extremidades da barra, $dT(x_1)/dx$ e $dT(x_5)/dx$, representam incógnitas. Portanto, as equações podem ser representadas como

$$\begin{array}{cccccc} \dfrac{dT}{dx}(x_1) & -0{,}4T_2 & & & = & -3{,}5 \\ 0{,}8T_2 & -0{,}4T_3 & & & = & 41 \\ -0{,}4T_2 & +0{,}8T_3 & -0{,}4T_4 & & = & 25 \\ & -0{,}4T_3 & +0{,}8T_4 & & = & 105 \\ & & -0{,}4T_4 & -\dfrac{dT}{dx}(x_5) & = & -67{,}5 \end{array} \tag{31.27}$$

31.2.5 Solução

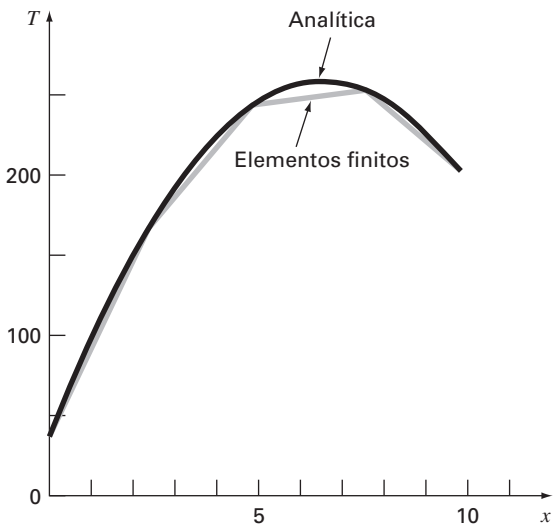
A Equação (31.17) pode ser resolvida por

$$\frac{dT}{dx}(x_1) = 66 \qquad T_2 = 173{,}75 \qquad T_3 = 245$$
$$T_4 = 253{,}75 \qquad \frac{dT}{dx}(x_5) = -34$$

31.2.6 Pós-processamento

Os resultados podem ser mostrados graficamente. A Figura 31.8 mostra os resultados por elementos finitos juntamente com a solução exata. Observe que os cálculos por elementos finitos capturam a tendência geral da solução exata e, na realidade, fornecem uma coincidência exata nos nós. Entretanto, existe uma discrepância nos pontos interiores de cada elemento em virtude da natureza linear das funções de forma.

FIGURA 31.8
Resultados da aplicação da abordagem por elementos finitos na barra aquecida. A solução exata também é mostrada.



31.3 PROBLEMAS BIDIMENSIONAIS

Embora a "contabilidade" matemática aumente acentuadamente, a extensão da abordagem por elementos finitos para duas dimensões é conceitualmente parecida com as aplicações unidimensionais discutidas até este ponto. Portanto, ela segue os mesmos passos que foram delineados na Seção 31.1.

31.3.1 Discretização

Diversos elementos simples, tais como triângulos e quadriláteros, são geralmente empregados para a grade de elementos finitos em duas dimensões. Na discussão presente, vamos nos limitar a elementos triangulares do tipo descrito na Figura 31.9.

31.3.2 Equações dos Elementos

Exatamente como no caso unidimensional, o próximo passo é desenvolver uma equação para aproximar a solução para o elemento. Para um elemento triangular, a abordagem mais simples é o polinômio linear [compare com a Equação (31.1)]

$$u(x, y) = a_0 + a_{1,1}x + a_{1,2}y \tag{31.28}$$

onde $u(x, y)$ é a variável dependente, os a's são os coeficientes e x e y são as variáveis independentes. Essa função deve passar pelos valores de $u(x, y)$ nos nós do triângulo (x_1, y_1), (x_2, y_2) e (x_3, y_3). Portanto,

$$\begin{aligned} u_1(x, y) &= a_0 + a_{1,1}x_1 + a_{1,2}y_1 \\ u_2(x, y) &= a_0 + a_{1,1}x_2 + a_{1,2}y_2 \\ u_3(x, y) &= a_0 + a_{1,1}x_3 + a_{1,2}y_3 \end{aligned}$$

ou, na forma matricial,

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_{1,1} \\ a_{1,2} \end{Bmatrix} = \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix}$$

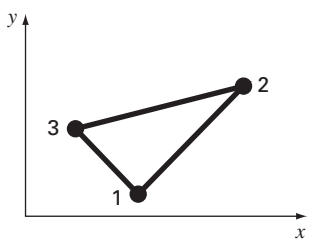
a qual pode ser resolvida por

$$a_0 = \frac{1}{2A_e}[u_1(x_2y_3 - x_3y_2) + u_2(x_3y_1 - x_1y_3) + u_3(x_1y_2 - x_2y_1)] \tag{31.29}$$

$$a_{1,1} = \frac{1}{2A_e}[u_1(y_2 - y_3) + u_2(y_3 - y_1) + u_3(y_1 - y_2)] \tag{31.30}$$

$$a_{1,2} = \frac{1}{2A_e}[u_1(x_3 - x_2) + u_2(x_1 - x_3) + u_3(x_2 - x_1)] \tag{31.31}$$

FIGURA 31.9
Um elemento triangular.



onde A_e é a área do elemento triangular,

$$A_e = \frac{1}{2}[(x_2y_3 - x_3y_2) + (x_3y_1 - x_1y_3) + (x_1y_2 - x_2y_1)]$$

As Equações (31.29) a (31.31) podem ser substituídas na Equação (31.28). Depois de agrupar as termos, o resultado pode ser expresso como

$$u = N_1u_1 + N_2u_2 + N_3u_3 \tag{31.32}$$

onde

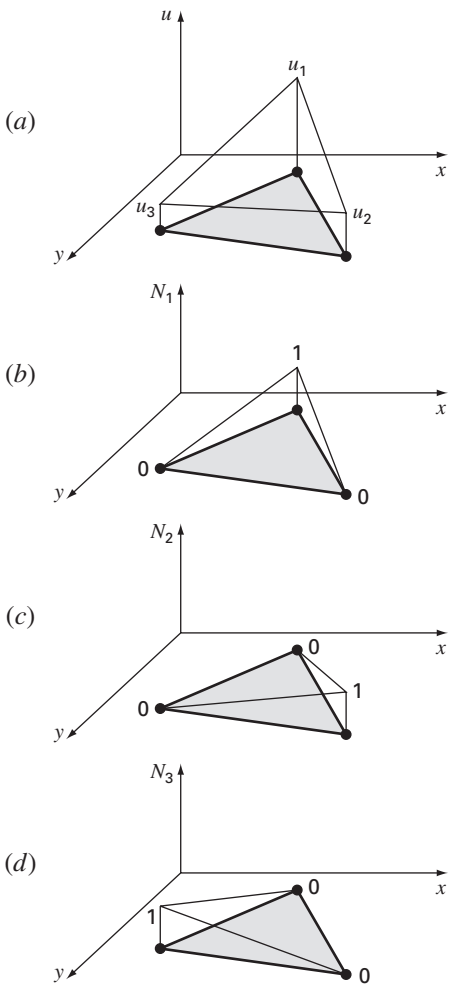
$$N_1 = \frac{1}{2A_e}[(x_2y_3 - x_3y_2) + (y_2 - y_3)x + (x_3 - x_2)y]$$

$$N_2 = \frac{1}{2A_e}[(x_3y_1 - x_1y_3) + (y_3 - y_1)x + (x_1 - x_3)y]$$

$$N_3 = \frac{1}{2A_e}[(x_1y_2 - x_2y_1) + (y_1 - y_2)x + (x_2 - x_1)y]$$

A Equação (31.32) fornece um meio de prever valores intermediários para o elemento com base nos valores em seus nós. A Figura 31.10 mostra a função de forma juntamente com as funções interpoladoras correspondentes. Observe que a soma das funções interpoladoras é sempre igual a 1.

FIGURA 31.10
(*a*) Uma função de aproximação linear para um elemento triangular. As funções interpoladoras correspondentes são mostradas em (*b*) a (*d*).



Como no caso unidimensional, diversos métodos estão disponíveis para desenvolver as equações dos elementos com base na EDP subjacente e nas funções de aproximação. As equações resultantes são consideravelmente mais complicadas que a Equação (31.26). Entretanto, como as funções de aproximação em geral são polinômios de grau baixo como a Equação (31.28), os termos da matriz de elementos final consistirá em polinômios de baixo grau e de constantes.

31.3.3 Condições de Contorno e Montagem

A incorporação das condições de contorno e da montagem da matriz do sistema também se torna muito mais complicadas quando a técnica de elementos finitos é aplicada a problemas bi e tridimensionais. Entretanto, como na dedução da matriz de elementos, a dificuldade se relaciona com a mecânica do processo em vez da complexidade conceitual.

FIGURA 31.11
Um esquema de enumeração para os nós e os elementos de uma aproximação por elementos finitos da placa aquecida que foi anteriormente caracterizada por diferenças finitas no Capítulo 29.

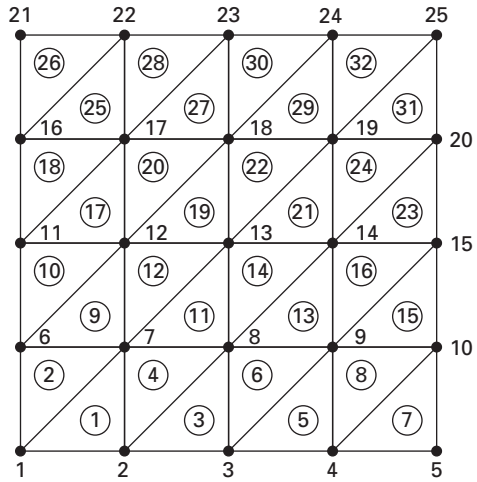
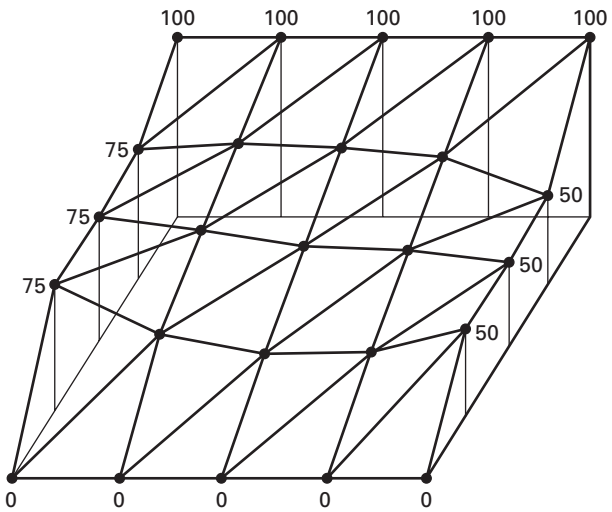


FIGURA 31.12
A distribuição de temperaturas de uma placa aquecida quando calculada com o método de elementos finitos.



Por exemplo, o estabelecimento da topologia do sistema, que era trivial para o caso unidimensional, se torna uma questão de grande importância em duas e três dimensões. Em particular, a escolha do esquema de enumeração determinará se a matriz do sistema resultante será ou não de banda e, portanto, a eficiência com a qual pode ser resolvida. A Figura 31.11 mostra um esquema que foi desenvolvido para a placa aquecida anteriormente resolvida pelos métodos de diferenças finitas no Capítulo 29.

31.3.4 Solução e Pós-processamento

Embora a mecânica seja complicada, a matriz do sistema é simplesmente um conjunto de n equações simultâneas que podem ser usadas para determinar os valores da variável dependente nos n nós. A Figura 31.12 traz uma solução que corresponde à solução por diferenças finitas da Figura 29.5.

31.4 RESOLVENDO EDPS COM BIBLIOTECAS E PACOTES

As bibliotecas e pacotes de software têm alguns recursos para resolver diretamente EDPs. Entretanto, como descrito nas seções seguintes, muitas das soluções são limitadas a problemas simples. Isso é particularmente verdade nos casos bi e tridimensionais. Para essas situações, pacotes genéricos (isto é, os que não são expressamente desenvolvidos para resolver EDPs, tais como os pacotes de elementos finitos), são freqüentemente limitados a domínios retangulares simples.

Ainda que isso possa parecer uma limitação, as aplicações simples podem ser de grande utilidade em um sentido pedagógico. Isso é particularmente verdade quando as ferramentas de visualização dos pacotes são usadas para exibir o resultado dos cálculos.

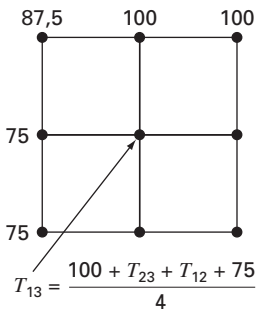
31.4.1 Excel

Embora o Excel não tenha um recurso direto para resolver EDPs, é um bom ambiente para desenvolver soluções de EDPs elípticas simples. Por exemplo, a disposição ortogonal das células da planilha (Figura 31.13*b*) é diretamente análoga à grade no Capítulo 29 para modelar a placa aquecida (Figura 31.13*a*).

Como na Figura 31.13*b*, as condições de contorno de Dirichlet podem ser inseridas inicialmente na periferia do bloco de células. A fórmula para o método de Liebman pode ser implementada inserindo a Equação (29.11) em uma das células no interior (como a célula B2 na Figura 31.13*b*). Logo, o valor para a célula pode ser calculado como uma função de suas células adjacentes. Então, a célula pode ser copiada para as outras células interiores. Por causa da natureza relativa do comando de cópia do Excel, todas as outras células serão apropriadamente dependentes de suas células adjacentes.

Uma vez que tenha copiado a fórmula, você vai obter uma mensagem de erro: **Cannot resolve circular references**. Você pode resolver isso indo para o menu *T*(ools) e escolhendo ***O*(ptions)**. A seguir, escolha a **Calculation tab** e clique no quadro de **Iteration**.

FIGURA 31.13
A analogia entre (*a*) uma grade retangular e (*b*) as células da planilha.



(*a*) Grade

	A	B	C	D	E
1	87,5	100	100	100	75
2	75	78,57	76,12	69,64	50
3	75	63,17	56,25	52,46	50
4	75	42,86	33,26	33,93	50
5	37,5	0	0	0	25

$B2 = \frac{B1 + C2 + B3 + A2}{4}$

(*b*) Planilha

Isso permitirá que a planilha recalcule (o padrão é cem iterações) e resolva o método de Liebman iterativamente. Após isso ocorrer, aperte a tecla F9 para recalcular manualmente a planilha até que as respostas não variem. Isso significa que a solução convergiu.

Uma vez que o problema tenha sido resolvido, as ferramentas gráficas do Excel podem ser utilizadas para visualizar os resultados. Um exemplo é mostrado na Figura 31.14*a*. Para esse caso,

- Usamos uma grade mais fina.
- Mantivemos a borda inferior isolada.
- Adicionamos uma fonte de calor de 150 ao meio da placa (célula E5).

Os resultados numéricos da Figura 31.14*a* podem então ser exibidos com o Chart Wizard do Excel. A Figura 31.14*b* e *c* exibem os gráficos da superfície 3D. A orientação entre *y* destes são em geral o inverso da planilha. Assim, a aresta de cima com alta temperatura (100) seria normalmente mostrada na parte de baixo do gráfico. Invertemos os valores de *y* na nossa planilha antes de traçar os gráficos de forma que eles fossem consistentes com a planilha.

Observe como os gráficos ajudam você a visualizar o que está acontecendo. O calor flui para baixo a partir da fonte de calor em direção às bordas formando a figura de uma montanha. O calor também flui das bordas de alta temperatura para as duas arestas laterais. Observe como o calor flui preferencialmente em direção à aresta de temperatura mais baixa (50). Finalmente, observe como o gradiente de temperatura na dimensão *y* vai a zero na aresta inferior isolada ($\partial T/\partial y \rightarrow 0$).

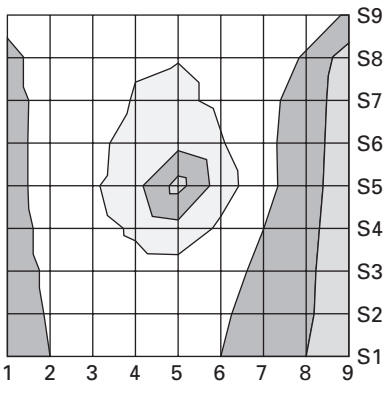
31.4.2 MATLAB

Embora o pacote de software MATLAB padrão não tenha atualmente grandes recursos para resolver EDPs, certamente podem ser desenvolvidos M-files e funções para esse propósito. Além disso, seus recursos de visualização são muito bons, particularmente para a visualização de problemas em duas dimensões espaciais.

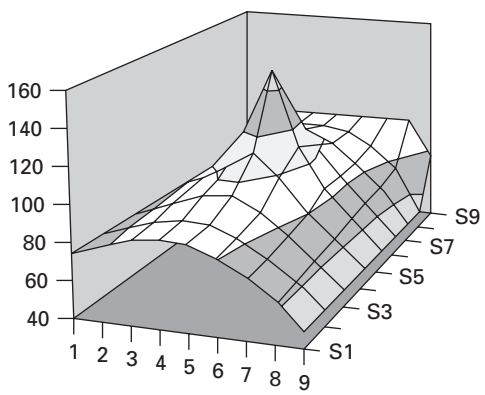
FIGURA 31.14
(*a*) Solução do Excel para a equação de Poisson para uma placa com a aresta inferior isolada e uma fonte de calor. (*b*) Um "mapa topográfico" e (*c*) um gráfico 3D das temperaturas.

	A	B	C	D	E	F	G	H	I
1	87,5	100,0	100,0	100,0	100,0	100,0	100,0	100,0	75,0
2	75,0	89,2	95,8	99,1	99,7	96,6	89,9	77,6	50,0
3	75,0	86,2	94,7	100,9	103,1	96,7	85,5	70,3	50,0
4	75,0	85,7	96,1	106,7	115,3	101,4	85,2	68,2	50,0
5	75,0	85,5	97,4	114,3	150,0	108,6	85,6	67,3	50,0
6	75,0	84,0	93,4	103,4	111,6	97,4	81,3	65,6	50,0
7	75,0	82,2	88,9	94,2	95,6	88,1	76,6	63,6	50,0
8	75,0	80,9	85,9	88,9	88,4	82,8	73,5	62,2	50,0
9	75,0	80,4	84,9	87,3	86,3	81,1	72,4	61,7	50,0

(*a*)



(*b*)



(*c*)

Para ilustrar esse recurso, vamos inicialmente montar um planilha Excel na Figura 31.14*a*. Esses resultados podem ser salvos como um arquivo de texto (delimitado por Tab) com um nome como **plate.txt**. Esse arquivo pode então ser movido para o diretório do MATLAB.

Uma vez no MATLAB, o arquivo pode ser carregado digitando-se

```
>> load plate.txt
```

A seguir, os gradientes podem ser calculados de forma simples por

```
>> [px,py]=gradient(plate);
```

Observe que isso é o método mais simples de calcular gradientes, usando valores padrão de dx = dy = 1. Portanto, as direções e os módulos relativos estarão corretos.

Finalmente, pode ser usada uma série de comandos para desenvolver um gráfico. O comando **contour** desenvolve um gráfico de contorno dos dados. O comando **clabel** adiciona legendas nos contornos do gráfico. Por fim, o comando **quiver** toma os dados do gradiente e os adiciona ao gráfico como setas,

```
>> cs=contour(plate);clabel(cs);hold on
>> quiver(-px,-py);hold off
```

Observe que os sinais de menos são adicionados por causa do sinal de menos na lei de Fourier [Equação (29.4)]. Como visto na Figura 31.15, o gráfico resultante fornece uma representação excelente da solução.

Observe que qualquer arquivo em formato adequado pode ser inserido no MATLAB e exibido dessa forma. Por exemplo, os cálculos IMSL, descritos a seguir, poderiam ser programados para gerar um arquivo para exibição no MATLAB (ou mesmo no Excel). Esse compartilhamento de arquivos entre as ferramentas está se tornando comum. Além disso, podem ser criados arquivos em uma posição em uma ferramenta, transmitidos pela Internet para outra posição, na qual o arquivo possa ser exibido com outra ferramenta. Este é um dos aspectos excitantes das aplicações numéricas modernas.

FIGURA 31.15
Gráficos de contorno gerados pelo MATLAB para a placa aquecida calculada com o Excel (Figura 31.14).

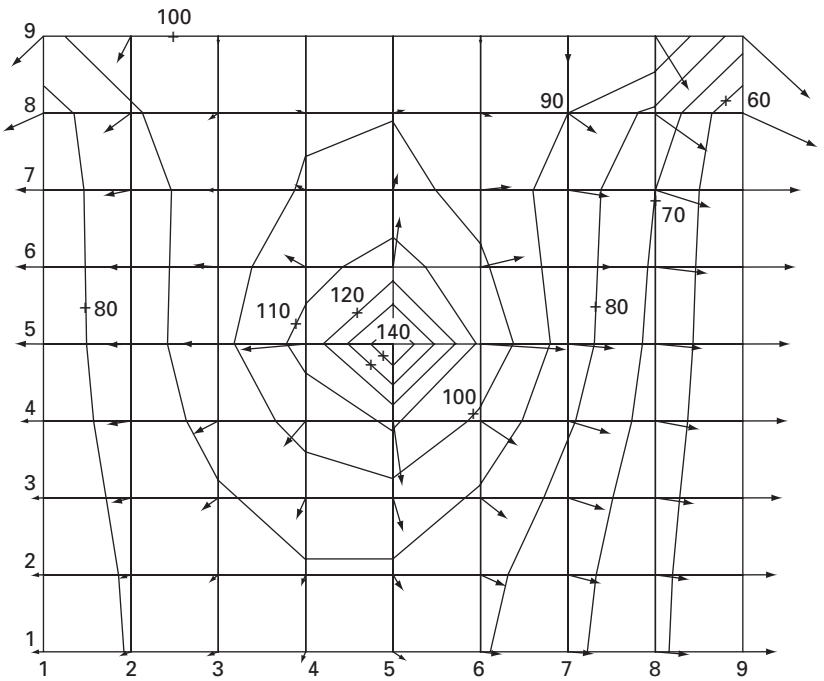


TABELA 31.2 Rotinas IMSL para resolver EDPs.

Categoria	Rotinas	Recursos
Solução de sistemas de EDPs em uma dimensão	MOLCH	Método das linhas com uma base cúbica de Hermite
Solução de uma EDP em duas e três dimensões	FPS2H	Ferramenta para a solução rápida de Poisson bidimensional
	FPS3H	Ferramenta para a solução rápida de Poisson tridimensional

31.4.3 IMSL

O IMSL tem algumas poucas rotinas para a resolução de EDPs (Tabela 31.2). Na discussão presente, nos concentraremos na rotina **fps2h**. Essa rotina resolve a equação de Poisson ou de Helmholtz em um retângulo bidimensional, utilizando uma ferramenta de resolução rápida de Poisson em uma grade uniforme.

A sub-rotina **fps2h** é implementada pelo seguinte comando CALL:

```
CALL FPS2H(PRH,BRH,COEF,NX,NY,AX,BX,AY,BY,IBCT,IORD,U,LDU)
```

onde

PRH = FUNCTION fornecida pelo usuário para calcular o lado direito da equação diferencial parcial. A forma é PRH(X, Y), em que
X = valor da coordena X. (Entrada)
Y = valor da coordena Y. (Entrada)
PRH deve ser declarada EXTERNAL no programa principal.

BRH = FUNCTION fornecida pelo usuário para calcular o lado direito das condições de contorno.
A forma é BRHS(ISIDE, X, Y), em que
ISIDE = Número de lados. (Entrada) Ver IBCTY a seguir para a definição de números de lados.
X = valor da coordena X. (Entrada)
Y = valor da coordena Y. (Entrada)
BRH deve ser declarada EXTERNAL no programa principal.

COEF = Valor do coeficiente de U na equação diferencial. (Entrada)

NX = Número de retas na grade na direção X. (Entrada) NX deve ser pelo menos 4. Ver Comentário 2 para mais restrições em NX.

NY = Número de retas na grade na direção Y. (Entrada) NY deve ser pelo menos 4. Ver Comentário 2 para mais restrições em NY.

AX = O valor de X ao longo do lado esquerdo do domínio. (Entrada)

BX = O valor de X ao longo do lado direito do domínio. (Entrada)

AY = O valor de Y ao longo do lado de baixo do domínio. (Entrada)

BY = O valor de Y ao longo do lado de cima do domínio. (Entrada)

IBCT = Vetor de tamanho 4 indicando o tipo de condição de contorno em cada lado do domínio ou que a solução é periódica. (Entrada) Os lados são numerados de 1 a 4 como segue:

Lado	Posição
1—Direito	(X = BX)
2—De baixo	(Y = AY)
3—Esquerdo	(X = AX)
4—De cima	(Y = BY)

Existem três tipos de condição de contorno.

IBCTY	Condição de contorno
1	São dados os valores de U. (Dirichlet)
2	São dados os valores de dU/dX (lados 1 e/ou 3). (Neumann) São dados os valores de dU/dY (lados 2 e/ou 4).
3	Periódica.

IORD = Ordem de acurácia da aproximação por diferenças finitas. (Entrada) Pode ser 2 ou 4. Geralmente, é utilizado IORD = 4.

U = Matriz de tamanho NX por NY contendo a solução nos pontos da grade. (Saída)

LDU = Dimensão dominante de U exatamente como especificado na declaração de dimensão no programa principal. (Entrada)

EXEMPLO 31.3 Usando o IMSL para Determinar a Temperatura de uma Placa Aquecida

Enunciado do Problema. Use fps2h para determinar as temperaturas da placa quadrada com condições de contorno fixas do Exemplo 29.1.

Solução. Um exemplo de um programa principal em Fortran 90 e de uma função usando o fps2h para resolver esse problema pode ser escrito como

```
Program Plate
USE msimsl
IMPLICIT NONE
INTEGER ::ncval, nx, nxtabl, ny, nytabl
PARAMETER (ncval=11, nx=33, nxtabl=5, ny=33, nytabl=5)
INTEGER :: i, ibcty(4), iorder, j, nout
REAL :: QD2VL,ax,ay,brhs,bx,by,coefu,prhs,u(nx,ny),utabl,x,xdata(nx),y,ydata(ny)
EXTERNAL brhs, prhs
ax = 0.0
bx = 40.
ay = 0.0
by = 40.
ibcty(1) = 1
ibcty(2) = 1
ibcty(3) = 1
ibcty(4) = 1
coefu = 0.0
iorder = 4
CALL FPS2H(prhs,brhs,coefu,nx,ny,ax,bx,ay,by,ibcty,iorder,u,nx)
DO i=1, nx
  xdata(i) = ax + (bx-ax)*FLOAT(i-1)/FLOAT(nx-1)
END DO
DO j=1, ny
  ydata(j) = ay + (by-ay)*FLOAT(j-1)/FLOAT(ny-1)
END DO
CALL UMACH (2, nout)
WRITE (nout,'(8X,A,11X,A,11X,A)') 'X', 'Y', 'U'
DO j=1, nytabl
  DO i=1, nxtabl
    x     = ax + (bx-ax)*FLOAT(i-1)/FLOAT(nxtabl-1)
    y     = ay + (by-ay)*FLOAT(j-1)/FLOAT(nytabl-1)
    utabl = QD2VL(x,y,nx,xdata,ny,ydata,u,nx,.FALSE.)
    WRITE (nout,'(4F12.4)') x, y, utabl
  END DO
END DO
END PROGRAM
```

```
FUNCTION prhs (x, y)
IMPLICIT NONE
REAL :: prhs, x, y
prhs = 0.0
END FUNCTION

REAL FUNCTION brhs (iside, x, y)
IMPLICIT NONE
INTEGER :: iside
REAL :: x, y
IF (iside == 1) then
  brhs = 50.
ELSEIF (iside == 2) THEN
  brhs = 0.
ELSEIF (iside == 3) THEN
  brhs = 75.
ELSE
  brhs = 100.
END IF
END FUNCTION
```

Um exemplo de execução produz a seguinte saída:

```
x        y         u         x         y         u
   .0000     .0000   37.5000   30.0000   20.0000    52.3849
 10.0000     .0000     .0000   40.0000   20.0000    50.0000
 20.0000     .0000     .0000     .0000   30.0000    75.0000
 30.0000     .0000     .0000   10.0000   30.0000    79.0032
 40.0000     .0000   25.0000   20.0000   30.0000    76.8058
   .0000   10.0000   75.0000   30.0000   30.0000    69.9017
 10.0000   10.0000   42.5976   40.0000   30.0000    50.0000
 20.0000   10.0000   32.2945     .0000   40.0000    87.5000
 30.0000   10.0000   33.4962   10.0000   40.0000   100.0000
 40.0000   10.0000   50.0000   20.0000   40.0000   100.0000
   .0000   20.0000   75.0000   30.0000   40.0000   100.0000
 10.0000   20.0000   63.5128   40.0000   40.0000    75.0000
 20.0000   20.0000   56.2493
```

PROBLEMAS

31.1 Repita o Exemplo 31.1, mas para $T(0, t) = 75$ e $T(10, t) = 150$ e uma fonte de calor uniforme de 15.

31.2 Repita o Exemplo 31.2, porém para condições de contorno de $T(0, t) = 75$ e $T(10, t) = 150$ e uma fonte de calor de 15.

31.3 Aplique os resultados do Problema 31.2 para calcular a distribuição de temperaturas para a barra toda, usando a abordagem por elementos finitos.

31.4 Use o método de Galerkin para desenvolver uma equação de elemento para uma versão estacionária da equação de advecção-difusão descrita no Problema 30.7. Expresse o resultado final no formato da Equação (31.26) de modo que cada termo tenha uma interpretação física.

31.5 Uma versão da equação de Poisson que ocorre na mecânica é o seguinte modelo para a deflexão vertical de uma barra com uma carga distribuída $P(x)$:

$$A_c E \frac{\partial^2 u}{\partial x^2} = P(x)$$

em que A_c é a área da secção transversal, E é o módulo de Young, u é a deflexão e x é a distância ao longo do comprimento da barra. Se a barra for rigidamente fixa ($u = 0$) em ambas as extremidades, use o método dos elementos finitos para modelar sua deflexão para $A_c = 0{,}1$ m^2, $E = 200 \times 10^9$ N/m^2, $L = 10$ m, e $P(x) = 1000$ N/m. Use o valor de $\Delta x = 2$ m.

31.6 Desenvolva um programa amigável ao usuário para modelar a distribuição estacionária de temperaturas em uma barra com uma fonte de calor constante, utilizando o método dos elementos finitos. Faça o programa de modo que nós desigualmente espaçados possam ser usados.

31.7 Use o Excel para fazer os mesmos cálculos que na Figura 31.14, mas com a extremidade à direita isolada e adicione um sumidouro de calor de -150 na célula C7.

31.8 Use o MATLAB para desenvolver um gráfico de contorno com setas de fluxo para a solução do Excel para o Problema 31.7.

31.9 Use o Excel para modelar a distribuição de temperatura da laje mostrada na Figura P31.9. A laje tem 0,02 m de espessura e possui uma condutividade térmica de 3 W/(m · ° C).

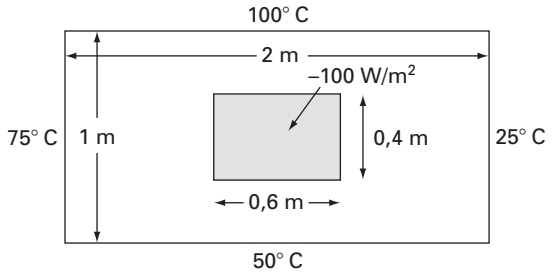


Figura P31.9

31.10 Use o MATLAB para desenvolver um gráfico de contorno com setas de fluxo para a solução do Excel para o Problema 31.9.
31.11 Use o IMSL para fazer os mesmos cálculos que no Exemplo 31.3, mas isole a aresta inferior da placa.
31.12 Encontre a distribuição de temperatura em uma barra (Figura P31.12) com geração de calor interna usando o método de elementos finitos. Deduza as equações nodais dos elementos usando a condução de calor de Fourier

$$q_k = -kA\frac{dT}{\partial x}$$

e a relação de conservação de calor

$$\sum [q_k + f(x)] = 0$$

onde q_k é o fluxo de calor (W), k é a condutividade térmica (W/(m · °C) A é a área da secção transversal (m^2) e $f(x)$ é a fonte de calor (W/cm). A barra tem um valor de $kA = 100$ W m/° C. A barra tem 50 cm de comprimento, a coordenada x é zero na extremidade esquerda e positiva para a direita. Divida a barra em cinco elementos (seis nós, cada um com 10 cm de comprimento). A extremidade esquerda da barra tem gradiente de temperatura fixo e a temperatura é variável. A extremidade direita tem uma temperatura fixa e o gradiente é uma variável. A fonte de calor $f(x)$ tem um valor constante. Logo, as condições são

$$\left.\frac{dT}{\partial x}\right|_{x=0} = 0{,}25^\circ\text{C/m} \qquad T|_{x=50} = 100^\circ\text{C} \qquad f(x) = 30 \text{ W/cm}$$

Desenvolva as equações nodais que precisam ser resolvidas para as temperaturas e os gradientes de temperatura em cada um dos seis nós. Monte as equações, insira as condições de contorno e resolva o conjunto resultante para determinar as incógnitas.

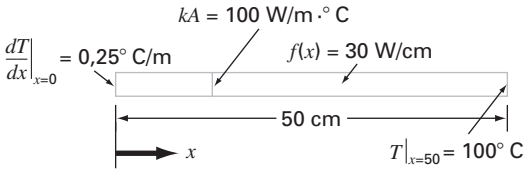


Figura P31.12

31.13 Encontre a distribuição de temperatura para uma barra (Figura P31.13) com geração de calor interno, usando o método dos elementos finitos. Deduza as equações nodais dos elementos, utilizando a condição de calor de Fourier

$$q_k = -kA\frac{dT}{\partial x}$$

e as relações de conservação de calor

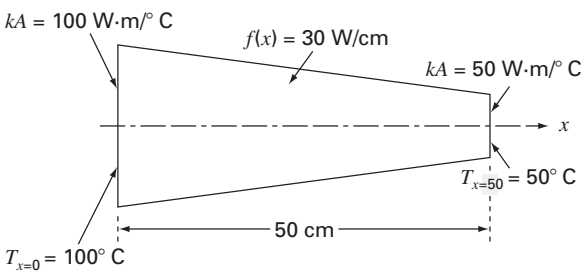
$$\sum [q_k + f(x)] = 0$$

onde q_k é o fluxo de calor (W), k é a condutividade térmica (W/(m · °C), A é a área da secção transversal (m^2) e $f(x)$ é a fonte de calor (W/cm). A barra tem 50 cm de comprimento, a coordenada x é zero na extremidade esquerda e positiva na direita. A barra também é linearmente afunilada com um valor de $kA = 100$ e 50 W m/° C em $x = 0$ e em $x = 50$, respectivamente. Divida a barra em cinco elementos (seis nós, cada um com 10 cm de comprimento). Ambas as extremidades da barra têm temperatura fixas. A fonte de calor $f(x)$ tem um valor constante. Logo, as condições são

$$T|_{x=0} = 100^\circ\text{C} \qquad T|_{x=50} = 50^\circ\text{C} \qquad f(x) = 30 \text{ W/cm}$$

As áreas afuniladas devem ser tratadas como se fossem constantes no comprimento de um elemento. Portanto, calcule a média dos valores de kA em cada extremidade do elemento e tome a média como o valor constante no elemento. Desenvolva as equações nodais que devem ser resolvidas para determinar a temperatura e o gradiente de temperatura em cada um dos seis nós. Monte as equações, insira as condições de contorno e resolva o conjunto resultante para determinar as icógnitas.

Figura P31.13



CAPÍTULO 32

Estudo de Casos: Equações Diferenciais Parciais

O propósito deste capítulo é aplicar os métodos da Parte Oito em problemas da engenharia prática. Na *Seção 32.1*, uma EDP parabólica é usada para calcular a distribuição variável no tempo de um composto químico ao longo de eixo longitudinal de um reator retangular. Esse exemplo ilustra como a instabilidade de uma solução pode ser devida a natureza da EDP, em vez das propriedades do método numérico.

As Seções 32.2 e 32.3 envolvem aplicações das equações de Poisson e de Laplace a problemas de engenharia civil e elétrica, respectivamente. Entre outras coisas, isso vai lhe permitir perceber as semelhanças bem como as diferenças entre problemas de campo nessas áreas da engenharia. Além disso, eles podem ser contrastados com o problema da placa aquecida que serviu como nosso sistema protótipo nesta parte do livre. A *Seção 32.2* trata da deflexão de uma placa quadrada, enquanto a *Seção 32.3* é dedicada ao cálculo da distribuição de voltagem e fluxo de carga em uma superfície bidimensional com bordas curvas.

A *Seção 32.4* apresenta uma análise por elementos finitos aplicada a uma série de molas. Essa aplicação está mais próxima em espírito das aplicações de elementos finitos em mecânicas e estruturas do que do problema do campo de temperaturas usado para ilustrar a abordagem no Capítulo 31.

32.1 BALANÇO DE MASSA UNIDIMENSIONAL DE UM REATOR (ENGENHARIA QUÍMICA/BIOENGENHARIA)

Fundamentos. Os engenheiros químicos fazem uso extensivo de reatores idealizados em seus trabalhos de projetos. Nas Seções 12.1 e 28.1 nos concentramos em um único reator ou em diversos reatores de mistura acoplados. Estes são exemplos de *sistemas de parâmetro distribuído* (lembre-se da Seção PT 3.1.2).

A Figura 32.1 descreve um reator alongado com um único ponto de entrada e de saída. Esse reator pode ser caracterizado como um *sistema de parâmetro distribuído*. Se for suposto que o composto químico modelado está sujeito a decaimento de primeira ordem[1] e que o tanque seja bem misturado vertical e horizontalmente, pode ser feito uma balanço de massa em um segmento finito de comprimento Δx, como em

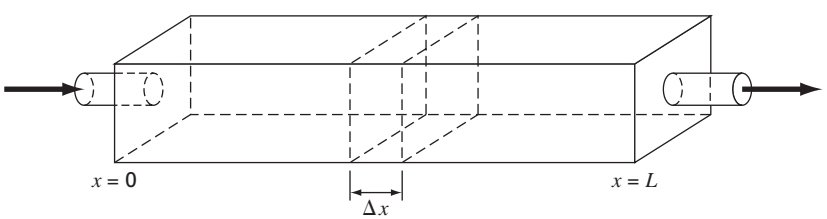


FIGURA 32.1
Um reator alongado com um único ponto de entrada e de saída. Um balanço de massa é desenvolvido em torno de um segmento finito ao longo do eixo longitudinal do tanque para deduzir uma equação diferencial para a concentração.

[1] N.R.T.: Isto é, o composto químico decai com uma taxa que é linearmente proporcional à quantidade de composto químico presente.

$$V\frac{\Delta c}{\Delta t} = \underbrace{Qc(x)}_{\text{Escoamento de entrada}} - \underbrace{Q\left[c(x) + \frac{\partial c(x)}{\partial x}\Delta x\right]}_{\text{Escoamento de saída}} - \underbrace{DA_c\frac{\partial c(x)}{\partial x}}_{\text{Dispersão para dentro}}$$

$$+ \underbrace{DA_c\left[\frac{\partial c(x)}{\partial x} + \frac{\partial}{\partial x}\frac{\partial c(x)}{\partial x}\Delta x\right]}_{\text{Dispersão para fora}} - \underbrace{kVc}_{\text{Reação de decaimento}} \tag{32.1}$$

onde V é o volume (m^3), Q é a vazão (m^3/h), c é a concentração (mols/m^3), D é coeficiente de dispersão (m^2/h), A_c é área da secção transversal do tanque (m^2) e k é o coeficiente de decaimento de primeira ordem (h^{-1}). Observe que os termos de dispersão são baseados na *primeira lei de Fick,*

$$\text{Escoamento} = -D\frac{\partial c}{\partial x} \tag{32.2}$$

que é diretamente análoga à lei de condução de calor de Fourier [lembre-se da Equação (29.4)]. Ela especifica que a mistura turbulenta tende a mover massa das regiões de alta para baixa concentração. O parâmetro D, portanto, reflete o módulo da mistura turbulenta.

Se for permitido que Δx e Δt tendam a zero, a Equação (32.1) se torna

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - U\frac{\partial c}{\partial x} - kc \tag{32.3}$$

onde $U = Q/A_c$ é a velocidade da água escoando através do tanque. O balanço de massa para a Figura 32.1 é, portanto, expresso agora como uma equação diferencial parcial parabólica. A Equação (32.3), algumas vezes, é chamada *equação de advecção-dispersão* com reação de primeira ordem. No estado estacionário, ela se reduz a uma EDO de segunda ordem,

$$0 = D\frac{d^2c}{dx^2} - U\frac{d^2c}{dx} - kc \tag{32.4}$$

Antes de $t = 0$, o tanque está cheio de água livre de compostos químicos. Em $t = 0$, o composto químico é injetado no escoamento de entrada do reator a um nível constante de c_{in}. Assim, as seguintes condições de contorno são válidas:

$$Qc_{in} = Qc_0 - DA_c\frac{\partial c_0}{\partial x}$$

e

$$c'(L, t) = 0$$

A segunda condição especifica que o composto químico deixa o reator exclusivamente como uma função do escoamento através do tubo de saída. Isto é, supõe-se que a dispersão no reator não afete a taxa de saída. Sob essas condições use métodos numéricos para resolver a Equação (32.4) para os níveis estacionários de um reator. Observe que este é um problema de contorno para uma EDO. A seguir, resolva a Equação (32.3) para caracterizar a resposta transiente — isto é, como os níveis variam com o tempo, conforme o sistema tende ao estado estacionário. Essa aplicação envolve uma EDP.

Solução. Uma solução estacionária pode ser desenvolvida substituindo-se a primeira e segunda derivadas na Equação (32.4) por diferenças finitas centradas para obter

$$0 = D\frac{c_{i+1} - 2c_i + c_{i-1}}{\Delta x^2} - U\frac{c_{i+1} - c_{i-1}}{2\Delta x} - kc_i$$

Agrupando os termos, temos

$$-\left(\frac{D}{U\Delta x} + \frac{1}{2}\right)c_{i-1} + \left(\frac{2D}{U\Delta x} + \frac{k\Delta x}{U}\right)c_0 - \left(\frac{D}{U\Delta x} - \frac{1}{2}\right)c_{i+1} = 0 \tag{32.5}$$

Essa equação pode ser escrita para cada nó do sistema. Nas extremidades do reator, esse processo introduz os nós que estão fora do sistema. Por exemplo, no nó de entrada ($i = 0$),

$$-\left(\frac{D}{U\Delta x}+\frac{1}{2}\right)c_{-1}+\left(\frac{2D}{U\Delta x}+\frac{k\Delta x}{U}\right)c_0-\left(\frac{D}{U\Delta x}-\frac{1}{2}\right)c_1=0 \tag{32.6}$$

O c_{-1} pode ser removido através da primeira condição de contorno. Na entrada, o seguinte balanço de massa deve ser satisfeito:

$$Qc_{\text{in}} = Qc_0 - DA_c\frac{\partial c_0}{\partial x}$$

onde c_0 é a concentração em $x = 0$. Logo, essa condição de contorno especifica que a quantidade de composto químico carregado para dentro do tanque pela advecção através do tubo deve ser igual à quantidade carregada para longe de entrada, tanto por advecção quanto por dispersão turbulenta no tanque. A derivada pode ser substituída por uma diferença dividida finita.

$$Qc_{\text{in}} = Qc_0 - DA_c\frac{c_1 - c_{-1}}{2\Delta x}$$

a qual pode ser resolvida por

$$c_{-1} = c_1 + \frac{2\Delta xU}{D}c_{\text{in}} - \frac{2\Delta xU}{D}c_0$$

que pode ser substituída na Equação (32.6) para fornecer

$$\left(\frac{2D}{U\Delta x}+\frac{k\Delta x}{U}+2+\frac{\Delta xU}{D}\right)c_0-\left(\frac{D}{U\Delta x}\right)c_1=\left(2+\frac{\Delta xU}{D}\right)c_{\text{in}} \tag{32.7}$$

Um exercício parecido pode ser feito na saída, na qual a equação de diferença original é

$$-\left(\frac{D}{U\Delta x}+\frac{1}{2}\right)c_{n-1}+\left(\frac{2D}{U\Delta x}+\frac{k\Delta x}{U}\right)c_n-\left(\frac{D}{U\Delta x}-\frac{1}{2}\right)c_{n+1}=0 \tag{32.8}$$

A condição de contorno na saída é

$$Qc_n - DA_c\frac{dc_n}{dx} = Qc_n$$

Como na entrada, podemos usar uma diferença dividida para aproximar a derivada.

$$Qc_n - DA_c\frac{c_{n+1} - c_{n-1}}{2\Delta x} = Qc_n \tag{32.9}$$

Uma inspeção dessa equação nos leva a concluir que $c_{n+1} = c_{n-1}$. Em outras palavras, a inclinação na saída deve ser nula para que a Equação (32.9) seja válida. A substituição desse resultado na Equação (32.8) e sua simplificação fornece

$$-\left(\frac{D}{U\Delta x}\right)c_{n-1}+\left(\frac{2D}{U\Delta x}+\frac{k\Delta x}{U}\right)c_n=0 \tag{32.10}$$

As Equações (32.5), (32.7) e (32.10) agora formam um sistema tridiagonal de n equações com n incógnitas. Por exemplo, se $D = 2$, $U = 1$, $\Delta x = 2{,}5$, $k = 0{,}2$ e $c_{\text{in}} = 100$, o sistema é

$$\begin{bmatrix} 5{,}35 & -1{,}6 & & & \\ -1{,}3 & 2{,}1 & -0{,}3 & & \\ & -1{,}3 & 2{,}1 & -0{,}3 & \\ & & -1{,}3 & 2{,}1 & -0{,}3 \\ & & & -1{,}6 & 2{,}1 \end{bmatrix}\begin{Bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{Bmatrix} = \begin{Bmatrix} 325 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}$$

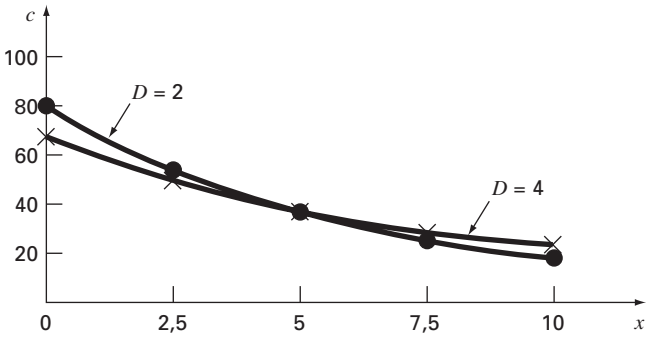


FIGURA 32.2
Concentração em função da distância ao longo do eixo longitudinal de um reator retangular para um composto químico que decai com cinética de primeira ordem.

o qual pode ser resolvido por

$$c_0 = 76{,}44 \qquad c_1 = 52{,}47 \qquad c_2 = 36{,}06$$
$$c_3 = 25{,}05 \qquad c_4 = 19{,}09$$

Esses resultados estão traçados na Figura 32.2. Como esperado, a concentração decresce em virtude da reação de decaimento, conforme o composto químico escoa através do tanque. Além dos cálculos anteriores, a Figura 32.2 mostra outro caso com $D = 4$. Observe como o aumento da mistura turbulenta tende a achatar a curva.

Em contraste, se a dispersão diminuir, a curva se tornaria mais íngreme conforme a mistura se tornasse menos importante em relação à advecção e ao decaimento. Deve ser observado que se a dispersão diminuir muito, os cálculos se tornarão sujeitos a erros numéricos. Esse tipo de erro é conhecido como *instabilidade estática,* para diferenciar da *instabilidade dinâmica* devida a um passo no tempo muito grande durante o cálculo da dinâmica. O critério para evitar essa instabilidade estática é

$$\Delta x \leq \frac{2D}{U}$$

Logo, o critério se torna mais restritivo (menor Δx) para os casos nos quais a advecção domina a dispersão.

Além dos cálculos de estado estacionário, os métodos numéricos podem ser usados para gerar soluções dependentes do tempo da Equação (32.3). A Figura 32.3 mostra resultados para $D = 2$, $U = 1$, $\Delta x = 2.5$, $k = 0{,}2$ e $c_{\text{in}} = 100$, em que a concentração no tanque é 0 no instante zero. Como esperado, o impacto imediato está próximo da entrada. Com o passar do tempo, a solução eventualmente tenderá ao nível estacionário.

Deve ser observado que, em tais cálculos da dinâmica, o passo no tempo é restrito por um critério de estabilidade expresso por (Chapra, 1997)

$$\Delta t \leq \frac{(\Delta x)^2}{2D + k(\Delta x)^2}$$

Assim, o termo de reação age para tornar o passo no tempo menor.

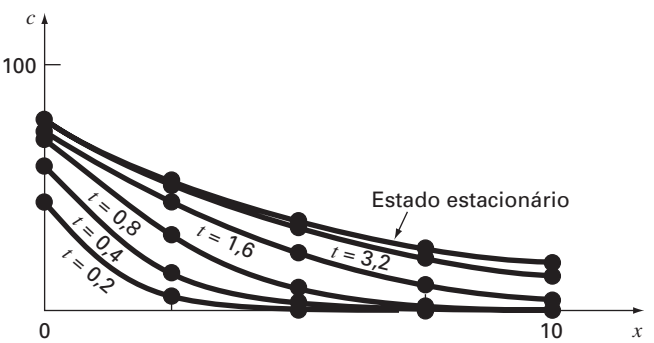


FIGURA 32.3
Concentração em função da distância em diferentes instantes durante o crescimento de um composto químico em um reator.

32.2 DEFLEXÕES EM UMA PLACA (ENGENHARIA CIVIL/AMBIENTAL)

Fundamentos. Uma placa quadrada com as arestas apoiadas está sujeita a uma carga de superfície q (Figura 32.4). A deflexão na dimensão z pode ser determinada resolvendo-se a EDP elíptica (ver Carnahan, Luther e Wilkes, 1969)

$$\frac{\partial^4 z}{\partial x^4} + 2\frac{\partial^4 z}{\partial x^2 \partial y^2} + \frac{\partial^4 z}{\partial y^4} = \frac{q}{D} \tag{32.11}$$

sujeita a condições de contorno que, nas arestas, a deflexão e a inclinação normal à borda são nulas. O parâmetro D é a rigidez de flexão,

$$D = \frac{E\,\Delta z^3}{12(1-\sigma^2)} \tag{32.12}$$

onde E é o módulo de elasticidade, Δz é a espessura da placa e σ é a razão de Poisson,

Se for definida uma nova variável como

$$u = \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2}$$

a Equação (32.11) pode ser reescrita como

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{q}{D} \tag{32.13}$$

Portanto, o problema se reduz a resolver sucessivamente duas equações de Poisson. Primeiro, a Equação (32.13) pode ser resolvida para determinar u sujeito às condições de contorno que $u = 0$ nas bordas. Então, os resultados podem ser empregados em conjunto com

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = u \tag{32.14}$$

para determinar z sujeito à condição que $z = 0$ nas bordas.

Desenvolva um programa computacional para determinar a deflexão de uma placa quadrada sujeita a uma carga de superfície constante. Teste seu programa para uma placa com bordas de 2 m de comprimento, $q = 33{,}6$ kN/m^2, $\sigma = 0{,}3$, $\Delta z = 10^{-2}$ m, e $E = 2 \times 10^{11}$ Pa. Use $\Delta x = \Delta y = 0{,}5$ m para seu teste de execução.

Solução. Diferenças divididas finitas podem ser substituídas na Equação (32.13) para produzir

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{\Delta x^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{\Delta y^2} = \frac{q}{D} \tag{32.15}$$

A Equação (32.12) pode ser usada para calcular $D = 1{,}832 \times 10^4$ N/m. Esse resultado, juntamente com os outros parâmetros do sistema, pode ser substituído na Equação (32.15) para fornecer

$$u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j} = 0{,}458$$

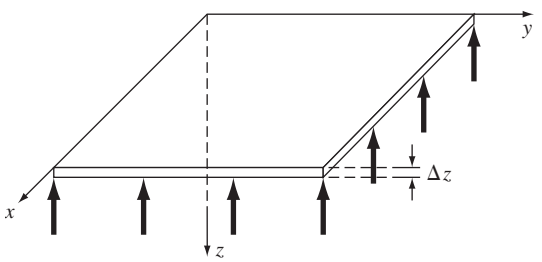


FIGURA 32.4
Uma placa quadrada apoiada sujeita a uma carga de superfície.

Essa equação pode ser escrita para todos os nós com as bordas fixas em $u = 0$. As equações resultantes são

$$\begin{bmatrix} -4 & 1 & & 1 & & & & & \\ 1 & -4 & 1 & & 1 & & & & \\ & 1 & -4 & & & 1 & & & \\ 1 & & & -4 & 1 & & 1 & & \\ & 1 & & 1 & -4 & 1 & & 1 & \\ & & 1 & & 1 & -4 & & & 1 \\ & & & 1 & & & -4 & 1 & \\ & & & & 1 & & 1 & -4 & 1 \\ & & & & & 1 & & 1 & -4 \end{bmatrix} \begin{Bmatrix} u_{1,1} \\ u_{2,1} \\ u_{3,1} \\ u_{1,2} \\ u_{2,2} \\ u_{3,2} \\ u_{1,3} \\ u_{2,3} \\ u_{3,3} \end{Bmatrix} = \begin{Bmatrix} 0{,}458 \\ 0{,}458 \\ 0{,}458 \\ 0{,}458 \\ 0{,}458 \\ 0{,}458 \\ 0{,}458 \\ 0{,}458 \\ 0{,}458 \end{Bmatrix}$$

as quais podem ser resolvidas por

$$\begin{aligned} u_{1,1} &= -0{,}315 & u_{1,2} &= -0{,}401 & u_{1,3} &= -0{,}315 \\ u_{2,1} &= -0{,}401 & u_{2,2} &= -0{,}515 & u_{2,3} &= -0{,}401 \\ u_{3,1} &= -0{,}315 & u_{3,2} &= -0{,}401 & u_{3,3} &= -0{,}315 \end{aligned}$$

Esses resultados podem, por sua vez, ser substituídos na Equação (32.14), a qual pode ser escrita na forma de diferenças finitas e resolvida por

$$\begin{aligned} z_{1,1} &= 0{,}063 & z_{1,2} &= 0{,}086 & z_{1,3} &= 0{,}063 \\ z_{2,1} &= 0{,}086 & z_{2,2} &= 0{,}118 & z_{2,3} &= 0{,}086 \\ z_{3,1} &= 0{,}063 & z_{3,2} &= 0{,}086 & z_{3,3} &= 0{,}063 \end{aligned}$$

32.3 PROBLEMAS DE CAMPOS ELETROSTÁTICOS BIDIMENSIONAIS (ENGENHARIA ELÉTRICA)

Fundamentos. Do mesmo modo como a lei de Fourier e o balanço de calor podem ser usados para caracterizar distribuições de temperaturas, relações análogas estão disponíveis para modelar problemas de campo em outras áreas da engenharia. Por exemplo, os engenheiros elétricos usam uma abordagem parecida quando modelam campos eletrostáticos.

Sob diversas hipóteses simplificadoras, um análogo da lei de Fourier pode ser representado na forma unidimensional por

$$D = -\varepsilon \frac{dV}{dx}$$

onde D é chamado vetor densidade de fluxo eletrostático, ε é a permissividade do material e V é o potencial eletrostático.

De maneira semelhante, a equação de Poisson para campos eletrostáticos pode ser representada em duas dimensões por

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = -\frac{\rho_v}{\varepsilon} \tag{32.16}$$

onde ρ_v é a densidade de carga volumétrica.

Finalmente, para regiões que não contenham cargas livres (isto é, $\rho_v = 0$), a Equação (32.16) se reduz à equação de Laplace,

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0 \tag{32.17}$$

Empregue métodos numéricos para resolver a Equação (32.17) para a situação descrita na Figura 32.5. Calcule ambos os valores de V e de D se $\varepsilon = 2$.

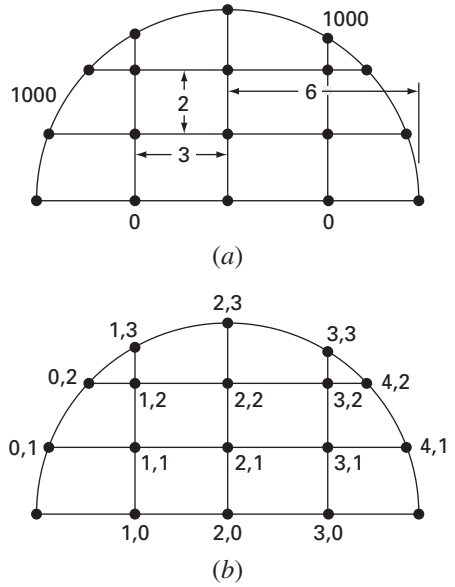


FIGURA 32.5
(*a*) Um sistema bidimensional com uma voltagem de 1000 ao longo da fronteira circular e uma voltagem de 0 ao longo da base. (*b*) O esquema de enumeração dos nós.

Solução. Usando a abordagem delineada na Seção 29.3.2, a Equação (29.24) pode ser escrita para o nó (1, 1) como

$$\frac{2}{\Delta x^2}\left[\frac{V_{1,1}-V_{0,1}}{\alpha_1(\alpha_1+\alpha_2)}+\frac{V_{1,1}-V_{2,1}}{\alpha_2(\alpha_1+\alpha_2)}\right]+\frac{2}{\Delta y^2}\left[\frac{V_{1,1}-V_{0,1}}{\beta_1(\beta_1+\beta_2)}+\frac{V_{1,1}-V_{2,1}}{\beta_2(\beta_1+\beta_2)}\right]=0$$

De acordo com a geometria descrita na Figura 32.5, $\Delta x = 3$, $\Delta y = 2$, $\beta_1 = \beta_2 = \alpha_2 = 1$ e $\alpha_1 = 0{,}94281$. Substituindo esses valores, obtemos

$$0{,}12132V_{1,1} - 121{,}32 + 0{,}11438V_{1,1} - 0{,}11438V_{2,1} + 0{,}25V_{1,1} + 0{,}25V_{1,1} - 0{,}25V_{1,2} = 0$$

Agrupando os termos, resulta

$$0{,}73570V_{1,1} - 0{,}11438V_{2,1} - 0{,}25V_{1,2} = 121{,}32$$

Uma abordagem parecida pode ser aplicada aos nós interiores remanescentes. As equações simultâneas resultantes podem se expressar na forma matricial como

$$\begin{bmatrix} 0{,}73570 & -0{,}11438 & & -0{,}25000 & & \\ -0{,}11111 & 0{,}72222 & -0{,}11111 & & -0{,}25000 & \\ & -0{,}11438 & 0{,}73570 & & & -0{,}25000 \\ -0{,}31288 & & & 1{,}28888 & -0{,}14907 & \\ & -0{,}25000 & & -0{,}11111 & 0{,}72222 & -0{,}11111 \\ & & -0{,}31288 & & -0{,}14907 & 1{,}28888 \end{bmatrix} \times \begin{Bmatrix} V_{1,1} \\ V_{2,1} \\ V_{3,1} \\ V_{1,2} \\ V_{2,2} \\ V_{3,2} \end{Bmatrix} = \begin{Bmatrix} 121{,}32 \\ 0 \\ 121{,}32 \\ 826{,}92 \\ 250 \\ 826{,}92 \end{Bmatrix}$$

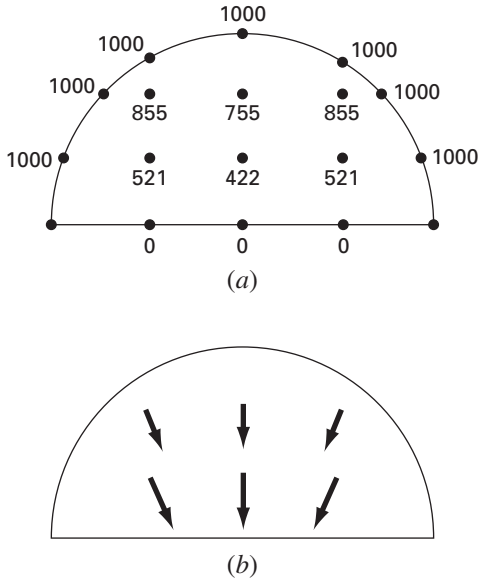


FIGURA 32.6
Os resultados da solução da equação de Laplace com fatores de correção para as fronteiras irregulares. (*a*) Potencial e (*b*) fluxo.

que pode ser resolvida por

$$V_{1,1} = 521{,}19 \qquad V_{2,1} = 421{,}85 \qquad V_{3,1} = 521{,}19$$
$$V_{1,2} = 855{,}47 \qquad V_{2,2} = 755{,}40 \qquad V_{3,2} = 855{,}47$$

Esses resultados estão descritos na Figura 32.6*a*.

Para calcular o fluxo (lembre-se da Seção 29.2.3), as equações (29.14) e (29.15) podem ser modificadas para levar em conta as fronteiras irregulares. No presente exemplo, as modificações resultam em

$$D_x = -\varepsilon \frac{V_{i+1,j} - V_{i-1,j}}{(\alpha_1 + \alpha_2)\,\Delta x}$$

e

$$D_y = -\varepsilon \frac{V_{i,j+1} - V_{i,j-1}}{(\beta_1 + \beta_2)\Delta y}$$

Para o nó (1, 1), essas fórmulas podem ser usadas para calcular as componentes x e y do fluxo

$$D_x = -2\frac{421{,}85 - 1000}{(0{,}94281 + 1)3} = 198{,}4$$

e

$$D_y = -2\frac{855{,}47 - 0}{(1 + 1)2} = -427{,}7$$

as quais, por sua vez, podem ser usadas para calcular o vetor de densidade do fluxo elétrico

$$D = \sqrt{198{,}4^2 + (-427{,}7)^2} = 471{,}5$$

com a direção de

$$\theta = \text{tg}^{-1}\left(\frac{-427{,}7}{198{,}4}\right) = -65{,}1^\circ$$

Os resultados para os outros nós são

Nó	D_x	D_y	D	θ
2, 1	0,0	−377,7	377,7	−90
3, 1	−198,4	−427,7	471,5	245,1
1, 2	109,4	−299,6	281,9	−69,1
2, 2	0,0	−289,1	289,1	−90,1
3, 2	−109,4	−299,6	318,6	249,9

Os fluxos estão mostrados na Figura 32.6*b*.

32.4 SOLUÇÃO POR ELEMENTOS FINITOS DE UMA SÉRIE DE MOLAS (ENGENHARIA MECÂNICA/AEROESPACIAL)

Fundamentos. A Figura 32.7 mostra uma série de molas conectadas. Uma extremidade é fixa a uma parede, enquanto a outra está sujeita a uma força constante F. Usando o procedimento passo a passo delineado no Capítulo 31, pode ser empregada uma abordagem por elementos finitos para determinar os deslocamentos das molas.

Solução.

Discretização. A maneira de decompor esse sistema é, obviamente, tratar cada mola como um elemento. Assim, o sistema consiste em quatro elementos e cinco nós (Figura 32.7*b*).

Equações dos elementos. Como esse sistema é tão simples, suas equações dos elementos podem ser escritas diretamente sem recorrer a aproximações matemáticas. Este é um exemplo da abordagem direta para a dedução dos elementos.

A Figura 32.8 mostra um elemento individual. A relação entre a força F e o deslocamento x pode ser representada matematicamente pela lei de Hooke:

$$F = kx$$

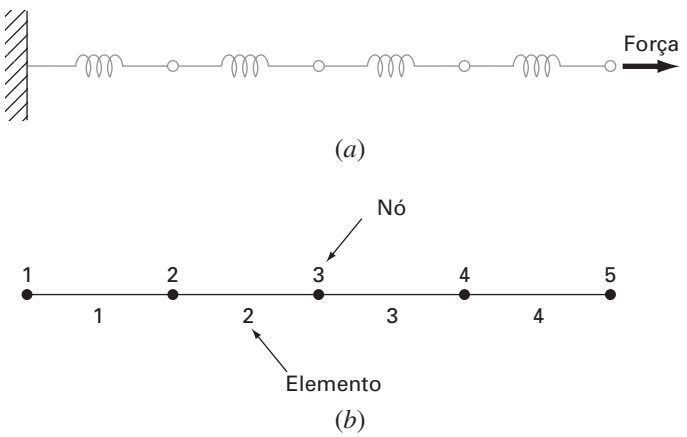


FIGURA 32.7
(*a*) Uma série de molas conectadas. Uma extremidade está fixa em uma parede, ao passo que a outra está sujeita a uma força constante F. (*b*) A representação por elementos finitos. Cada mola representa um elemento. Portanto, o sistema consiste em quatro elementos e cinco nós.

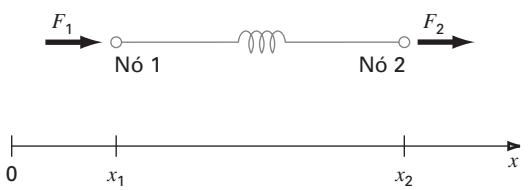


FIGURA 32.8
Um diagrama de corpo livre para uma mola do sistema.

onde k é a constante da mola, que pode ser interpretada como a força necessária para produzir um deslocamento unitário. Se a força F_1 for aplicada no nó 1, o seguinte balanço de forças deve ser válido:

$$F = k(x_1 - x_2)$$

onde x_1 é o deslocamento do nó 1 de sua posição de equilíbrio e x_2 é o deslocamento do nó 2 de sua posição de equilíbrio. Logo, $x_2 - x_1$ representa quanto a mola está esticada ou comprimida em relação ao equilíbrio (Figura 32.8).

Essa equação também pode ser escrita como

$$F_1 = kx_1 - kx_2$$

Para um sistema estacionário, um balanço de forças também exige que $F_1 = -F_2$ e, portanto,

$$F_2 = -kx_1 + kx_2$$

Essas duas equações simultâneas especificam o comportamento do elemento em resposta às forças prescritas. Elas podem ser escritas na forma matricial como

$$\begin{bmatrix} k & -k \\ -k & k \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \end{Bmatrix}$$

ou

$$[k]\{x\} = \{F\} \tag{32.18}$$

onde a matriz $[k]$ é a matriz de propriedades do elemento. Observe que a Equação (32.18) foi colocada na forma da Equação (31.9). Assim, conseguimos gerar uma equação matricial que descreve o comportamento de um elemento típico em nosso sistema.

Antes de prosseguir para o próximo passo — a montagem da solução total — vamos introduzir alguma notação. Os elementos de $[k]$ e $\{F\}$ convencionalmente são escritos com sobrescritos e subscritos, como em

$$\begin{bmatrix} k_{11}^{(e)} & -k_{12}^{(e)} \\ -k_{21}^{(e)} & k_{22}^{(e)} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} F_1^{(e)} \\ F_2^{(e)} \end{Bmatrix}$$

onde o sobrescrito (e) designa que estas são equações dos elementos. Os k's também tem subscritos como em k_{ij} para denotar sua posição na i-ésima linha e j-ésima coluna da matriz. Para o caso presente, eles também podem ser fisicamente interpretados como re-presentando a força necessária no nó i para induzir um deslocamento unitário no nó j.

Montagem. Antes das equações dos elementos serem montadas, todos os elementos e nós devem ser numerados. Esse esquema de enumeração global especifica uma configuração ou topologia do sistema (observe que o caso presente usa um esquema idêntico ao apresentado na Tabela 31.2). Isto é, ela documenta quais nós pertencem a quais elementos. Uma vez que a topologia seja especificada, as equações para cada elemento podem ser escritas em relação a coordenadas globais.

As equações dos elementos podem ser adicionadas uma a uma para montar o sistema total. O resultado final pode ser expresso na forma matricial como [lembre-se da Equação (31.10)]

$$[k]\{x'\} = \{F'\}$$

onde

$$[k] = \begin{bmatrix} k_{11}^{(1)} & -k_{12}^{(1)} & & & \\ -k_{21}^{(1)} & k_{22}^{(1)} + k_{11}^{(2)} & -k_{12}^{(2)} & & \\ & -k_{21}^{(2)} & k_{22}^{(2)} + k_{11}^{(3)} & -k_{12}^{(3)} & \\ & & -k_{21}^{(3)} & k_{22}^{(3)} + k_{11}^{(4)} & -k_{12}^{(4)} \\ & & & -k_{21}^{(4)} & k_{22}^{(4)} \end{bmatrix} \tag{32.19}$$

e

$$\{F'\} = \begin{Bmatrix} F_1^{(1)} \\ 0 \\ 0 \\ 0 \\ F_2^{(4)} \end{Bmatrix}$$

e $\{x'\}$ e $\{F'\}$ são deslocamentos expandidos e vetores de força, respectivamente. Observe que, conforme as equações foram montadas, as forças internas se cancelaram. Assim, o resultado final para $\{F'\}$ tem zeros em todos, exceto o primeiro e o último nó.

Antes de prosseguir para o próximo passo, devemos comentar sobre a estrutura da matriz de propriedades de montagem [Equação (32.19)]. Observe que a matriz é tridiagonal. Este é um resultado direto do esquema de enumeração global particular que foi escolhido (Tabela 31.1) antes da montagem. Embora não seja muito importante no presente contexto, atingir tal sistema esparso, de banda, pode ser uma vantagem decisiva para problemas mais complicados. Isso se deve aos esquemas eficientes disponíveis para resolver tais sistemas.

Condições de Contorno. O presente sistema está sujeito a uma única condição de contorno, $x_1 = 0$. A introdução dessa condição e a aplicação do esquema de enumeração global reduz o sistema para (k's $= 1$)

$$\begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 1 \end{bmatrix} \begin{Bmatrix} x_2 \\ x_3 \\ x_4 \\ x_5 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ F \end{Bmatrix}$$

O sistema está agora na forma da Equação (31.11) e está pronto para ser resolvido.

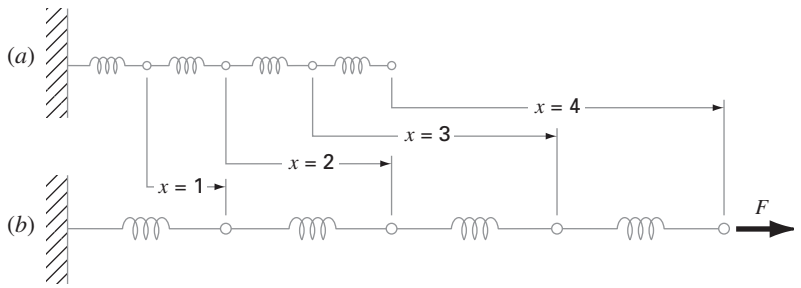
Embora a redução das equações seja certamente uma abordagem válida para incorporar as condições de contorno, em geral é preferível deixar o número de equações intacto quando a solução for obtida pelo computador. Qualquer que seja o método, uma vez que as condições de contorno sejam incorporadas, podemos prosseguir para o próximo passo — a solução.

Solução Geral. Usando uma das abordagens da Parte Três, tal como a eficiente técnica de solução tridiagonal delineada no Capítulo 11, o sistema pode ser resolvido por (com todos os k's $= 1$ e $F = 1$)

$$x_2 = 1 \qquad x_3 = 2 \qquad x_4 = 3 \qquad x_5 = 4$$

Pós-processamento. Os resultados podem agora ser mostrados graficamente. Como na Figura 32.9, os resultados são os esperados. Cada mola está esticada por um deslocamento unitário.

FIGURA 32.9
(*a*) O sistema de molas original. (*b*) O sistema depois da aplicação da força constante. Os deslocamentos estão indicados no espaço entre os dois sistemas.



PROBLEMAS

Engenharia Química/Bioengenharia

32.1 Faça os mesmos cálculos que na Seção 32.1, mas use $\Delta x = 1{,}25$.

32.2 Desenvolva uma solução por elementos finitos para o sistema estacionário da Seção 32.1.

32.3 Calcule os fluxos de massa para a solução estacionária da Seção 32.1 usando a primeira lei de Fick.

32.4 Calcule a distribuição estacionária da concentração para o tanque mostrado na Figura P32.4. A EDP que rege esse sistema é

$$D\left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2}\right) - kc = 0$$

e as condições de contorno são como mostradas. Use um valor de 0,5 para D e 0,1 para k.

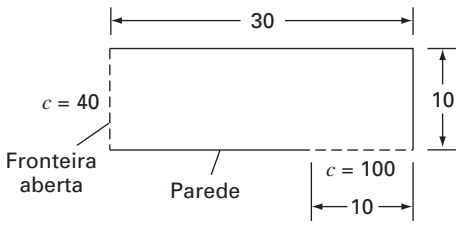


Figura P32.4

32.5 Duas placas estão separadas por 10 cm, como mostrado na Figura P32.5. Inicialmente, ambas as placas e o fluido estão parados. Em $t = 0$, a placa de cima começa a se mover a uma velocidade constante de 8 cm/s. As equações que regem o movimento dos fluidos são

$$\frac{\partial v_{\text{óleo}}}{\partial t} = \mu_{\text{óleo}} \frac{\partial^2 v_{\text{óleo}}}{\partial x^2} \quad \text{e} \quad \frac{\partial v_{\text{água}}}{\partial t} = \mu_{\text{água}} \frac{\partial^2 v_{\text{água}}}{\partial x^2}$$

e as seguintes relações são válidas na interface óleo-água:

$$v_{\text{óleo}} = v_{\text{água}} \quad \text{e} \quad \mu_{\text{óleo}} \frac{\partial v_{\text{óleo}}}{\partial x} = \mu_{\text{água}} \frac{\partial v_{\text{água}}}{\partial x}$$

Qual é a velocidade das duas camadas de fluidos em $t = 0{,}5$, 1, e 1,5 s nas distâncias de $x = 2$, 4, 6 e 8 cm da placa inferior? Observe que $\mu_{\text{água}}$ e $\mu_{\text{óleo}}$ são iguais a 1 e 3, respectivamente.

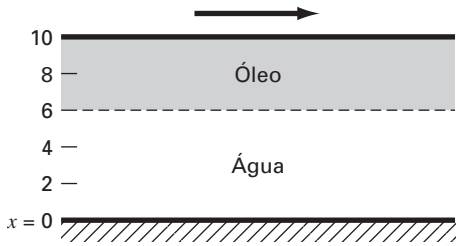


Figura P32.5

Engenharia Civil/Ambiental

32.6 Faça os mesmos cálculos que na Seção 32.2, mas use $\Delta x = \Delta y = 0{,}4$ m.

32.7 O escoamento através de um meio poroso pode ser descrito pela equação de Laplace

$$\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} = 0$$

onde h é a carga. Use métodos numéricos para determinar a distribuição de carga para o sistema mostrado na Figura P32.7.

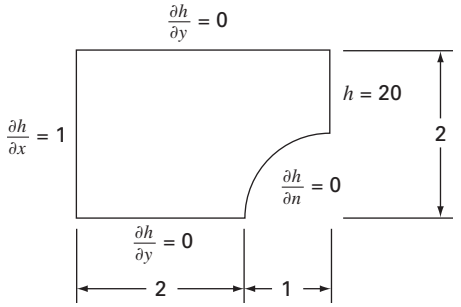


Figura P32.7

32.8 A velocidade do escoamento de água por um meio poroso pode ser relacionada com a carga pela lei de D'Arcy

$$q_n = -K\frac{dh}{dn}$$

onde K é a condutividade hidráulica e q_n é a velocidade de descarga na direção n. Se $K = 5 \times 10^{-4}$ cm/s, calcule a velocidade da água no Problema 32.7.

Engenharia Elétrica

32.9 Faça os mesmos cálculos que na Seção 32.3, mas para o sistema descrito na Figura P32.9.

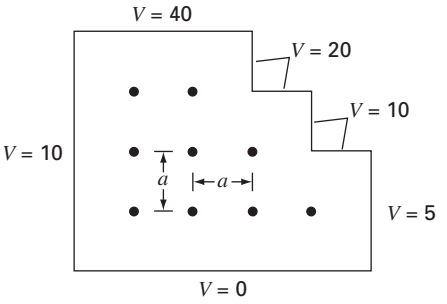


Figura P32.9

32.10 Faça aos mesmos cálculos que na Seção 32.3, mas para o sistema descrito na Figura P32.10.

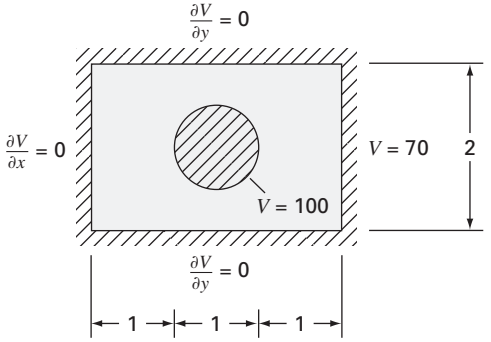


Figura P32.10

Engenharia Mecânica/Aeroespacial

32.11 Faça os mesmos cálculos que na Seção 32.4, mas mude a força para 1,5 e as constantes das molas para

Mola	1	2	3	4
k	0,75	1,5	0,5	2

32.12 Faça os mesmos cálculos que na Seção 32.4, mas mude a força para 2 e use cinco molas com

Mola	1	2	3	4	5
k	0,25	0,5	1,5	0,75	1

32.13 Uma barra composta isolada é formada por duas partes encostadas nas extremidades e ambas as partes têm comprimentos iguais. A parte a possui condutividade térmica k_a, para $0 \le x \le 1/2$, e a parte b tem condutividade térmica k_b, para $1/2 \le x \le 1$. As equações de condução de calor transientes adimensionais que descrevem a temperatura u ao longo do comprimento x da barra composta são

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t} \qquad 0 \le x \le 1/2$$

$$r\frac{\partial^2 u}{\partial x^2} = \frac{\partial u}{\partial t} \qquad 1/2 \le x \le 1$$

onde u é a temperatura, x é a coordenada axial, t é o tempo e $r = k_a/k_b$. As condições de contorno e iniciais são

Condições de contorno	$u(0, t) = 1$	$u(1, t) = 1$
	$\left(\frac{\partial u}{\partial x}\right)_a = \left(\frac{\partial u}{\partial x}\right)_b$	$x = 1/2$
Condições iniciais	$u(x, 0) = 0$	$0 < x < 1$

Resolva esse conjunto de equações para a distribuição de temperaturas como função do tempo. Use diferenças finitas com acurácia de segunda ordem para as derivadas com uma formulação de Crank-Nicolson para integrar no tempo. Escreva um programa computacional para a solução e escolha valores de Δx e Δt para uma boa acurácia. Trace a temperatura u *versus* o comprimento x para diversos valores de tempo t. Gere uma curva separada para cada um dos seguintes valores do parâmetro $r = 1$, 0,1, 0,01, 0,001 e 0.

32.14 Resolva a equação de condução de calor transiente adimensional em duas dimensões que representa a distribuição de temperaturas transiente em uma placa isolada. A equação é

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial t}$$

onde u é a temperatura, x e y são as coordenadas espaciais e t é o tempo. As condições de contorno e iniciais são

Condições de contorno	$u(x, 0, t) = 0$	$u(x, 1, t) = 1$	
	$u(0, y, t) = 0$	$u(1, y, t) = 1$	
Condições iniciais	$u(x, y, 0) = 0$	$0 \le x < 1$	$0 \le y < 1$

Resolva usando a técnica implícita de direções alternadas. Trace os resultados usando uma rotina tridimensional de gráficos em que o plano horizontal contenha os eixos x e y e o eixo z é a variável independente u. Construa diversos gráficos em diversos instantes de tempo, incluindo o seguinte: (**a**) as condições iniciais; (**b**) um instante intermediário, aproximadamente na metade do caminho para o estado estacionário e (**c**) a condição estacionária.



EPÍLOGO: PARTE OITO

PT8.3 PRÓS E CONTRAS

Os prós e contras principais associados aos métodos numéricos para a solução de equações diferenciais parciais envolvem a escolha entre abordagens por *diferenças finitas* e por *elementos finitos*. Os métodos de diferenças finitas são conceitualmente mais fáceis de entender. Além disso, são fáceis de programar para sistemas que possam ser aproximados por grades uniformes. Entretanto, são difíceis de aplicar a sistemas com geometrias complicadas.

As abordagens por diferenças finitas podem ser divididas em categorias, dependendo do tipo de EDP que está sendo resolvida. As *EDPs elípticas* podem ser aproximadas por um conjunto de equações algébricas lineares. Conseqüentemente, o *método de Liebmann* (que, na realidade, é Gauss-Seidel) pode ser empregado para obter a solução iterativamente.

As *EDPs parabólicas unidimensionais* podem ser resolvidas de duas maneira fundamentalmente diferentes: abordagens explícitas ou implícitas. O *método explícito* avança no tempo de uma forma parecida com a técnica de Euler para resolver EDOs. Possui a vantagem de ser simples de programar, mas tem a desvantagem de um critério de estabilidade muito restritivo. Em contraste, estão disponíveis métodos implícitos estáveis. Uma dessas abordagens, o *método de Crank-Nicolson*, é tanto acurada quanto estável e, portanto, amplamente usada para problemas parabólicos lineares unidimensionais.

As *EDPs parabólicas bidimensionais* também podem ser modeladas explicitamente. Contudo, suas restrições de estabilidade são ainda mais severas do que para o caso unidimensional. Abordagens implícitas especiais, que são conhecidas em geral como métodos de decomposição, têm sido desenvolvidas para contornar esse problema. Essas abordagens são tanto eficientes quanto estáveis. Uma das mais comuns é o método *ADI* ou *implícito de direções alternadas*.

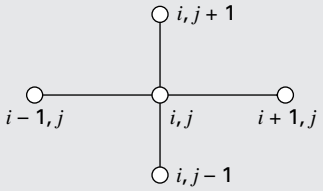
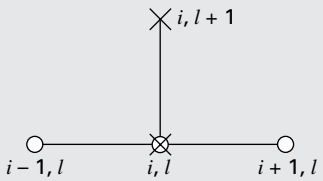
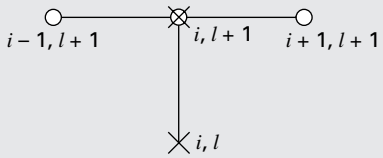
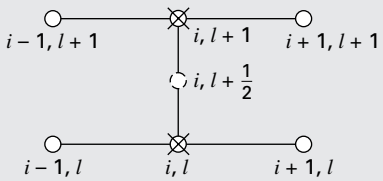
Todas as abordagens por *diferenças finitas* anteriores se tornam incômodas quando aplicadas a sistemas envolvendo formas não uniformes e condições heterogêneas. Os métodos de elementos finitos estão disponíveis para tratar tais sistemas de um modo superior.

Embora o *método dos elementos finitos* seja baseado em algumas idéias bastante simples, a mecânica de gerar um bom código de elementos finitos para problemas bi e tridimensionais não é um exercício trivial. Além disso, pode ser computacionalmente caro para problemas grandes. Entretanto, é amplamente superior às abordagens por diferenças finitas para sistemas envolvendo formas complicadas. Como conseqüência, seu custo e complicação conceitual adicionais são, em geral, justificados pelos detalhes da solução final.

PT8.4 RELAÇÕES E FÓRMULAS IMPORTANTES

A Tabela PT8.3 resume informações importantes que foram apresentadas com relação aos métodos de diferenças finitas na Parte Oito. Essa tabela pode ser consultada para ter acesso rapidamente a relações e fórmulas importantes.

TABELA PT8.3 Resumo dos métodos de diferenças finitas.

	Molécula Computacional	Equação
EDPs elípticas Método de Liebmann		$T_{i,j} = \frac{T_{i+1,j} + T_{i-1,j} + T_{i,j+1} + T_{i,j-1}}{4}$
EDPs parabólicas (unidimensional) Método explícito		$T_i^{l+1} = T_i^l + \lambda(T_{i+1}^l - 2T_i^l + T_{i-1}^l)$
Método implícito		$-\lambda T_{i-1}^{l+1} + (1 + 2\lambda)T_i^{l+1} - \lambda T_{i+1}^{l+1} = T_i^l$
Método de Crank-Nicolson		$-\lambda T_{i-1}^{l+1} + 2(1 + \lambda)T_i^{l+1} - \lambda T_{i+1}^{l+1}$ $= \lambda T_{i-1}^l + 2(1 - \lambda)T_i^l + \lambda T_{i+1}^l$

PT8.5 MÉTODOS AVANÇADOS E REFERÊNCIAS ADICIONAIS

Carnahan, Luther e Wilkes (1969), Rice (1983), Ferziger (1981) e Lapidus e Pinder (1982) fornecem uma análise geral dos métodos e dos softwares para resolver EDPs. Você pode consultar também Ames (1977), Gladwell e Wait (1979), Vichnevetsky (1981, 1982) e Zienkiewicz (1971) para tratamentos com mais profundidade. Informações adicionais sobre os métodos de elementos finitos podem ser encontradas em Allaire (1985), Huebner e Thornton (1982), Stasa (1985) e Baker (1983). Além das EDPs elípticas e parabólicas, também estão disponíveis os métodos numéricos para a resolução de EDPs hiperbólicas. Tais introduções e resumos desses métodos podem ser encontrados em Lapidus e Pinder (1981), Ferziger (1981), Forsythe e Wasow (1960) e Hoffman (1992).

APÊNDICE A

A Série de Fourier

A série de Fourier pode ser expressa em diversas formas diferentes. Duas expressões trigonométricas equivalentes são

$$f(t) = a_0 + \sum_{k=1}^{\infty} [a_k \cos(k\omega_0 t) + b_k \operatorname{sen}(k\omega_0 t)]$$

ou

$$f(t) = a_0 + \sum_{k=1}^{\infty} [c_k \cos(k\omega_0 t + \theta_k)]$$

onde os coeficientes estão relacionados por (ver a Figura A.1)

$$c_k = \sqrt{a_k^2 + b_k^2}$$

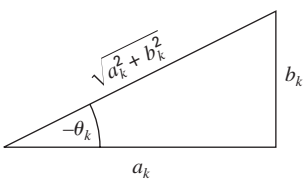
e

$$\theta_k = -\operatorname{tg}^{-1}\left(\frac{b_k}{a_k}\right)$$

Além das formas trigonométricas, a série também pode ser expressa em termos da função exponencial como

$$f(t) = \tilde{c}_0 + \sum_{k=1}^{\infty} [\tilde{c}_k e^{ik\omega_0 t} + \tilde{c}_{-k} e^{-ik\omega_0 t}] \qquad \text{(A.1)}$$

FIGURA A.1
Relações entre as formas cartesiana e polar dos coeficientes da série de Fourier.



onde (ver a Figura A.2)

$$\tilde{c}_0 = a_0$$

$$\tilde{c}_k = \frac{1}{2}(a_k - ib_k) = |\tilde{c}_k| e^{i\phi} k$$

$$\tilde{c}_{-k} = \frac{1}{2}(a_k + ib_k) = |\tilde{c}_k| e^{-i\phi} k$$

onde $|\tilde{c}_0| = a_0$ e

$$|\tilde{c}_k| = \frac{1}{2}\sqrt{a_k^2 + b_k^2} = \frac{c_k}{2}$$

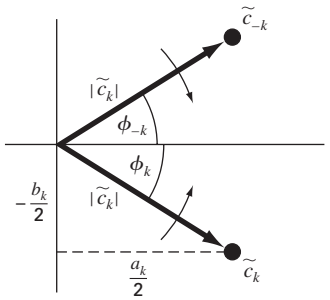
e

$$\phi_k = \operatorname{tg}^{-1}\left(\frac{-b_k}{a_k}\right)$$

Observe que o til significa que o coeficiente é um número complexo.

Cada termo na Equação (A.1) pode ser visualizado como um *phasor* em rotação (as setas na Figura A.2). Termos com um subscrito positivo giram no sentido anti-horário, enquanto aqueles com um subscrito negativo giram no sentido horário. Os coeficientes de $\tilde{c}_k$ e $\tilde{c}_{-k}$ especificam a posição do *phasor* em $t = 0$. A soma infinita dos *phasors* em rotação, os quais podem girar em $t = 0$, é então igual a $f(t)$.

FIGURA A.2
Relações entre a exponencial complexa e os coeficientes reais da série de Fourier.



APÊNDICE B

Iniciação ao MATLAB

O software MATLAB é um programa computacional que fornece ao usuário um ambiente conveniente para muitos tipos de cálculos — em particular, aqueles relacionados com a manipulação de matrizes. O MATLAB opera interativamente, executando os comandos do usuário um a um, conforme eles sejam inseridos. Uma série de comandos pode ser salva como um roteiro e executada como um programa interpretativo. O MATLAB tem um grande número de funções de biblioteca; entretanto, os usuários podem construir suas próprias funções com os comandos e funções do MATLAB. As características principais do MATLAB cálculos de vetores e matrizes, incluindo:

- Aritmética de vetores e matrizes.
- Inversão de matrizes e análise de autovalores/autovetores.
- Operações aritméticas e polinomiais complexas.
- Cálculos estatísticos.
- Exibição gráfica.
- Projeto de sistemas de controle.
- Ajuste de modelos de processos aos dados de teste.

O MATLAB tem diversas ferramentas (*toolboxes*) adicionais que fornecem funções especializadas. Essas incluem: processamento de sinais, sistemas de controle, identificação de sistemas, otimização e estatística.

O MATLAB está disponível em versões executáveis em PCs, Macs e estações de trabalho. A versão moderna executável em PCs o faz no ambiente do Windows. O propósito dos sete exercícios a seguir é lhe dar uma idéia de como calcular com o MATLAB; eles não constituem um tutorial abrangente. Existem materiais tutoriais adicionais nos manuais do MATLAB. Diversos livros texto apresentam exercícios em MATLAB. Além disso, estão disponíveis informações on-line para qualquer comando ou função digitando: `help` *`name`*, em que *`name`* identifica o comando. Não dê apenas uma olhada nestes exercícios; tente-os todos e tente as variações que lhe ocorram. Verifique as respostas que o MATLAB lhe fornece e certifique-se de que você as entendeu e de que elas estejam corretas. Esta é a forma efetiva de aprender MATLAB.

1. Atribuição de Valores a Nomes de Variáveis

A atribuição de valores a variáveis escalares é parecida com outras linguagens computacionais. Tente digitar

```
a = 4
```

e

```
A = 6
```

Observe como a atribuição é respondida para confirmar o que você fez. Isso é característico do MATLAB. A resposta pode ser suprimida, terminando-se a linha de comando com um ponto-e-vírgula (`;`). Tente digitar

```
b = -3;
```

O MATLAB trata os nomes de modo sensível a maiúsculas e minúsculas, isto é, o nome `a` não é o mesmo que o nome `A`. Para ilustrar, insira

```
a
```

e

```
A
```

Veja como seus valores são distintos. Eles são nomes distintos.

Os nomes de variáveis no MATLAB, em geral, representam quantidades matriciais. Um vetor linha pode ser atribuído como segue:

```
a = [ 1 2 3 4 5 ]
```

A resposta confirma a atribuição novamente. Observe como a nova atribuição de `a` domina. Um vetor coluna pode ser inserido de diversas formas. Tente-as.

```
b = [ 1 ; 2 ; 3 ; 4 ; 5 ]
```

ou

```
b = [ 1;
      2;
      3;
      4;
      5 ]
```

ou transpondo um vetor linha com o operador `'`.

```
b = [ 1 2 3 4 5 ]'
```

Uma matriz bidimensional de valores pode ser atribuída como segue:

```
A = [ 1 2 3 ; 4 5 6 ; 7 8 8 ]
```

ou

```
A = [ 1 2 3 ;
      4 5 6 ;
      7 8 8  ]
```

Os valores armazenados por uma variável podem ser examinados a qualquer momento, digitando-se só o nome, por exemplo

```
b
```

ou

```
A
```

Além disso, uma lista de todas as variáveis atuais pode ser obtida, inserindo-se o comando

```
who
```

ou, com mais detalhe, insira

```
whos
```

Existem diversas variáveis predefinidas, como o *pi*.

Também é possível atribuir valores complexos às variáveis, já que o MATLAB trata a aritmética complexa automaticamente. Para fazer isso, é conveniente atribuir a raiz quadrada de -1 ao nome de uma variável, geralmente `i` ou `j`.

```
i = sqrt(-1)
```

Então, um valor complexo pode ser atribuído como

```
x = 2 + i*4
```

2. Operações Matemáticas

As operações com quantidades escalares são manipuladas de forma direta, análoga às linguagens computacionais. Os operadores comuns, em ordem de prioridade são

```
^     Exponenciação
* /   Multiplicação e divisão
\     Divisão à esquerda (se aplica a matrizes)
+ -   Adição e subtração
```

Esses operadores funcionarão como em uma calculadora. Tente

```
2 * pi
```

Além disso, variáveis reais escalares podem ser incluídas:

```
y = pi / 4
y ^ 2.45
```

Resultados de cálculos podem ser atribuídos a uma variável, como no penúltimo exemplo, ou simplesmente exibidos, como no último exemplo.

Os cálculos também podem envolver quantidades complexas. Usando x definido anteriormente, tente

```
3 * x
1 / x
x ^ 2
x + y
```

O poder real do MATLAB é ilustrado por sua habilidade de efetuar cálculos matriciais. O produto interno de dois vetores (produto escalar) pode ser calculado usando-se o operador *,

```
a * b
```

e, do mesmo modo, o produto exterior

```
b * a
```

Para ilustrar a multiplicação de vetores e matrizes, primeiro redefina `a` e `b`,

```
a = [ 1 2 3 ]
```

e

```
b = [ 4 5 6 ]'
```

Agora, tente

```
a * A
```

ou

```
A * b
```

O que acontece quando as dimensões não são as exigidas pelas operações? Tente

```
A * a
```

A multiplicação de matriz por matriz é feita de modo parecido:

```
A * A
```

Operações mistas com escalares também são possíveis:

```
A / pi
```

É importante lembrar sempre que o MATLAB aplicará os operadores aritméticos simples na forma vetor matriz sempre que possível. Às vezes, você pode querer fazer os

cálculos item a item em uma matriz ou vetor. O MATLAB permite isso também. Por exemplo,

```
A ^ 2
```

resulta na multiplicação de A por si mesma. E se você quiser elevar cada elemento de A ao quadrado? Isso pode ser feito com

```
A .^ 2
```

O . precedendo o operador ^ significa que a operação deve ser feita item a item. O manual do MATLAB chama esse tipo de operações *arrays operations*.

Quando o operador divisão (/) é utilizado com matrizes, está implícito o uso de uma matriz inversa. Portanto, se A for uma matriz quadrada não singular, então B/A corresponde à multiplicação à direita de B pela inversa de A. Uma forma mais longa de fazer isso é usar a função *inv*, isto é B*inv(A); entretanto, o uso do operador divisão é mais eficiente já que X = B/A é, na realidade, resolvido como um conjunto de equações X*A=B, usando-se um esquema de decomposição/eliminação.

O operador de "divisão à esquerda" (\ , o caractere contrabarra) também é usado nas operações com matrizes. Como descrito anteriormente, A\B corresponde à multiplicação à esquerda de B pela inversa de A. Isso é, na verdade, resolvido como um conjunto de equações A*X=B, um cálculo comum na engenharia.

Por exemplo, se c for um vetor coluna com valores 0,1,1,0 e 10, a solução de A * x = c , em que A foi definida anteriormente, pode ser obtida digitando-se

```
c = [ 0.1 1.0 10 ]'
x = A \ c
```

Tente fazer isso.

3. O Uso de Funções de Biblioteca

O MATLAB e suas Toolboxes têm uma rica coleção de funções de biblioteca. Podemos usar ajuda on-line para descobrir mais sobre elas. Uma de suas propriedades importantes é que elas operarão diretamente em quantidades vetoriais ou matriciais. Por exemplo, tente

```
log(A)
```

e você verá que a função logaritmo natural é aplicada no estilo array, elemento a elemento, à matriz A. A maioria das funções, como *sqrt, abs, sin, acostanh, exp*, opera na forma de array. Certas funções, como a exponencial e a raiz quadrada, também têm definições matriciais. O MATLAB calculará a versão matricial quando a letra m for acrescentada ao nome da função. Tente

```
sqrtm(A)
```

Um uso comum das funções é calcular uma fórmula para uma série de argumentos. Crie um vetor coluna t que contém valores de 0 a 100, em passos de 5,

```
t = [ 0 : 5 : 100 ]'
```

Verifique o número de itens no vetor t com a função *Lenght*,

```
length(t)
```

Agora, digamos que você queira calcular y = f(t), em que a fórmula é calculada para cada valor de t no vetor, e atribuir o resultado à posição correspondente no vetor y. Por exemplo,

```
y = t .^ 0.34 - log10(t) + 1 ./ t
```

Feito! [Observe o uso dos operadores array adjacentes aos pontos decimais.] Isso é análogo a criar uma coluna de valores de t em uma planilha e copiar uma fórmula em uma coluna adjacente para calcular os valores de y.

4. Gráficos

Os recursos gráficos do MATLAB são parecidos com aqueles de um programa de planilha. Os gráficos podem ser criados rápida e convenientemente; entretanto, não há muita flexibilidade para mudá-los.

Por exemplo, para criar um gráfico dos vetores `t,y` dos dados anteriores, insira

```
plot(t, y)
```

Está pronto! Você pode modificar o gráfico um pouco com comandos como os seguintes:

```
title('Grafico de y versus t')
xlabel('Valores de t')
ylabel('Valores de y')
grid
```

O gráfico aparece em uma janela diferente e pode ser impresso ou transferido pelo teclado (PCs com Windows ou Macs) para outros programas.

Existem outras características dos gráficos que são úteis, por exemplo, traçar objetos em vez de retas, gráficos de famílias de curvas, gráficos no plano complexo, janelas múltiplas de gráficos, gráficos log-log ou semilog, gráficos de grades tridimensionais e gráficos de contorno.

5 Polinômios

No MATLAB existem muitas funções que lhe permitem operar em vetores como se suas entradas fossem coeficientes ou raízes de equações polinomiais. Por exemplo, entre

```
c = [ 1 1 1 1 ]
```

e então

```
r = roots(c)
```

e as raízes do polinômio $x^3 + x^2 + x + 1 = 0$ seriam impressas e também armazenadas no vetor *r*. Os coeficientes dos polinômio podem ser calculados a partir das raízes com a função *poly*,

```
poly(r)
```

e um polinômio pode ser calculado em dado valor de `x`. Por exemplo,

```
polyval(c, 1.32)
```

Se outro polinômio, $2x^2 - 0{,}4x - 1$, for representado pelo vetor `d`,

```
d = [ 2 -0.4 -1 ]
```

os dois polinômios podem ser multiplicados simbolicamente com a função de convolução, *conv*, para produzir os coeficientes do produto dos polinômios,

```
cd = conv(c,d)
```

A função deconvolução, *deconv*, pode ser usada para dividir um polinômio por outro, por exemplo,

```
[ q,r ] = deconv(c,d)
```

O resultado `q` é o quociente e o resultado `r` é o resto.

Existem outras funções polinomiais que podem ser úteis para você, tal como a função resíduo para a expansão em frações parciais.

6. Análise Estatística

A Statistics Toolbox (caixa de ferramentas estatísticas) contém muitos recursos para a análise estatística; entretanto, cálculos estatísticos comuns podem ser feitos com o conjunto das funções básicas do MATLAB. Você pode gerar uma série de números (pseudo)

aleatórios com a função *rand*. Estão disponíveis tanto a distribuição uniforme quanto a normal:

```
rand('normal')
n = 0 : 5 : 1000 ;
```

(Você esqueceu o ; !!!)

```
num = rand(size(n)) ;
```

Você provavelmente entendeu por que o uso do ponto-e-vírgula no final dos comandos dados anteriormente é importante, especialmente se esqueceu de fazê-lo.

Se quiser ver um gráfico do ruído, tente

```
plot(num)
```

Esses números são normalmente distribuídos por hipótese com média zero e variância (e desvio padrão) de 1. Verifique com

```
mean(num)
```

e

```
std(num)
```

Ninguém é perfeito! Você pode encontrar os valores máximo e mínimo,

```
min(num)
max(num)
```

Há uma função conveniente para traçar um histograma dos dados:

```
hist(num,20)
```

onde 20 é o número de degraus.

Se quiser ajustar um polinômio a alguns dados por mínimos quadrados, você pode usar a função *polyfit*. Tente o seguinte exemplo:

```
t = 0 : 5
y = [ -0.45 0.56 2.34 5.6 9.45 24.59 ]
coef = polyfit( t, y, 3 )
```

Os valores em `coef` são os coeficientes do polinômio ajustado. Para gerar o valor calculado de `y`,

```
yc = polyval( coef,t )
```

e para traçar os dados junto com a curva ajustada,

```
plot ( t,yc,t,y,'o' )
```

O gráfico da curva contínua é linear por partes; portanto, ela não parece muito lisa. Melhore isso como segue:

```
t1 = [ 0 : 0.05 : 5 ] ;
yc = polyval(coef, t1)
plot(t1, yc, t, y, 'o')
```

7. Isto e Aquilo

Existem muitos outros recursos do MATLAB. Alguns deles você vai achar úteis; outros, talvez nunca use. Nós o encorajamos a explorar e experimentar.

Para salvar uma cópia de sua seção, o MATLAB tem um recurso útil chamado *diary*. Você digita o comando

```
diary problem1
```

e o MATLAB abre um arquivo no disco que armazena todos os comandos subseqüentes e os resultados (não os gráficos) de sua seção. Você pode desativar esse recurso com:

```
diary off
```

e ativá-lo novamente com o mesmo arquivo:

```
diary on
```

Depois de deixar o MATLAB, o arquivo *diary* estará disponível para você. É comum usar um editor ou processador de texto para limpar esse arquivo (livrando-se de todos os erros que você fez antes que alguém possa vê-los!) e então imprimi-lo para obter uma cópia rígida das partes importantes da sua seção de trabalho, por exemplo, resultados numéricos chave.

Saia do MATLAB com os comandos `quit` ou `exit`. É possível salvar o estado atual de sua seção com o comando `save`. Também é possível recarregar esse estado com o comando `load`.



BIBLIOGRAFIA

AL-KHAFAJI, A. W.; TOOLEY, J. R. *Numerical methods in engineering practice*. Nova York: Holt, Rinehart and Winston, 1986.

ALLAIRE, P. E. *Basics of the finite element method*. Dubuque, IA: William C. Brown, 1985.

AMES, W. F. *Numerical methods for partial differential equations*. Nova York: Academic Press, 1977.

ANG, A. H-S.; TANG, W. H. *Probability concepts in engineering planning and design, Vol. 1*: Basic Principles, Nova York: Wiley, 1975.

APHA (American Public Health Association), 1992. *Standard Methods for the Examination of Water and Wastewater*, 18. ed. Washington, DC.

ATKINSON, K. E. *An introduction to numerical analysis*. Nova York: Wiley, 1978.

ATKINSON, L. V.; HARLEY, P. J. *An Introduction to numerical methods with pascal*. Reading, MA: Addison-Wesley, 1983.

BAKER, A. J. *Finite element computational fluid mechanics*. Nova York: McGraw-Hill, 1983.

BATHE, K. J.; WILSON, E. L. *Numerical methods in finite element analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1976.

BOOTH, G. W.; PETERSON, T. L. Nonlinear Estimation. *I.B.M*. Share *Program Pa*. n. 687 WLNL1, 1958.

BOYCE, W. E.; DIPRIMA, R. C. *Elementary differential equations and boundary value problems*. 5. ed. Nova York: Wiley, 1992.

BRANSCOMB, L. M. Electronics and Computers: An Overview. *Science*. v. 215, p. 755, 1982.

BRIGHAM, E. O. *The fast fourier transform*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

BURDEN, R. L.; FAIRES, J. D. *Numerical analysis*. 5. ed. Boston: PWS Publishing, 1993.

BUTCHER, J. C. On Runge-Kutta Processes of Higher Order. *J. Austral. Math. Soc*., v. 4, p. 179, 1964.

CARNAHAN, B. H. et al. *Applied numerical methods*. Nova York: Wiley, 1969.

CASH, J. R.; KARP, A. H. *ACM Transactions on mathematical software*. v. 16, p. 201–222, 1990.

CHAPRA, S. C. *Surface water-quality modeling*. Nova York: McGraw-Hill, 1997.

CHAPRA, S. C.; CANALE, R. P. *Introduction to computing for engineers*. 2. ed. Nova York: McGraw-Hill, 1994.

_____. *Numerical methods for engineers with personal computers*. Nova York: McGraw-Hill, 1985.

CHENEY, W.; KINCAID, D. *Numerical mathematics and computing*. 2. ed. Monterey, CA: Brooks/Cole, 1994.

CHIRLIAN, P. M. *Basic network theory*. Nova York: McGraw-Hill, 1969.

COOLEY, J. W. et al. Historical notes on the fast fourier transform. *IEEE trans. Audio Electroacoust. AU*, v. 15, n. 2, p. 76–79, 1977.

DANTZIG, G. B. *Linear programming and extensions*. Princeton, NJ: Princeton University Press, 1963.

DAVIS, H. T. *Introduction to nonlinear differential and integral equations*. Nova York: Dover, 1962.

DAVIS, L. *Handbook of genetic algorithms*. Nova York: Van Nostrand Reinhold, 1991.

DAVIS, P. J.; RABINOWITZ, P. *Methods of numerical integration*. Nova York: Academic Press, 1975.

DENNIS, J. E.; SCHNABEL, R. B. *Numerical methods for unconstrained optimization and nonlinear equations*. Filadélfia, PA: Society for Industrial and Applied Mathematics (SIAM), 1996.

DIJKSTRA, E. W. Go to statement considered harmful. *Commun. ACM*. v. 11, n. 3, p. 147–148, 1968.

DRAPER, N. R.; APPLIED, H. Smith. *Regression analysis*. 2. ed. Nova York: Wiley, 1981.

ENRIGHT, W. H. et al. Comparing numerical methods for stiff systems of ODE's. *BIT*, v.15, p. 10, 1975.

FADEEV, D. K.; FADEEVA, V. N. *Computational methods of linear algebra*. São Francisco: Freeman, 1963.

FERZIGER, J. H. *Numerical methods for engineering application*. Nova York: Wiley, 1981.

FLETCHER, R. *Practical methods of optimization: 1*: unconstrained optimization. Chichester, UK: Wiley, 1980.

_____. *Practical methods of optimization: 2*: constrained optimization. Chichester: Wiley, 1981.

FORSYTHE, G. E.; WASOW, W. R. *Finite-difference methods for partial differential equations*. Nova York: Wiley, 1960.

FORSYTHE, G. E. et al. *Computer methods for mathematical computation*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

FYLSTRA, D. et al. Design and use of the Microsoft Excel Solver. *Interfaces*, v. 28, n. 5, p. 29–55, 1998.

GABEL, R. A.; ROBERTS, R. A. *Signals and linear systems*. Nova York Wiley: 1987.

GEAR, C. W. *Numerical initial-value problems in ordinary differential equation*s. Englewood Cliffs, NJ: Prentice-Hall, 1971.

GERALD, C. F.; WHEATLEY, P. O. *Applied numerical analysis*. 3. ed. Reading, MA: Addison-Wesley, 1989.

GILL, P. E. et al, *Practical optimization*, Londres: Academic Press, 1981.

GLADWELL, J.; WAIT, R. *A survey of numerical methods of partial differential equations*. Nova York: Oxford University Press, 1979.

GOLDBERG, D. E. *Genetic algorithms in search*, o*ptimization and machine learning*. Reading, MA: Addison-Wesley, 1989.

GUEST, P. G. *Numerical methods of curve fitting*. Nova York: Cambridge University Press, 1961.

HAMMING, R. W. *Numerical methods for scientists and engineers*. 2. ed. Nova York: McGraw-Hill, 1973.

HARTLEY, H. O. The modified Gauss-Newton Method for fitting non-linear regression functions by least squares. *Technometrics* v. 3, p. 269–280, 1961.

HAYT, W. H.; KEMMERLY, J. E. *Engineering circuit analysis*. Nova York: McGraw-Hill, 1986.

HEIDEMAN, M. T. et al. Gauss and the history of the fast Fourier transform. *IEEE ASSP Mag*, v. 1, n. 4, p. 14–21, 1984.

HENRICI, P. H. *Elements of numerical analysis*. Nova York: Wiley, 1964.

HILDEBRAND, F. B. *Introduction to numerical analysis*. 2. ed. Nova York: McGraw-Hill, 1974.

HOFFMAN, J. *Numerical methods for engineers and scientists*. Nova York: McGraw-Hill, 1992.

HOLLAND, J. H. *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press, 1975.

HORNBECK, R. W. *Numerical methods,* Nova York: *Quantum*, 1975.

HOUSEHOLDER, A. S. *Principles of numerical analysis*. Nova York: McGraw-Hill, 1953.

_____. *The theory of matrices in numerical analysis,* Nova York: Blaisdell, 1964.

HUEBNER, K. H.; THORNTON, E. A. *The finite element method for engineers*. Nova York: Wiley, 1982.

HULL, T. E.; CREEMER, A. L. The Efficiency of predictor-corrector procedures. *J. Assoc. Comput. Mach*. v. 10, p. 291, 1963.

ISAACSON, E.; KELLER, H. B. *Analysis of numerical methods*. Nova York: Wiley, 1966.

JACOBS, D. (Ed.). *The state of the art in numerical analysis*. Londres: Academic Press, 1977.

JAMES, M. L. et al. *Applied numerical methods for digital computations with FORTRAN and CSMP*. 3. ed. Nova York: Harper & Row, 1985.

KELLER, H. B. *Numerical methods for two-point boundary-value problems*. Nova York: Wiley, 1968.

LAPIDUS, L.; PINDER, G. F. *Numerical solution of partial differential equations in science and engineering*. Nova York: Wiley, 1981.

LAPIDUS, L.; SEINFIELD, J. H. *Numerical solution of ordinary differential equations*. Nova York: Academic Press, 1971.

LAPIN, L. L. *Probability and statistics for modern engineering*. Monterey, CA: Brooks/Cole, 1983.

LASDON, L. S.; SMITH, S. Solving large nonlinear programs using GRG. *ORSA Journal on computing*, v. 4, n. 1, p. 2–15, 1992.

LASDON, L. S. et al. Design and testing of a generalized reduced gradient code for nNonlinear programming. *ACM transactions on mathematical software*. v. 4, n. 1, p. 34–50, 1978.

LAWSON, C. L.; HANSON, R. J. *Solving least squares Problems*. Englewood Cliffs, NJ: Prentice-Hall. 1974.

LUENBERGER, D. G. *Introduction to linear and nonlinear programming*. Reading, MA: Addison-Wesley, 1984.

LYNESS, J. M. Notes on the adaptive Simpson quadrature routine. *J. Assoc. Comput. Mach*. v.16, p. 483, 1969.

MALCOLM, M. A.; SIMPSON, R. B. Local versus global strategies for adaptive quadrature. *ACM trans. math. software*, v.1, p.129, 1975.

MARON, M. J. *Numerical analysis, a practical approach*. Nova York: Macmillan, 1982.

MILTON, J. S.; ARNOLD, J. C. *Introduction to probability and statistics*: principles and applications for engineering and the computing sciences. 3. ed. Nova York: McGraw-Hill, 1995.

MULLER, D. E. A Method for solving algebraic equations using a digital computer. *Math. tables aids comput*. v. 10, p. 205, 1956.

NA, T. Y. *Computational methods in engineering boundary value problems*. Nova York: Academic Press, 1979.

NOYCE, R. N. Microelectronics. *Sci. Am*. v. 237, p. 62, 1977.

OPPENHEIM, A. V.; SCHAFER, R. *Digital signal processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

ORTEGA, J. RHEINBOLDT, W. *Iterative solution of nonlinear equations in several variables*. Nova York: Academic Press, 1970.

ORTEGA, J. M. *Numerical analysis* — A second course. Nova York: Academic Press, 1972.

PRENTER, P. M. *Splices and variational methods*. Nova York: Wiley, 1975.

PRESS, W. H. et al. *Numerical recipes*: the art of scientific computing. Cambridge: Cambridge University Press, 1992.

RABINOWITZ, P. Applications of linear programming to numerical analysis. *SIAM Rev*, v. 10, p. 121–159, 1968.

RALSTON, A. Runge-Kutta methods with minimum error bounds. *Match. Comp*., v. 16, p. 431, 1962.

RALSTON, A.; RABINOWITZ, P. *A first course in numerical analysis*. 2. ed. Nova York: McGraw-Hill, 1978.

RAMIREZ, R. W. *The FFT, fundamentals and concepts*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

RAO, S. S. *Engineering optimization*: theory and practice. 3. ed. Nova York: Wiley-Interscience, 1996.

REVELLE, C. S. et al. *Civil and environmental systems engineering*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

RICE, J. R. *Numerical methods, software and analysis*. Nova York: McGraw-Hill, 1983.

RUCKDESCHEL, F. R. *BASIC scientific subroutine*, v. 2, Peterborough, NH: Byte/McGraw-Hill, 1981.

SCARBOROUGH, J. B. *Numerical mathematical analysis*. 6. ed. Baltimore, MD: Johns Hopkins Press, 1966.

SCOTT, M. R.; WATTS, H. A. A Systematized collection of codes for solving two-point boundary-value problems. In: LAPIDUS, L.; SCHIESSER, W. E. (Eds.). *Numerical methods for differential equations*. Nova York: Academic Press, 1976.

SHAMPINE, L. F.; ALLEN, JR., R. C. *Numerical computing*: An introduction. Filadélfia: Saunders, 1973.

SHAMPINE, L. F.; GEAR, C. W. A User's view of solving stiff ordinary differential equations. *SIAM Review*, v. 21, p. 1, 1979.

SIMMONS, E. F. *Calculus with analytical geometry*. Nova York: McGraw-Hill, 1985.

STARK, P. A. *Introduction to numerical methods,* Nova York: Macmillan. 1970.

STASA, F. L. *Applied finite element analysis for engineers.* Nova York: Holt, Rinehart and Winston, 1985.

STEWART, G. W. *Introduction to matrix computations*. Nova York: Academic Press, 1973.

SWOKOWSKI, E. W. *Calculus with analytical geometry*. 2. ed. Boston: Prindle, Weber and Schmidt, 1979.

TAYLOR, J. R. *An introduction to error analysis*. Mill Valley, CA: University Science Books, 1982.

TEWARSON, R. P. *Sparse matrices.* Nova York: Academic Press. 1973.

THOMAS, JR., G. B.; FINNEY, R. L. *Calculus and analytical geometry*. 5. ed. Reading, MA: Addison-Wesley, 1979.

VAN VALKENBURG, M. E. *Network analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

VARGA, R. *Matrix iterative analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1962.

VICHNEVETSKY, R. *Computer Methods for partial differential equations. V. 1*: elliptical equations and the finite element method. Englewood Cliffs, NJ: Prentice-Hall, 1981.

_____. *Computer methods for partial differential eauations. Vol. 2*: initial value problems. Englewood Cliffs, NJ: Prentice-Hall, 1982.

WILKINSON, J. H. *The algebraic eigenvalue problem*. Oxford Fair Lawn, NJ: University Press, 1965.

WILKINSON, J. H.. REINSCH, C. *Linear algebra*: handbook for automatic computation. Berlim: Springer-Verlag, v. 11, 1971.

WOLD, S. Spline functions in data analysis. *Technometrics*, v. 16, n. 1, p. 1–11, 1974.

YAKOWITZ, S.; SZIDAROVSKY, F. *An introduction to numerical computation*. Nova York: Macmillan, 1986.

YOUNG, D. M. *Iterative solution of large linear systems*. Nova York: Academic Press, 1971.

ZIENKIEWICZ, O. C. *The finite element method in engineering science*. Londres: McGraw-Hill, 1971.



ÍNDICE REMISSIVO

A

Abordagem por volume de controle, equações elípticas, 720-721
Abordagem preditor-corretor, 598
Abordagens/métodos diretos
  equações de elementos, método dos elementos finitos em uma dimensão, 741-742
  otimização multidimensional, 306-310
Agoritmo FFT de Sande-Dukey, 454-458
Água de dejeto, otimização do custo mínimo no tratamento de, 349-353
Ajuste
  de uma função sinosoidal, aproximação de Fourier, 442-444
  de uma linha reta, 382-383
Ajuste de curvas, 4, 366. *Ver também* Interpolação
  aproximação de Fourier. *Ver* Aproximação de Fourier
  como bibliotecas e pacotes, 460-467
  dados experimentais, análise de, 476-477
  estudo de casos em, 470-477
  Exel para, 460-463
  IMSL, para, 465-467
  interpolação por splines, 426-436
  MATLAB para, 463-465
  métodos avançados, 489
  polinômio interpolador de Newton. *Ver* Polinômios interpoladores por diferenças divididas de Newton
  prós e contras, 486-487
  regressão dos mínimos quadrados. *Ver* Mínimos quadrados
  regressão linear. *Ver* Regressão linear
  regressão não linear, 403-406
  regressão polinomial. *Ver* Regressão Polinomial
  regressão. *Ver* Regressão
  resumo de fórmulas, 487-488
  splines, 465
Ajuste ótimo de funções a solução, método dos elementos finitos, equações dos elementos, 739
Algarismos significativos, 44-45, 372
  critério de parada, 48-49
  eliminação de Gauss e, 215
  no computador, 56
Algoritmo da diferença do quociente (QD), 185
Algoritmo de Davidon-Fletcher-Powell (DFP), otimização, 321
Algoritmo de Thomas, sistema tri-diagonal com, 247-248
Algoritmo FFT de Cooley-Tukey, 454, 458-459
Algoritmo genético, 308
Algoritmo(s). *Ver também* Fluxogramas; Pseudocódigo
  Bairstow, 151
  controle do tamanho do passo e métodos de passos múltiplos, EDOs, 637
  FFT Cooley-Tukey, 454, 458-459
  FFT Sande-Tukkey, 454, 458
  otimização de busca aleatória, 307
  para a aplicação múltipla da regra dos trapézios, 511-512, 529
  para a busca da razão áurea, 296-297
  para a decomposição LU, 234
  para a iteração de um ponto (ponto fixo), 120
  para a regressão linear, 386-389
  para bissecção, 106-107

para eliminação de Gauss, 220
para o método da secante, 127
para o método de Euler, 594-595
para o método de Gauss-Newton, 498
para o método de Gauss-Seidel, 254-255
para o método de Heun, 603
para o método do ponto médio, 603
para o polinômio interpolador de Newton, 417-418
para os métodos de Newton-Raphson, 124-125
para os métodos de Runge-Kuta, 617-620
para raízes de quadráticas, 27-30, 151
para regressão polinomial, 395-396
para splines cúbicos, 436
para transformação de Fourier discreta, 452
pra o polinômio de Lagrange, 421
regras de Simpson, 520
sistemas de EDOs, 615-616
Allaire, P.E., 772
Allen, R.C., Jr., 489
Ames, W.F., 772
Amplitude, 440-441
fator de aumento para, 173
Análise de erro e sistema condicionado, 239-247
Análise de Fourier, estudo de casos em, 475-476
Análise de sensibilidade, 17
Análise de vibração,
harmônica, 444
raízes de equações, 170-175
Análise estatística, MATLAB, 778-779
Analysis Tollpack, Exel, 461
Ang, A.H-S., 911
Ângulo de fase de retardamento, 440
Ângulo de fase, 440-441
Aplicação única da regra do trapézio, 507
comparação com a quadratura de Gauss, 535
relação com as EDOs, 696
Aplicações múltiplas da regra do trapézio, 509-511, 529
algoritmos para, 512
extrapolação de Richardson, 530-531
Aproximação de Fourier, 438-467
ajuste de curvas, 439-444
ajuste por mínimos quadrados de uma função sinosoidal, 442-444
domínio de freqüência versus domínio de tempo, 447-451
domínio de tempo versus domínio de freqüência, 447-451
e regressão, 442-444
espectro de potência, 458-459
FFT. *Ver* Transformada rápida de Fourier
forma complexa, 447
funções sinoidais, 439-444
série de Fourier contínua, 444-450
série de Fourier, 444-450, 773-774
TFD. *Ver* Transformada de Fourier discreta
transformada e integral de Fourier, 450-452
Aproximação de ordem zero, 63
Aproximação de primeira ordem, 63
Aproximação funcional, 489
Aproximação por diferença centrada, da primeira derivada, 75
Aproximação regressiva por diferenças, 73-74
Aproximações temporais de ordem superior, equações parabólicas, 725-726
Aproximações\pontos iniciais, 87
Arnold, J.C., 373, 401
Arredondamento, 54
Atkinson, K.E., 234
Atratores estranhos, 681
Autovalores, 139, 585, 654-670
bibliotecas e pacotes para, 662-668
definição de, 654
estudo de casos em, 676-678, 683-686
Excel para, 662
IMSL para, 665-667, 685
MATLAB para, 663-667
método da potência, 661-663
método polinomial, 660-661
Auto-vetores, 657

B

Baker, A.J., 772
Balanço de forças, 17, 85
  pára-quedista caindo, 11
Balanços. *Ver* Leis de conservação
Bashforth. *Ver* Fórmulas de integração aberta de Adams-Bashforth
BASIC, 89. *Ver também* Algoritmo(s); VBA
Bernoulli, J., 286
BFGS. *Ver* Algoritmo de Bryoden-Fletcher-Goldfarb-Shanno
Bibliotecas e pacotes, 21-22, 152-154, 256-261
  Excel, 152-155
  IMSL, 158-160
  MATLAB, 155-158
  para ajuste de curvas, 460-467
  para análise de Fourier, 475-476
  para autovalores, 684-685
  para derivação e integração numérica, 551-553
  para equações diferenciais ordinárias, 677-678
  para equações diferenciais parciais, 751-756
  para localização de raízes, 152-160
  para matrizes especiais, 256-261
  para o método de Gauss-Seidel, 256-261
  para otimização, 333-343
  para sistemas de equações algébricas lineares, 256-261
  programação, 21-22
Bits (dígitos binários), 49, 457
Booth, G.W., 406
Boyce, W.E., 140
Branscomb, L.M., 911
Bringham, E.O., 460
Bryoden-Fletcher-Goldfarb-Shanno (BFGS)
  algoritmo de, 321, 333
B-splines, 489
"Buraco" no zero, 54
Burden, R.L., 78, 81
Busca da razão áurea, 296-300
Buscas de padrão e de uma única variável, otimização, 306-310
Buscas incrementais, 113

C

C++, 38
Cálculos de resposta a estímulos, 238-239
Cálculos dependentes do tempo, 15
Cálculos estacionários, 15-16, 264-266
Campo eletrostático bidimensional, 763-764
Canale, R.B., 89
Cancelamento na subtração, 59-60
Caos, estudo de casos em, 679-680
Carga, conservação da, 17
Carnahan, B., 89, 381, 489, 592, 696, 900
CASE, 25
  Exel, 33
  MATLAB, 36
Cash, J.R., 619
Caso criticamente anortecido, 40, 172
Caso sub-amortecido, 140
Casos superamortecido, 140, 171
Cauchy. *Ver* Método de Euler-Cauchy
Chapra, S.C., 89, 677, 761
Charnes , 287
Chart Wizard, Excel, 34
Cheney, W., 89, 433, 489
Chirlian, P.M., 460
Cinética da reação, 305-308, 544, 786, 884-888
Circuito de resistores, 269-271
Circuitos
  corrente elétrica em, 269-270, 682-686
  projeto de circuito elétrico, 168-170
  transferência de potência máxima para, 353-355
Coeficiente(s)
  de amortecimento crítico, 172
  de arrasto, 11
  de condutividade térmica, 706
  de correlação *(r),* 385, 388, 393

de determinação *(r")*, 385, 393, 395
de difusividade térmica, 706
de um polinômio interpolador, 423
de variação, 369
método dos indeterminados, 534-535
Coeficientes indeterminados, método dos, 534-535
Colocação, 743. *Ver também* Interpolação
Colunas carregadas axialmente
autovalores, 660-661
Computador(es)
aritmética em, 56-62
representação de números, 49-56
software. *Ver* Software
Condição de contorno de Neumann, 712
Condição de contorno natural, 712
Condições auxiliares, 583
Condições de contorno fixas, placa aquecida, 709-710
Condições de contorno nas derivadas
equações elípticas, 712-714
métodos de diferença finita, equações parabólicas, 725
Condições de contorno
derivada, diferença finita, parabólica, 725
equações elípticas, 712-716
fixada, temperatura de uma placa aquecida com, 709-710
Condições de Dirichlet, 444
contorno, 780, 712
Condutividade térmica, coeficiente de, 706
Constante de meia saturação, 470
Constantes, de integração, 582
Contando operações. *Ver* Operações em ponto flutuante
Controle do tamanho do passo, 620
adaptativo, 620
método de passo múltiplo, EDOs, 636
Convergência
de interação de um ponto (ponto fixo), 118-120
de sistemas não lineares, 135
do método da falsa posição, 110
do método de Gauss-Seidel, critério para, 252-254
dos métodos de Newton-Raphson, 122-124
e estabilidade, equações parabólicas, 725
taxa de, 79
Convergência quadrática, 122
Cooley, J.W., 454, 458
Corrente eficaz, 561-563
Corrente elétrica em circuitos
resistores e, 269-271
simulação de, EDOs e, 679-684
Corrente, simulação em um circuito elétrico, estudo de casos em, 682-686
Cotes. *Ver* Fórmulas de integração de Newton-Cotes
Covariância, 401
Creemer, A.L., 912
Critério de parada. *Ver* Critério de parada
Critério do melhor ajuste, regressão linear, 380-381
Custo do pára-quedas, otimização do, 287-290

D

Dados
distribuição de, 370
incerteza nos, 82
Dados desigualmente espaçados, 549-550
com bibliotecas e pacotes, 551-553
derivação, 549-550
estudo de casos em, 557-574
integração, 521-523
Dados experimentais, análise de, 476-417
Danielson, 454
Dantzig, G.B., 287
Davis, L., 308
Davis, P.J., 460
Decomposição. *Ver* Decomposição *LU*
Decomposição de Cholesky, 248-249
e regressão, 400
Decomposição de Crout, 234-236
Decomposição de Doolittle, 234

Decomposição em uma única variável, 489
Decomposição *LU,* 228-248, 400
algoritmo, 234
análise de erro, 239-244
decomposição de Cholesky, 248-250
decomposição de Crout, 234-236
decomposição de Doolittle, 234
eliminação de Gauss como, 230-234
inversão de matrizes, 236-239
pseudocódigo para, 231, 233, 235
Deflação polinomial, 141-143
Deflação progressiva, 143-144
Deflação regressiva, 143
Deflexão de uma placa, 762-763
Dennis, J.E., 364
Derivação e integração numérica, 76-77, 491-574
alta acurácia, 545-548
dados com erros, 550-551
derivação, 553-554. *Ver também* Derivação
erro, sensibilidade ao, 550-551
extrapolação de Richardson, 548-549
fórmula de integração de Newton-Cotes. *Ver* Fórmula de integração de Newton-Cotes
incerteza nos dados, 550
integração, comparação com, 551
prós e contras, 667-668
Derivação gráfica por área igual, 493-495
Derivada, 73-77. *Ver também* Derivação e integração numérica
analítica, 499
definição de, 491
erros de arredondamento, 81-82
integração, comparação com, 492-493
polinomial, 140-141
Derivadas direcionais, 311
Derivadas, 491. *Ver também* Derivação e integração numérica
direcional, 311
tabela de, 500
Descriminante, 140
Deslocamento na fase, 440-441
Desvio padrão normalizado, 542
Desvio padrão, 369
Determinantes
cálculo de, eliminação de Gauss, 201-202
escala/por em escala e, 213-214
DFP. Algoritmo de Davidos-Flecher-Powell, otimização
Diagonal
dominante, 254
matrizes, 191
Diferença dividida finita, 13, 73-77
derivadas, aproximação de, 75-76
método da secante, 127
polinômio interpolador por diferenças divididas de Newton, 412-413
Diferença dividida finita segunda, 76-77
Diff. MATLAB, 552
Difusividade térmica, coeficiente de, 706
Dígitos binários (bits), 49, 457
Dinâmica, 761
estática, 761
Direções conjugadas, 309
Direções padrão, 309
Discretização, método dos elementos finitos, 736, 766
em uma dimensão, 741-742
em duas dimensões, 748
Distribuição de fluxo para uma placa aquecida, 711-712
Distribuições normais cumulativas, 542
Divergência. *Ver* Convergência
Divisão por zero, eliminação de Gauss e, 211
Divisão sintética, 142
DOEXIST, 26
Excel, 33
MATLAB, 37
Domínio de freqüência, 447-451
Domínio do tempo, 447-450
Draper, N.R., 383, 386, 393, 403, 489
Duas dimensões espaciais, equações parabólicas em, 732-735

E

Economização de Chebyshev, 489
Editor de Visual Basic (VBE), 32
EDO de segunda ordem, 578
EDOs. *Ver* Equações diferenciais ordinárias
  estudo de casos em, 673-679
EDPs. *Ver* Equações diferencias parciais
Efeitos da escala/mudança de escala, na eliminação de Gauss, 214-215, 217-220
Eliminição de Gauss, 197, 199-225
  algarismos significativos e, 215
  armadilhas da, 211-215
  calculo do determinante, 201-202
  contando as operações, 209-210
  efeito da escala\mudança de escala, 213-214, 218-220
  erros de arredondamento, efeito dos, 211
  ingênua, 205-210
  melhorando as soluções, 215-221
  pivotamento, 216-217
  sistemas complexos, 221-222
  sistemas de equações lineares algébricas, 220-221
  sistemas de equações não lineares e, 222-224
  sistemas mal condicionados, 211-215
  sistemas singulares e, 215
Enganos, 82-83
Engenharia
  e leis de conservação, 15-17
  três fases da solução de problemas, 3
Enright, W.H., 696
Epilímnio, 473
Epsilon da máquina, 55
Equação de condução do calor, 723-724
  método de Crank-Nicolson, 730
  método implícito simples, 726-728
  unidimensional, 724
Equação de Helmholdz, 754
Equação de Laplace elíptica, 705-707. *Ver também* Métodos de diferenças finitas
  equações de diferença, 707-708
Equação de Manning, 165
Equação de Poisson, 754, 890
Equação de potencia
  linearização da, 390-391
  regressão linear múltipla, 398
Equação de taxa de crescimento de saturação, 390
Equação de Van der Pool, 658
Equação de Van der Waals, 162-164
Equações algébricas. *Ver* Sistemas de equações algébricas lineares
Equações características, 139
Equações corretoras, 598
Equações de advecção-dispersão, 759
Equações de elementos, método sos elementos finitos, 736-740
  escolha da função de aproximação, 737-739
Equações de Laplace, 698, 763
  elíptica, 705-707
  equações de diferença, 707-709
Equações de Lorenz, 679
Equações de Lotka-Volterra, 679
Equações de taxas, 578
Equações diferenciais ordinárias (EDOs), 138-140, 673-696
  Adams de quarta ordem, 643-646
  estudos de casos em, 673-689
  método de euler, 587-597
  método de Heun, 598-601
  método de Milnie, 679
  métodos de passo múltiplo. *Ver* Métodos de passo múltiplos, EDOs
  métodos de Runge-Kutta. *Ver* Métodos de Runge-Kutta
  otimização e, 679
  para sistemas rígidos, 665-667
  problemas de contorno, 648-653
  prós e contras, 695-696
  sistemas de, 612-617
  solução analítica, 12-14

Equações diferenciais parciais (EDPs), 6, 578
aplicações em engenharia, 758-766
bibliotecas e pacotes para, 751-756
diferenças finitas. *Ver* Métodos de diferenças finitas
elementos finitos. *Ver* Métodos de elementos finitos
equações elípticas. *Ver* Métodos de diferenças finitas
equações parabólicas. *Ver* Métodos de diferenças finitas
hiperbólica, 701
prós e contras, 771
Equações diferenciais parciais hiperbólicas, 702
Equações lineares algébricas. *Ver também* Sistemas de equações algébricas lineares
estudo de caso em, 264-293
Equações não lineares, 132-136, 222-223
Equações normais
mínimos quadrados linear geral, 399-400
regressão linear múltipla, 396-399
regressão linear, 383-384, 396-399
regressão polinomial, 393-395
Equações parabólicas, 701, 723-735. *Ver também* Métodos de diferenças finitas
aproximações temporais de ordem superior, 726
condições de contorno na derivada, 725-726
convergência e estabilidade, 725
em duas dimensões espaciais, 732-735
equação de condução de calor, 723-724
esquema ADI, 732-735
esquemas explícito e implícito, 732
método de Crank-Nicolson, 730-732
método implícito simples, 726-729
métodos explícitos, 723-724
métodos unidimensionais, comparação de, 731-732
Equações simultâneas. *Ver* Sistemas de equações algébricas lineares
Equações, 699, 705-720. *Ver também* Equações dos elementos; Equações parabólicas
abordagem por volume de controle, 717-720
condições de contorno, 712-717
distribuição de fluxo para uma placa aquecida, 711-712
equação de diferença de Laplace, 707-708
equação de Laplace, 705-707
fórmula de Gauss-Legendre de dois pontos, 536-540
fórmulas de Newton-Cotes para, 528-529
fronteiras irregulares, 714-717
integrais impróprias, 540-542
método de Liebmann, 709-710
placa aquecida com aresta isolada, 713-714
quadratura de Gauss, 534-540
Romberg, 530-534
software para, 720
técnicas de solução, 707-712
temperatura de uma placa aquecida com condições de contorno fixas, 709-710
variáveis secundárias, 711
Erro de formulação, 82-83
Erro de truncamento global, EDOs, 685
Erro de truncamento local, EDOs, 590-594
método de Euler, 590-594
métodos de passo múltiplo, 630-633
Erro de truncamento, 43, 63-82
global, 590
local, 590, 593
método de Euler, 590-594
métodos de passo múltiplo, 631
polinômio interpolador de Newton, 414-419
propagado, 590
Erro do modelo, 83
Erro numérico total, 81-82
Erro padrão da estimativa por regressão linear, 384
múltipla, 398
Erro padrão da estimativa
para a regressão linear, 384
para a regressão polinomial, 394
Erro relativo, 46-49, 79-80
Erro relativo aproximado (*), 43, 46-49
integração de Romberg, 533
iteração de um único ponto (ponto fixo), 120

método da falsa posição, 108
método de Gauss-Seidel, 250-251
método de Heun de início não automático , 683
método de Newton-Raphson, 122
para a bisecção, 104-106
regressão não linear, 403
Erro relativo verdadeiro (*), 48
Erro residual, 379
Erro. *Ver também* Critério de parada
correções, 250, 529-530. *Ver também* Modificadores de arredondamento. *Ver* Erro de arredondamento de truncamento. *Ver* Erro de truncamento
equações de. *Ver* Métodos iterativos, refinamento
formulação de, 82-83
numérico total, 81-83
padrão. *Ver* Erro padrão da estimativa
propagação de, 77-80
relativo aproximado. *Ver* Erro relativo aproximado (*)
relativo, 47-48
resíduo, 379
sensibilidade ao, derivação numérica, 550-551
superposição *LU,* 239-244
Erros de aproximações e de arredondamento, 43-62, 736, 739
Erros de arredondamento, 46
definido, 43
derivação numérica, 80-82
do computador, 49-62
efeito na eliminação de Gauss, 211-212
método de Euler, 590
polinômio interpolador de Newton, 415-419
regressão polinomial, 395
Escoamento de canal aberto, raízes de equações, estudo de casos em, 164-168
Espaço de solução viável, 325
Espectro de Potencia, aproximação de Fourier, 460, 551
Espectro discreto, 449-450
Esquema ADI, equações parabólicas, 732-734
Esquemas implícitos, equações parabólicas, 732
Estabilidade, 79-80, 87, 696
convergência e, equações parabólicas, 725
incondicional, 626
Estatísticas, simples, 368-371
Estimação, 371
Estudo de casos
em ajuste de curvas, 470-477
em análise de Fourier, 475
em dados espaçados desigualmente, 557-574
em equações algébricas lineares, 264-293
em equações de elementos, sistema de elementos finitos, 766-767
em montagem, método dos elementos finitos, 766-767
em otimização, 345-357
em regressão linear, 470-474
em regressão polinomial, 461-463
em sistemas massa-mola, 271-272, 766-768
em splines, 473-474
em um pêndulo oscilante, 686-689
no método de Runge-Kuta de quarta ordem, 609, 695
Estudo de casos em EDOs, 673-689
correntes, circuitos elétricos, simulação de, 682-686
modelos predador-presa e caos, 679-681
pêndulo oscilante, 686-689
reatores, resposta transiente de, 673-678
Estudo de casos em raízes de equações, 162-175
análise de vibração, 170-175
escoamento de canal aberto, 164-168
lei dos gases ideais e não ideais, 162-164
projeto de circuitos elétricos, 168-170
Euler, L., 286
Exatidão, 45-46, 88
Excel, 10, 21, 31-35, 55, 96
análises Toolpack, 461
chart wizard, 34
Editor de Visual Basic, 32. *Ver também* EVB
equações algébricas lineares e, 256-257
para ajuste de curvas, 460-463
para autovalores, 573

para equações diferenciais ordinárias, 612, 677-678
para equações diferenciais parciais, 875-877
para localização de raízes, 152-154
para otimização, 333-338
raízes de equações com, 152-154
Exemplo da barra aquecida, método dos elementos finitos, 740-741, 746
Expansão em série de MacLaurin, 48
Expressões de teste, 25
Extrapolação de Richardson
derivação, 548-549
integração, 528-532
Extrapolação, 425

F

Fadeeva, D.K., 755
Fadeeva, V.N., 755
Faires, J.D., 86, 89
Fator de amortecimento, 172
Fatorização. *Ver* Decomposição *LU*
Ferziger, J.H., 772
FFT.*Ver* Transformada rápida de Fourier
Finney, R.L., 81
Fletcher, R., 364.*Ver também* Algoritimo de Broyden-Fletcher-Goldfarb-Shanno; Algoritmo de Davidon-Fletcher-Powell
Fletcher-Reeves (método do gradiente conjugado), otimização, 320
Fluxogramas
símbolos usados em, 23-26
Força efetiva, 559-561, 567, 569
Forma de Hessenberg, 677
Forma de Lagrange do resto da série de Taylor, 64
Forma integral do resto da série de Taylor, 63
Forma simétrica, Polinômio de Lagrange, 422
Fórmula de Gauss-Legendre, 623-627
de dois pontos, 536-538
Fórmulas de integração aberta de Adams-Bashforth, 639-641
Fórmulas de integração de Newton-Cotes, 504-524
de ordem superior, 520-521
integração com segmentos desiguais, 521-523
integrais múltiplas, 524-525
métodos de passo múltiplo, EDOs, 636-640
para integração aberta, 524
regra do trapézio, 505-514
regras de Simpson. *Ver* Regras de Simpson
Fórmulas de integração fechada de Adams-Bashforth, 628, 641-642
Forsythe, G.E., 489, 772
Fortran, 38-39, 89 *Ver também* Algoritmo(s)
Freqüência angular, 440
Freqüência natural, 172
Fronteira irregulares, equações elípticas, 714-717
Função de forma, 737-741
Função força externa, 09, 189
Função ímpar, 447
Função incremento, 604
Função objetivo, 290
Função par, 446
Função unidimensional, ao longo da direção gradiente, 317
Funções de penalizações, 332
Funções embutidas, MATLAB, 777
Funções explícitas, 92
Funções implícitas, 92
Funções quadráticas, algoritmos para raízes de, 27-30, 151
Funções transcendentais, 93
Funções, no computador, 30-32
Funções-base, 489

G

Gabel, R. A. 460
Gauss, C. F., 454
Gear, C. W., 628, 696
Gerald, C. F., 89, 185, 489
Gill, P. E., 364
Gladwell, J., 900

Goal seek, Excel, 152-155
Goldberg, D. E., 308
Goldfarb. *Ver* Algoritmo de Broyden-Fletcher-Goldfarb-Shanno
Gradiente conjugado, 333
método de otimização (Flecther-Reeves), 320
Gradientes, definidos, 311-3122
Graus de liberdade, 369
GRG. *Ver* Método da busca gradiente reduzida generalizado
Guest, P. G., 489

H

Hamming, R. W., 86
Harmônicos, 444
Hartley, H. O., 406
Hayt, W. H., 460
Hessianas, gradientes e , 310-315
Hildebrand, F. B., 696
Hipolímnio, 738
Histogramas, 370
Hoffman, J., 84, 726, 772
Holland, J. H., 308
Householder, A. S., 152, 696
Huebner, K. H., 772

I

IF\THEN, 24
Excel, 33
MATLAB, 35
IF\THEN\ELSE, 24
Excel, 33
IF\THEN\ELSEIF, 24
Excel, 33
MATLAB, 35
Imprecisão, 46
IMSL, 39, 97
derivação e integração numérica com, 553-554
equações algébricas lineares e, 259-261
equações diferenciais parciais, 754-756
para ajuste de curva, 465-467
para autovalores, 751
para equações diferencias ordinárias, 751-753
para otimização, 341-342
raízes de equações com, 158-159
Inclinação máxima, 312-319
Incondicionalmente estável, 626
Instabilidade dinâmica, 761
Instabilidade estática, 761
Integração aberta, 504
Adams-Pashforth, 639-642
fórmulas para, 637-642
Newton-Cotes, 523
Integração de Romberg, 532-534
Integração definida, 492
Integração fechada, 505
de Adams-Moulton, 640-641
de Newton-Cotes, 492-494
Integração indefinida, 492, 572
Integração, 6-8. *Ver também* Integrais; Derivação e integração numérica
aberta, 505, 524, 696
cúbica, 427, 432-435
de equações, 528-543
de fórmulas, 520-521, 636-642
definição de, 491
definida, 492
estimador do intervalo, 371
fechada, 505
funções, 738
imprópia, 540-543
indefinida, 492
interpolação, 366, 378, 426-436
linear, 429-430
média de funções contínuas, 497
polinomial, 409-437
prós e contras, 667-668
quadrática, 429-432

quadratura de Gauss. *Ver* Fórmulas fechadas de ordem superior para quadratura de Gauss, 520-521
regra 1/3 de Simpson, 514-517, 520
regra do trapézio. *Ver* Fórmulas de integração de Newton-Cotes
regra estrela de Simpson, 519-521
segmentos desiguais, 521-523
solução analítica, 499-500
Splines-B, 489
teorema fundamental, 499
Integrais imprópias, 540-543
Integrais múltiplas, 524-525
Integrais. *Ver também* Integração
de área, 498
definida, 492
indefinida, 492
múltipla, 524-525
tabela de, 500
volume, 498
Integral de área, 496
Integral dupla, 56, 60
Integral no Volume, 498, 718
Inteiros, 50-51
Interpolação cúbica, 427, 433-434
derivação de, 433
Interpolação linear, 409-411, 428-429
fórmula, 410
método, 108
Interpolação polinomial, 409, 435
coeficientes da, 423
dados igualmente espaçados, 426
inversa, 424
polinômio interpolador de Newton. *Ver* Polinômios interpoladores por diferenças divididas de Newton
polinômios de Lagrange, 409, 419-423
spline, 426-435
Interpolação quadrática, 411-413, 429-432
Intervalo bilateral, 371-372
Intervalo unilateral, 372
Intervalos de confiança
estimativa de, 371-376
na média, 375
para regressão linear, 401-402
Inversa, 194
interpolação, 423-424
transformada de Fourier, 450-451
Inversão de matrizes, 194-197, 236-239, 400
análise de erro e condicionamento do sistema, 239-245
cálculos da resposta a estímulos, 238-239
e mal condicionamento, 240
normas de vetores e matrizes, 240-242
número de condicionamento da matriz, 242-243
resolução de equações lineares usando o computador, 244
Isaacson, E., 696
Isolando raiz no intervalo, otimização e, 303
Isolando raízes de equações em intervalos, 98-113
aproximações iniciais, 113
buscas incrementais, 113
método da bissecção. *Ver* Método da bissecção
método da falsa posição. *Ver* Método da falsa posição
Iteração de Jacobi, 252
Iteração de ponto fixo simples, 117-121
Iteração de ponto fixo, 116-121, 167-168
abordagem gráfica, 118-120
algoritmo para, 120-121
convergência de, 118-120
simples, 116-120
sistemas não lineares, 133-134

J

Jacobi, C. G., 676
Jacobianos, 135
Jordan. *Ver também* Método de Gauss Jordan

K

Kantorovich, L. V., 287
Karp, A. H., 619. *Ver também* Método RK de Cash-Karp
Keller, H. B., 696
Kemmerly, J. E., 460
Kincaid, T., 103, 433, 489
Koopmans, D. C., 287

L

Lagrange, J. L., 287
Lanczos, 454
Lapidus, L., 698, 735, 772
Lasdon, L. S., 332
Lawson, C. L., 489
Legendre. *Ver também* Fórmulas de Gauss-Legendre
Lei de Faraday, 581
Lei de Hook, 170
Lei dos gases ideais, 162-164
Lei dos gases não ideais, 16, 162-164
Leis de conservação
  da carga, 17
  da energia, 17, 706
  da massa, 17, 165, 264
  do momento, 10, 17, 165
  engenharia e, 15-17
Leis de Fick, 581, 759
Leis de Fourier, 706, 763
  do calor, 581
  para a condução do calor, 495, 723, 759
Leis de Kirchhoff, 92, 682
Leis de movimento de Newton, 92, 686
Leis dos Gases, 162-164
Lewis, P. A., 460
Linearização, 580
  de equações não lineares, 389-393
  de uma equação de potência, 391-392
  e regressão, 389-393, 399
Linguagens de computador, 31-39
  Excel, 31-34. *Ver também* VBA
  MATLAB, 35-38
  programas de computador, 22. *Ver também* Algoritmo(s); programas individuais
Little, J.N., 35
Localização de raízes. *Ver também* Raízes de equações; Raízes de polinômio; Raízes de funções quadráticas
Laço, 25-27
  com pós-teste, 26
  com pré-teste, 26
  com teste intermediário, 26
  decisão, 26
  interrupção, 26
  parada, 36
Laço com pós-teste, 25
Laços controlados por contadores, 26
  Excel, 33
  MATLAB, 35, 37
Laços com teste intermediário, 26
Laços de decisão, 26
Lorenz, E., 679-681
Lotka, A.J., 679
Luenberger, D.G., 364
Luther, H. A., 86, 89, 271, 381, 489, 809, 888, 900

M

Mal condicionamento, 79-80
  eliminação de Gauss e, 211-215
  inversão e, 240
  matriz inversa como medida, 240
  número de condicionamento, 80, 242-243
  regressão polinomial, 395
Manipulação aritmética no computador, 56-62
Manutenção, 882
Massa
  balanço de, 17, 92, 758-763
  conservação da, 77, 165, 264

MathWorks, The, 8
MATLAB, 8, 21, 35-39, 56, 774-779
análise de Fourier, 475-476
derivação e integração numérica com, 551-553
equações algébricas lineares e, 257-259
equações diferenciais ordinárias, 663-667
equações diferenciais parciais, 753-754
para ajuste de curvas, 463-465
para autovalores, 676, 684-685
para otimização, 339-340
polinômios, 157
raízes de equações, 155-158
Matriz das propriedades dos elementos, 739
Matriz positiva definida, 250
Matriz(es), 190-197. *Ver também* MATLAB
associativa, 192-193
aumentada, 196
comutativa, 192
de banda, 192
decomposição de Cholesky e, 249-250
diagonal, 192
equações algébricas lineares, 195-196
especial, 246-247
formulação geral para mínimos quadrados lineares, 399-450
identidade, 192
inversão, 194, 196, 236-239, 400-401
multiplicação de, 192-194
notação, 190-191
número de condicionamento, 242-243
quadrada, 191
regras operacionais, 191-195
rigidez, 739, 767
simétrica, 192
traço de, 195
transposta de, 195
triangular inferior, 192
triangular superior, 192
triangular, 192
tri-diagonal, 192, 247-248
Matrizes de banda, 246
Matrizes especiais, 246-250
bibliotecas e pacotes para, 256-261
Matrizes identidades, 192
Matrizes simétricas, 221, 246
Matrizes triangular inferiores, 192
Matrizes triangular superiores, 192
Média
aritmética, 368
intervalos de confiança na, 375-376
funções contínuas, 497
Método clássico de Runge-Kutta de quarta ordem, 609-610
Método da bissecção, 101-108
algoritmo para, 106
cálculos da função minimizadora, 106-107
corretor de Heun, 599
critério de parada, 103-108
estimativa de erro para, 103-106
falsa posição, comparação com, 110
integração de Romberg, 534
iteração de um ponto (ponto fixo), 116
método da falsa posição, 109-110
método de Gauss-Seidel, 251
método de Heun de inicio não-automático, 635
métodos de Newton-Raphson, 122-125
regressão não-linear, 495
Método da busca gradiente reduzida generalizada(GRG), 332
Método da falsa posição, 108-113
análise de erro, 109
armadilhas do 110-112
comparado com método da secante, 127-129
convergência do, 110-111
critério de parada, 110
desvio do, 109
fórmula para, 108
modificado, 112-113
Método da falsa posição modificado, 112-113
Método da potência para autovalores, 585, 663-666

intermediário, 665-666
maior, 663
mais alto, 663-665
mais baixo, 665
menor, 665
Método da potencia, 390-392
Método da secante, 125-130
algoritmo para, 129
convergência de, 127-128
falsa posição, comparação com, 127-128
modificado, 129-130
Método das linhas, 726
Método de Adam de quarta ordem
estabilidade, 645-646
Método de Bairstow, 147-151, 184
Método de Brent, 364
Método de busca aleatória, otimização, 307-308
Método de Crank-Nicolson, equações parabólicas, 730-733
Método de Euler, 14, 584-585, 587-604
algoritmo para, 594-595
análise de erro para, 590-594
erro de arredondamento, 590
erro de truncamento, 594
fórmula para, 140, 587
implícito, 627-628
método de Heun, 597-601
modificações e melhorias, 597-603
regressivo, 626-628
sistemas de EDOs, 613
Método de Euler-Cauchy, 587
Método de Gauss-Jordan, 223-225, 385-386
Método de Gauss-Newton, 401-403
algoritmo para, 489
regressão não linear, 403-406
Método de Gauss-Seidel, 250-259, 400, 709, 729. *Ver também* método de Libmann
critério de convergência para, 252-254
critério de parada, 250
dominância diagonal, 254
e iteração de Jacobi, 252
pseudo código, 255
relaxação, 254
Método de Given, 696
Método de Heun de início não automático, 585, 636-637
Método de Heun, 585, 597-601. *Ver também* Método de Heun de início não automático
algoritmo para, 603
correção do, critério de parada, 599
fórmulas, 598
método do ponto médio, 601-602
Método de Householder, 696
Método de Jacobi, 696
Método de Jenkins-Traub, 152, 215
Método de Laguerre, 152, 185
Método de Levenber-Marquardt, 406
Método de Libmann, 709-711, 729
Método de MacCormack, 726
Método de Maquardt. *Ver também* Método de Levenberg-Marquardt
otimização, 321
Método de Milne, 643-644
Método de Powell, otimização, 310. *Ver também* Algoritmo de Davidon-Fletcher-Powell, otimização
Método de Runge-Kuta Fehlberg, 617-623
Método de Runge-Kutta de quarta ordem, 609-610
equações diferenciais ordinárias, 612-613, 678-679
estudo de casos em, 622, 696
método de Runge Kutta Fehlberg, 619
Método dos coeficientes indeterminados, 534-535
Método dos elementos finitos, 736-757
abordagem geral, 740
de séries de molas, 766-768
discretização, 741, 748
em duas dimensões, 748-749
em uma dimensão, 740-748
equações de elementos, 736-739
montagem, 739, 745, 750, 767-768
pós-processamento, 740
solução, 739, 747, 751

Método dos resíduos ponderados (MRP)
método dos elementos finitos em uma dimensão, 742-745
Método implícito simples, equações parabólicas, 726-729
Método LR, 772
de Rutishauser, 771
Método polinomial, autovalores, 585, 662-663
Método QR, 573
de Francis, 573
Método RK de Cash-Carp, 619-623
Método RK embutido, 619
Método *shooting*, problemas de contorno, 585, 652-654
Método simplex da programação linear, 328-332
implementação, 329-332
variáveis de folga, 328
Método–inclinação, 682
Método/técnica do ponto médio, 585, 601-603
algoritmo para, 603
Métodos abertos
otimização e, 303
raízes de equações, 116-136
Métodos ascendentes, otimização multidimensional, 306
Métodos avançados
ajuste de curvas, 489
geral, 89
métodos de gradiente, otimização, 319-320
para raízes de equações, 185
Métodos de diferença finita, 585, 737
derivadas de ordem superior, aproximações de, 76-77
EDOs(valores de contorno), 653-655
equações do, 699, 705-720
equações parabólicas, 701, 723, 735
Métodos de eliminação, 211-215
Métodos de Newton, otimização, 302-303
Métodos de Newton-Raphson, 121-125, 144, 162-164. *Ver também* Método de Gauss-Newton
algoritmo para, 124-125
análise de erro para, 122-123
armadilhas, 123-124
aspectos computacionais, 124-125
dedução, 122
eliminação de Gauss e, 223
fórmula para, 121
para equações simultâneas não lineares, 135-136
para raízes múltiplas, 131-132
série de Taylor, 121
Métodos de passo múltiplo, EDOs, 585, 628-645
Adams de quarta ordem, 644-646
análise de erro de, 631-634
controle do tamanho do passo e programas computacionais, 636
dedução, 631-633
estabilidade de, 644-646
fórmulas de integração, 636-638
Heun de início não automático, 629-635
método de Milne, 643-644
modificadores, 634-635, 642
ordem superior, 642-645
Métodos de passo único, 584
Métodos de Runge-Kutta adaptativos, 617-623
método de divisão do passo na metade, 618
pseudocódigos, 621
Runge-Kutta Fehlberg, 619-620
Métodos de variável métrica, otimização, 321
Métodos descendentes, otimização multidimensional, 306
Métodos explícitos, equações parabólicas, 723-725, 732
Métodos gradientes, otimização multidimensional, 306, 310-321
Métodos interativos, 48-49. *Ver também* Iteração de ponto fixo, simples
refinamento, 243-244
Métodos não computacionais. *Ver também* Métodos/soluções gráficas
modelagem, computadores e análise de erros, 2

para sistemas de equações algébricas lineares, 189-190
Métodos não gradientes, otimização multidimensional, 306
Métodos numéricos, 12
Métodos Runge-Kuta (RK), 586, 587, 624
adaptativo, 617
clássico de quarta ordem, 609-610
de Butcher de quinta ordem, 611-612
de primeira ordem, 604
de quarta ordem, 609-610
de segunda ordem, 605-607
de terceira ordem, 608-609
método de Euler. *Ver também* Método de Euler
Métodos Runge-Kuta de segunda ordem, 605-606
controle do tamanho do passo, 620-621
de Ralston, 606-607
método de Heun com um único corretor, 606
Métodos unidimensionais
equações parabólicas, comparação de, 732
método dos elementos finitos. *Ver também* Método dos elementos finitos
Métodos/soluções gráficas
eliminação de Gauss, 99-201
para derivação, áreas iguais, 493-495
para integração, 493-495
para raízes de equações, 98-102, 113, 116-120
progamação linear, 325-327
Minimax, 489
critério, 381
princípio, 489
Minimização, 315
cálculos de funções, 106-108
Mínimos quadrados
ajuste, 382-383
regressão, 367, 379-406
Mínimos quadrados linear, 399-402
geral, 376, 399-462
Modelo exponencial, 389
Modelos predador-presa e caos, estudo de casos em, 679-682
Modificadores, 634-635, 687-688
Modificadores corretores, 686-688
Molas, series de, 766-768
Moler, C.B., 35
Momento, 10, 16, 165
abordagem direta, 742
ajuste ótimo da função à solução, 739
condições de contorno, 739
em duas dimensões, 748-749
em uma dimensão, 742-745
estudo de casos em, 766-768
montagem, 739
Montagem, métodos dos elementos finitos
em duas dimensões, 749-750
em uma dimensão, 745-747
equações dos elementos, 739
estudo de casos em, 767-768
Moulton. *Ver também* Fórmulas de integração fechadas de Adams-Moulton
MRP. *Ver também* Método dos resíduos ponderados

N

Na, T.Y., 809, 696
Nicolson. *Ver também* Método de Crank-Nicolson, equações parabólicas
Norma da soma da coluna, 241
Norma da soma da linha, 241-242
Norma de Frobenius, 241
Norma do máximo módulo, 241
Norma espectral, 241
Norma euclidiana, 241
Norma uniforme da matiz, 241
Norma uniforme do vetor, 241
Normalização, representação do número no computador, 51
Normas de matrizes e vetores, 240-242

Normas, vetores e matrizes, 240-242
Nós, 429
Notação posicional, 49
Número de condicionamento, 80, 197
  da matriz, 242-243
Número de manchas solares de Wolf, 475
Números complexos, aproximação de Fourier, 447
Números de Fibonacci, 296-297

O

Operações aritméticas, 56-57
Operações em ponto flutuante, 209-210, 224
Oppenhein, A. V., 460
Ordem de EDPs, 698
Otimização com restrições, 323-344
  programação linear. *Ver também* Programação linear
Otimização de uma única variável, 295
Otimização do custo mínimo
  projeto de um tanque, 345-348
  tratamento da água de dejeto, 349-352
Otimização irrestrita
  multidimensional, 306-321
  unidimensional, 294-303
Otimização multidimensional, 340
Otimização restrita não linear, 332
  Excel para, 335-339
Otimização sem restrições multidimensional, 306-322
  aproximações por diferenças finitas, 314-315
  busca aleatória, 306-308
  de uma variável e padrões de busca, 308-310
  Hessianas, gradientes e, 310-315
  MATLAB para, 340
  máxima inclinação ascendente, 315-319
  método de Marquardt, 321
  método de Newton, 320
  método do gradiente conjugado (Fletcher-Reeves), 320
  métodos da variável métrica, 321
  métodos descendentes, 306
  métodos diretos, 306-310
  métodos do gradiente avançados, 319-321
  métodos não-gradientes, 306
  métodos quasi-Newton, 321
Otimização sem restrições unidimensional, 294-303
  busca da razão áurea, 295-301
  interpolação quadrática, 301-302
  MATLAB para, 339-340
  método de Newton, 302-303
Otimização unidimensional, 339-340
Otimização, 5, 287-364. *Ver também* Otimização irrestrita multidimensional; otimização irrestrita unidimensional
  água de dejeto, tratamento com custo mínimo, 349-353
  com pacotes, 333-343
  custo mínimo, 345-353
  de circuitos, transferência de potencia máxima para, 353-356
  de EDOs, 679
  do custo do pára-quedas, 287-290
  estudo de casos em, 345-357
  Excel para, 333-339
  IMSL para, 341-343
  MATLAB para, 339-340
  multidimensional, 339
  projeto de montain bike, 356-357
  prós e contras, 363
  restrição não linear, 332-333
  restrito. *Ver também* Restrições
  transferência de potência máxima para um circuito, 353-356
  unidimensional, 339-340
Ótimo multimodal, 294
Overflow, 54

P

Pacotes, software. *Ver também* Bibliotecas e pacotes
Parâmetros, 725-726
  estimativas de, 663

Pêndulo oscilante, estudo de caso em, 686-689
Pequeno número de equações, 199-203, 215-216
Período, 439-440
Peterson, D. L., 406
Pinder, G. F., 699, 735, 772
Pivotamento
  coeficientes do, 206
  completo, 216
  eliminação de Gauss e, 215-217
  equações para, 206-207
  no lugar, 217
  parcial, 197, 215-216
  valores do lugar, 49
PL. *Ver também* Programação linear
Placa aquecida com bordas isoladas, equações elípticas, 712-714
Plano de freqüência, 447-449
Polimento de raiz, 143
Polinômios, 94
  MATLAB, 155, 157, 777
  raízes de. *Ver também* Raízes de polinômios
Polinômio interpolador de Lagrange, 378, 409, 419-423
  pseudocódigo, 421
Polinômio interpolador por diferenças divididas de Newton, 378, 409-410
  algoritmo(s) para, 417-419
  dedução do polinômio de Lagrange a partir de, 421
  diferença dividida finita, 413-415
  erro, 415-419
  forma geral, 413-415
  linear, 409-410
  quadrático, 411-413
Polinômios ortogonais, 489
Pontos extremos, 327
  viáveis, 327
População, 371, 470-474
Precisão, 45-46, 87-88
  do computador, 55-56
  dupla, 56, 60
  estendida. *Ver também* Precisão dupla
Preço sombra, 352
Preditores
  equação, 598-599
Prenter, P. M., 489
Press, W. H., 89, 143, 152, 364, 489
Primeira derivada por diferença regressiva, 74
Primeira diferença dividida finita, 73
Primitivação, 500
Princípio da probabilidade máxima, 383-384
Problemas de dois pontos não lineares, condição de contorno, 651-652
Problemas de propagação, 701
  erro de truncamento, 590
Problemas de valor de contorno, 649-655, 740
  autovalores, 658-660
  geral, 649-655
  método *shooting*, 650-652
  métodos de diferença finita, 653-655
  não linear de dois pontos, 651
Problemas ilimitados, 327
Problemas lineares *versus* não lineares, 17
Problemas multidimensionais, definidos, 291
Problemas não lineares, *versus* lineares, 16
Problemas unidimensionais, definidos, 291
Produtos internos, 61
Produtos, multiplicação de matriz e, 192
Programação estruturada, 22-30
Programação linear (PL), 291, 323-332
  Excel para, 333-336
  forma padrão, 323-324
  método simplex, 328-341
  solução gráfica, 325-327
Programação modular, 30-31
Programação não linear, 291
Programação quadrática, 292
Programação. *Ver também* Algoritmo(s)
  estruturada. *Ver também* Programação estruturada
  fluxogramas. *Ver também* Fluxograma
  linear. *Ver também* Programação linear modular
  modular. *Ver também* Programação modular

pacotes e, 21-22
quadrática, 292
Projeto, 16, 168, 290
de circuito elétrico, estudo de casos em, 168-170
de mountain bike, otimização, 356-358
método de, 144-147, 184
Prós e contras, 86-88
ajuste de curvas, 486-487
derivação e integração numérica, 575-576
equações algébricas lineares, 282-283
equações diferenciais ordinárias, 696
equações diferenciais parciais, 771
integração, 575-576
otimização, 363-364
raízes de equações, 184-185
Pseudocódigo, 23-25. *Ver também* Algoritmo(s)
decomposição de Cholesky, 250
eliminação de Gauss, 217-218
epsilon da máquina, 55
Gauss-Seidel, 255
integração de Romberg, 534
integração, segmentos desiguais, 521-523
inversão de matrizes, 238
método RK de Cash-Karp, 621-622
multiplicação de matrizes, 193-194
para a busca da seção áurea, 299-300
para a decomposição de Crout, 236
para a decomposição *LU*, 231, 233-234
para o método de Muller, 147
para resolver EDO por Runge-Kutta adaptativo, 621-622
pivotamento parcial, 217
polinômio de Lagrange, 422
regra do trapézio, aplicação múltipla, 509-512, 529
regras de Simpson, 520, 529
regressão múltipla, 396
regressão polinomial, 395
transformada de Fourier discreta, 452-453
transformada rápida de Fourier, 454
tri-diagonal, algoritmo de Thomas, 247

Q

QD (*ver também* Algoritmo da diferença do quociente)
Quadratura de Gauss, 534-535, 540
análise de erro para, 230
integração de equações, 534-540
método dos coeficiente indeterminados, 534-535
Quadratura, 495
de Gauss. *Ver também* Quadratura de Gauss
Quantidade de calor, 557-559
Splines, 473-474
Quebra
comando de, 38
de Laço 26

R

Rabinovitz, P., 86, 89, 152, 234, 400, 460, 489, 696
Raiz dupla, 130
Raízes de equações, 4, 84-185
buscas incrementais, 113
com bibliotecas e pacotes, 152-154
estudo de caso em, 162-175
iteração de ponto fixo (um ponto), 116-121
iteração de um ponto (ponto fixo), 116-121, 133-135
método da falsa posição para, 108-113
método da secante para estimar, 125-130
métodos abertos. *Ver também* Métodos abertos, otimização e
métodos avançados para, 185
métodos de Newton-Raphson. *Ver também* Métodos de Newton-Raphson
métodos gráficos para obter, 98-102, 113, 118-119
métodos intervalares. *Ver também* Intervalar
polinômios. *Ver também* Raízes de polinômios
prós e contras, 184-185
raízes múltiplas, 130-132
sistemas de equações não-lineares, 132-142

Raízes de funções quadráticas, algoritmos para, 27-30, 151
Raízes de polinômios, 138-159
calculando com polinômios, 140-143
cálculo e derivação polinomial, 140-141
com o MATLAB, 155, 157-158
deflação polinomial, 142-143
método de Bairstow, 147-151
método de Miller, 144-147
polinômios na engenharia e ciência, 138-141
Raízes múltiplas, 130-132
Ramirez, R. W., 452, 469
Rao, S.S., 320-321, 332
Raphson. *Ver também* Métodos de Newton-Raphson
Ratner, M., 912
Raulston, A., 86, 89, 152, 234, 400, 460, 489, 696
Razão áurea, 296
Reatores, estudo de casos em resposta transiente de, 673-678
Rede de borboleta, 457
Regra * de Simpson, 514-517
aplicação múltipla, 516-517, 529
aplicação única, 512
método de Runge-Kutta, relação com, 707
Regra 3/8 (de Simpson), 518-519
Regra da cadeia para derivação, 574
Regra de Boole, 521
Regra de Cramer, 135
Regra do trapézio, 505-513. *Ver também* Fórmulas de integração de Newton-Cotes
aplicação única, 507-509
aplicações múltiplas, 509-511, 529
dois segmentos, 525, 610
segmentos desiguais, 521
Regras de Simpsom, 513-521
regra 1/3, 513-518
regra 3/8, 518-519
pseudo-código para, 520
Regressão, 366-367, 379-406
linear (simples). *Ver também* Regressão linear
linear múltipla, 396-400. *Ver também* Regressão linear múltipla
linear. *Ver também* Regressão linear
mínimos quadrados linear geral, 398-400
não-linear, 403-406
polinomial. *Ver também* Regressão polinomial
trigonométrica, 441
Regressão linear, 376, 379-393
ajuste por mínimos quadrados de uma reta, 382-383
algoritmo para, 386, 389
coeficiente de correlação (<r>), 385
coeficiente de determinação (*), 385, 394-395
critério do melhor ajuste, 380-381
equações normais, 382
estudo de casos em, 470-473
linearização de equações não lineares, 389-393
Regressão linear múltipla, 376, 396-399
equação de potência, 398-399
equações normais, 396-399
pseudocódigo para equações normais, 398
Regressão não linear, 403-406, 463
Regressão polinomial, 376, 393-396
algoritmo para, 395-396
equações normais, 393
erros de arredondamento, 396
estimativas de erro padrão para, 394
estudo de casos em, 461-463
mal condicionamento, 396
pseudocódigo para as equações normais, 395
Regula falsi (falsa posição), 108
Relaxamento, 254
Repetição, 25
Representação binária (base dois), 52
Representação de um número em ponto flutuante, 51-56
Representação lógica, 23-30
Representação na base oito, 50
Representação no espaço dos estados, 680
Representações em base dois (binária), 52
Representações em base oito (octogonal), 50

Resíduos ponderados, método dos, 742-743
modelos matemáticos definidos, 9-10
resolução de problemas de engenharia e, 9-15
Resolvendo equações lineares usando o computador, 244
Respostas transientes, de reatores, estudo de casos em, 673-679
Ressonância, 173
Restrições, 291
limitantes, 326
não-limitantes, 326
Retas/planos nodais, 737
Revelle, C. S., 323
Reversão de bit, para transformada rápida de Fourier, 458
Rice, J. R., 89, 772
Rieeves. *Ver também* Fletcher-Reeves (método do gradiente conjugado)
Rigidez
matriz, 739, 767
EDOs e, 625-628
RK. *Ver também* Método de Runge-Kuta
Roberts, R.A., 460
Runge, 454

S

Schnabel, R.B., 364
Scott, M.R., 696
Segunda lei de Newton, 10, 581
Sela, 313
Sensibilidade ao ruído nos dados, 523-525
dados desigualmente espaçados, 522-525
Séries de Fourier, 444-451, 773-774
contínua, 444-449
espectros na reta, 449-451
na forma complexa, 447
Séries de Taylor, 63-82, 89
estabilidade e condicionamento, 79-80
fórmula de Newton-Raphson, 121
fórmulas de integração de Romberg, 639-642
método de Euler, 590-591
ordem superior, métodos de Runge-Kutta, 597
polinômio interpolador de Newton, 415-416
propagação de erro, 77-80
resto, 63, 68-70
versão para uma única variável, 77-78
versão para várias variáveis, 78-79, 134, 400
Séries infinitas, 60
Shafer, R., 460
Shampine, L.E., 489, 696
Shanno. *Ver também* Algoritmo de Broyden-Fletcher-Goldfarb-Shanno (BFGS)
Símbolos, usados em fluxograma, 23
Simmons, E.F., 89
Simulação de corrente elétrica, EDOs e, 682-686
Sistema sub-especificado, 328
Sistema tri-diagonal, 197
com o algoritmo de Thomas, 247-248
Sistemas complexos, eliminação de Gauss e, 220
Sistemas de equações algébricas lineares, 4-5, 217-329
bibliotecas e pacotes para, 256-261
decomposição *LU*. *Ver também* Decomposição *LU*
eliminação de Gauss. *Ver também* Eliminação de Gauss
forma matricial de, 226-227
Gauss-Seidel, 250-255
matrizes especiais, 246-247
métodos não computacionais para, 217-218
número de condicionamento, 242-243
refinamento iterativo, 243-244
tri-diagonal, 247-248
Sistemas de equações diferenciais ordinárias. *Ver também* Equações diferenciais ordinárias
Sistemas de parâmetros distribuídos, 758
Sistemas massa-mola, 271-273
aplicações em engenharia, 766-768
autovalores e autovetores para, 648-649, 653-655
estudo de caso em, 271-273, 766-767
Sistemas não homogêneos, 654

Sistemas numéricos, 49-51
Sistemas Rígido, 696
MATLAB para, 669-670
Sistemas singulares
funções senoidais, 439-444
integrais, 540-543
Smith, H., 383, 386, 393, 396, 401, 403, 489
Smith, S., 332
Sobre relaxamento simultâneo (SOR), 254
Sobre-relaxamento, 254
Software, 21-40. *Ver também* Algoritmo(s)
pacotes e bibliotecas. *Ver também* Bibliotecas e pacotes
para equações elípticas, 720
programação. *Ver também* Programação
Soluções básicas viáveis, 330
Soluções caóticas, 681
Solver, Excel, 152-154, 334-339, 348, 351-354, 678
Soma dos quadrados total, 383
SOR. *Ver também* Sobre-relaxação simultânea
Spline(s). *Ver também* B-Splines
estudo de casos em, 473-474
interpolação, 378, 426-436
MATLAB, 464-465
transferência de calor, 473-474
Splines cúbicos, 427, 432-434
algoritmos para, 435-436
derivação de, 433
Splines lineares, 428
Splines quadráticos, 429-432
Stasa, F.L., 772
Sub-relaxamento, 254
Subrotinas, computador, 30-31
Substituição regressiva, 205
Swokowski, E.W., 89

T

Tabelas
de derivadas, 500
de integrais, 500
Tanques, otimização do custo mínimo no projeto de, 345-348
Taxa de crescimento máxima possível, 471
Taylor, J.R., 89
Técnica de dizimação na freqüência, 454
Técnica de dizimação no tempo, 454, 458-459
Teorema de Taylor, 63-64
Teorema do Limite Central, 373
Teorema do valor média para integral, 64
Teorema do valor médio
para a derivada, 69, 120
para a integral, 64
Teoremas fundamentais do cálculo integral, 514
Teoremas, valor médio da integral, 64
segundo, 64
Termóclina, 473
Testando hipótese, 400
TFD. *Ver também* Transformada de Fourier discreta
Thomas, G.B., Jr., 89
Thornton, E.A., 772
Trabalho, cálculo do, 563-566
Traço, de matrizes, 195
Transferência de potência máxima, para um circuito, 353-355
Transformação rápida de Fourier(FFT), 454-460
algoritmo de Cooley-Tukey, 454
Sande-Tukey, 454-458
Transformada de Fourier discreta (TFD), 452-453
Transformada e integral de Fourier, 450-452
Transposta, de matrizes, 195
Tratamento de água de dejeto, otimização do custo mínimo em, 349-353
Traub. *Ver também* Método de Jenkins-Traub
Truncamento, 54
Tukey, J. W., 454. *Ver também* Algoritmo FFT de Cooley-Tukey, Algoritmo FFT de Sande-Tukey
Twiddle factor, 456

U

Underflow (“buraco” no zero), 54

V

Valores característicos. *Ver também* Autovalores
Van Valkenburg, M.E., 460
Variância, 369
Variáveis
básicas, 330
de entrada, 330
de folga, simplex LP, 328
dependentes, 9, 671
independentes, 9, 578
não básicas, 329
secundárias, equações elípticas, 711-712
VBA (Visual Basic para aplicações), 31-34, 355-356
VBE. *Ver também* Editor de Visual Basic
Vetor linha, 191
Vetores colunas, 191
Vichnevetsky, R., 699, 772
Viés, 46
Visual Basic para aplicações. *Ver também* VBA
Volterra, Vito, 679. *Ver também* Equações de Lotka-Volterra

W

Wait, R., 772
Wasow, W.R., 772
Watts, H.A., 696
Weatley, P.O., 89, 185, 489
Welch, P.D., 460
Wilkes, J.O., 86, 89, 185, 381, 489, 696, 772
Wilkinson, J.H., 696
Wold, S., 489

Z

Zero, “buraco no”, 54
Zeros, de funções, 295
Zienkiewich, O.C., 772