

Nonlinear Regression

Nonlinear Regression

G. A. F. SEBER and C. J. WILD

Department of Mathematics and Statistics
University of Auckland
Auckland, New Zealand



A JOHN WILEY & SONS, INC., PUBLICATION

A NOTE TO THE READER

This book has been electronically reproduced from digital information stored at John Wiley & Sons, Inc. We are pleased that the use of this new technology will enable us to keep works of enduring scholarly value in print as long as there is a reasonable demand for them. The content of this book is identical to previous printings.

Copyright © 2003 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, e-mail: permreq@wiley.com.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representation or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication is available.

ISBN 0-471-47135-6

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Preface

Some years ago one of the authors (G.A.F.S.) asked a number of applied statisticians how they got on with fitting nonlinear models. The answers were generally depressing. In many cases the available computing algorithms for estimation had unsatisfactory convergence properties, sometimes not converging at all, and there was some uncertainty about the validity of the linear approximation used for inference. Furthermore, parameter estimates sometimes had undesirable properties. Fortunately the situation has improved over recent years because of two major developments. Firstly, a number of powerful algorithms for fitting models have appeared. These have been designed to handle “difficult” models and to allow for the various contingencies that can arise in iterative optimization. Secondly, there has been a new appreciation of the role of curvature in nonlinear modeling and its effect on inferential procedures. Curvature comes in a two-piece suite: intrinsic curvature, which relates to the geometry of the nonlinear model, and parameter-effects curvature, which depends on the parametrization of the model. The effects of these curvatures have recently been studied in relation to inference and experimental design. Apart from a couple of earlier papers, all the published literature on the subject has appeared since 1980, and it continues to grow steadily. It has also been recognized that the curvatures can be regarded as quantities called connection coefficients which arise in differential geometry, the latter providing a unifying framework for the study of curvature. Although we have not pursued these abstract concepts in great detail, we hope that our book will at least provide an introduction and make the literature, which we have found difficult, more accessible.

As we take most of our cues for nonlinear modeling from linear models, it is essential that the reader be familiar with the general ideas of linear regression. The main results used are summarized in Appendix D. In this respect our book can be regarded as a companion volume to Seber [1977, 1984], which deal with linear regression analysis and multivariate methods.

We originally began writing this book with the intention of covering a wide range of nonlinear topics. However, we found that in spite of a smaller literature than that of linear regression or multivariate analysis, the subject is difficult and

diverse, with many applications. We have therefore had to omit a number of important topics, including nonlinear simultaneous equation systems, generalized linear models (and nonlinear extensions), and stochastic approximation. Also, we have been unable to do full justice to the more theoretical econometric literature with its detailed asymptotics (as in Gallant [1987]), and to the wide range of models in the scientific literature at large.

Because of a paucity of books on nonlinear regression when we began this work, we have endeavored to cover both the applied and theoretical ends of the spectrum and appeal to a wide audience. As well as discussing practical examples, we have tried to make the theoretical literature more available to the reader without being too entangled in detail. Unfortunately, most results tend to be asymptotic or approximate in nature, so that asymptotic expansions tend to dominate in some chapters. This has meant some unevenness in the level of difficulty throughout the book. However, although our book is predominantly theoretical, we hope that the balance of theory and practice will make the book useful from both the teaching and the research point of view. It is not intended to be a practical manual on how to do nonlinear fitting; rather, it considers broad aspects of model building and statistical inference.

One of the irritations of fitting nonlinear models is that model fitting generally requires the iterative optimization (minimization or maximization) of functions. Unfortunately, the iterative process often does not converge easily to the desired solution. The optimization algorithms in widespread use are based upon modifications of and approximations to the Newton(–Raphson) and Gauss–Newton algorithms. In unmodified form both algorithms are unreliable. Computational questions are therefore important in nonlinear regression, and we have devoted three chapters to this area. We introduce the basic algorithms early on, and demonstrate their weaknesses. However, rather than break the flow of statistical ideas, we have postponed a detailed discussion of how these algorithms are made robust until near the end of the book. The computational chapters form a largely self-contained introduction to unconstrained optimization.

In Chapter 1, after discussing the notation, we consider the various types of nonlinear model that can arise. Methods of estimating model parameters are discussed in Chapter 2, and some practical problems relating to estimation, like ill-conditioning, are introduced in Chapter 3. Chapter 4 endeavors to summarize some basic ideas about curvature and to bring to notice the growing literature on the subject. In Chapter 5 we consider asymptotic and exact inferences relating to confidence intervals and regions, and hypothesis testing. The role of curvature is again considered, and some aspects of optimal design close the chapter. Autocorrelated errors are the subject of Chapter 6, and Chapters 7, 8, and 9 describe in depth, three broad families of popular models, namely, growth-curve, compartmental, and change-of-phase and spline-regression models. We have not tried to cover every conceivable model, and our coverage thus complements Ratkowsky's [1988] broader description of families of parametric models. Errors-in-variables models are discussed in detail in Chapter 10 for both explicit

and implicit nonlinear models, and nonlinear multivariate models are considered briefly in Chapter 11. Almost by way of an appendix, Chapter 12 gives us a glimpse of some of the basic asymptotic theory, and Chapters 13 to 15 provide an introduction to the growing literature on algorithms for optimization and least squares, together with practical advice on the use of such programs. The book closes with five appendices, an author index, an extensive list of references, and a subject index. Appendix A deals with matrix results, Appendix B gives an introduction to some basic concepts of differential geometry and curvature, Appendix C outlines some theory of stochastic differential equations, Appendix D summarizes linear regression theory, and Appendix E discusses a computational method for handling linear equality constraints. A number of topics throughout the book can be omitted at first reading: these are starred in the text and the contents.

We would like to express our sincere thanks to Mrs. Lois Kennedy and her associates in the secretarial office for typing the difficult manuscript. Thanks go also to Betty Fong for the production of many of the figures. We are grateful to our colleagues and visitors to Auckland for many helpful discussions and suggestions, in particular Alastair Scott, Alan Rogers, Jock MacKay, Peter Phillips, and Ray Carroll. Our thanks also go to Miriam Goldberg, David Hamilton, Dennis Cook, and Douglas Bates for help with some queries in Chapter 4.

We wish to thank the authors, editors, and owners of copyright for permission to reproduce the following published tables and figures: Tables 2.1, 6.1, 6.2, 6.4, 6.5, 7.1, 8.3, 8.4, Figs. 2.1, 6.4, 7.4, 7.5, 7.6, 9.6, 9.10, 9.11 (copyright by the Biometric Society); Tables 2.2, 2.3, 4.1, 4.2, 9.4, Figs. 2.2, 2.3, 6.2, 9.1, 9.2, 9.4, 9.12, 9.16 (copyright by the Royal Statistical Society); Tables 2.5, 5.1, 5.3, 8.1, 9.1, 9.2, 9.5, 10.2, 10.3, 10.4, 10.5, 10.6, 11.1, 11.2, 11.4, 11.5, 11.6, Figs. 2.4, 2.5, 5.1, 5.2, 5.6, 5.7, 5.8, 5.9, 9.5, 9.7, 9.9, 9.17, 10.1, 11.1, 11.2, 11.6, 11.7 (copyright by the American Statistical Association); Table 3.5 (copyright by Akademie-Verlag, Berlin, DDR); Tables 9.3, 11.3, Figs. 5.4, 9.8, 9.14, 11.3, 11.4 (copyright by Biometrika Trust); Figs. 4.3, 5.3 (copyright by the Institute of Mathematical Statistics); Table 2.4 (copyright by the American Chemical Society); Table 3.1, Figs. 3.2, 3.3 (copyright by Plenum Press); Tables 8.2, 10.1, Figs. 8.12, 8.13, 8.14 (copyright by North-Holland Physics Publishing Company); Fig. 9.15 (copyright by Marcel Dekker); Table 5.2 (copyright by Oxford University Press); Figs. 7.2, 7.3 (copyright by Alan R. Liss); Table 7.2 and Fig. 7.7 (copyright by Growth Publishing Company); and Table 6.3 and Fig. 6.3 (copyright by John Wiley and Sons).

G. A. F. SEBER
C. J. WILD

Contents

1	Model Building	1
1.1	Notation, 1	
1.2	Linear and Nonlinear Models, 4	
1.3	Fixed-Regressor Model, 10	
1.3.1	Regressors measured without error, 10	
1.3.2	Conditional regressor models, 11	
1.4	Random-Regressor (Errors-in-Variables) Models, 11	
1.4.1	Functional relationships, 11	
1.4.2	Structural relationships, 12	
1.5	Controlled Regressors with Error, 13	
1.6	Generalized Linear Model, 14	
1.7	Transforming to Linearity, 15	
1.8	Models with Autocorrelated Errors, 18	
1.9	Further Econometric Models, 19	
2	Estimation Methods	21
2.1	Least Squares Estimation, 21	
2.1.1	Nonlinear least squares, 21	
2.1.2	Linear approximation, 23	
2.1.3	Numerical methods, 25	
2.1.4	Generalized least squares, 27	
*2.1.5	Replication and test of fit, 30	
2.2	Maximum-Likelihood Estimation, 32	
2.2.1	Normal errors, 32	

*Starred topics can be omitted at first reading.

- 2.2.2 Nonnormal data, 34
 - 2.2.3 Concentrated likelihood methods, 37
- *2.3 Quasi-likelihood Estimation, 42
- *2.4 LAM Estimator, 48
- *2.5 L_1 -norm Minimization, 50
- 2.6 Robust Estimation, 50
- 2.7 Bayesian Estimation, 52
 - 2.7.1 Choice of prior distributions, 53
 - 2.7.2 Posterior distributions, 55
 - 2.7.3 Highest-posterior-density regions, 63
 - 2.7.4 Normal approximation to posterior density, 64
 - 2.7.5 Test for model adequacy using replication, 65
 - *2.7.6 Polynomial estimators, 66
- 2.8 Variance Heterogeneity, 68
 - 2.8.1 Transformed models, 70
 - a. Box–Cox transformations, 71
 - b. John–Draper transformations, 72
 - 2.8.2 Robust estimation for model A, 73
 - 2.8.3 Inference using transformed data, 74
 - 2.8.4 Transformations to linearity, 75
 - 2.8.5 Further extensions of the Box–Cox method, 76
 - 2.8.6 Weighted least squares: model B, 77
 - 2.8.7 Prediction and transformation bias, 86
 - 2.8.8 Generalized least-squares model, 88

3 Commonly Encountered Problems

91

- 3.1 Problem Areas, 91
- 3.2 Convergence of Iterative Procedures, 91
- 3.3 Validity of Asymptotic Inference, 97
 - 3.3.1 Confidence regions, 97
 - 3.3.2 Effects of curvature, 98
- 3.4 Identifiability and Ill-conditioning, 102
 - 3.4.1 Identifiability problems, 102
 - 3.4.2 Ill-conditioning, 103
 - a. Linear models, 103
 - b. Nonlinear models, 110
 - c. Stable parameters, 117
 - d. Parameter redundancy, 118
 - e. Some conclusions, 126

4 Measures of Curvature and Nonlinearity 127

- 4.1 Introduction, 127
- 4.2 Relative Curvature, 128
 - 4.2.1 Curvature definitions, 129
 - 4.2.2 Geometrical properties, 131
 - *4.2.3 Reduced formulae for curvatures, 138
 - *4.2.4 Summary of formulae, 145
 - 4.2.5 Replication and curvature, 146
 - *4.2.6 Interpreting the parameter-effects array, 147
 - *4.2.7 Computing the curvatures, 150
 - *4.2.8 Secant approximation of second derivatives, 154
- 4.3 Beale's Measures, 157
- *4.4 Connection Coefficients, 159
- 4.5 Subset Curvatures, 165
 - 4.5.1 Definitions, 165
 - *4.5.2 Reduced formulae for subset curvatures, 168
 - *4.5.3 Computations, 170
- 4.6 Analysis of Residuals, 174
 - 4.6.1 Quadratic approximation, 174
 - 4.6.2 Approximate moments of residuals, 177
 - 4.6.3 Effects of curvature on residuals, 178
 - 4.6.4 Projected residuals, 179
 - a. Definition and properties, 179
 - b. Computation of projected residuals, 181
- 4.7 Nonlinearity and Least-Squares Estimation, 181
 - 4.7.1 Bias, 182
 - 4.7.2 Variance, 182
 - 4.7.3 Simulated sampling distributions, 184
 - 4.7.4 Asymmetry measures, 187

5 Statistical Inference 191

- 5.1 Asymptotic Confidence Intervals, 191
- 5.2 Confidence Regions and Simultaneous Intervals, 194
 - 5.2.1 Simultaneous intervals, 194
 - 5.2.2 Confidence regions, 194
 - 5.2.3 Asymptotic likelihood methods, 196
- 5.3 Linear Hypotheses, 197
- 5.4 Confidence Regions for Parameter Subsets, 202
- 5.5 Lack of Fit, 203

- *5.6 Replicated Models, 204
- *5.7 Jackknife Methods, 206
- *5.8 Effects of Curvature on Linearized Regions, 214
 - 5.8.1 Intrinsic curvature, 214
 - 5.8.2 Parameter-effects curvature, 218
 - 5.8.3 Summary of confidence regions, 220
 - 5.8.4 Reparametrization to reduce curvature effects, 222
 - 5.8.5 Curvature and parameter subsets, 227
- 5.9 Nonlinear Hypotheses, 228
 - 5.9.1 Three test statistics, 228
 - 5.9.2 Normally distributed errors, 229
 - 5.9.3 Freedom-equation specification, 232
 - 5.9.4 Comparison of test statistics, 234
 - 5.9.5 Confidence regions and intervals, 235
 - 5.9.6 Multiple hypothesis testing, 235
- 5.10 Exact Inference, 236
 - 5.10.1 Hartley's method, 236
 - 5.10.2 Partially linear models, 240
- 5.11 Bayesian Inference, 245
- *5.12 Inverse Prediction (Discrimination), 245
 - 5.12.1 Single prediction, 245
 - 5.12.2 Multiple predictions, 246
 - 5.12.3 Empirical Bayes interval, 247
- *5.13 Optimal Design, 250
 - 5.13.1 Design criteria, 250
 - 5.13.2 Prior estimates, 255
 - 5.13.3 Sequential designs, 257
 - 5.13.4 Multivariate models, 259
 - 5.13.5 Competing models, 260
 - 5.13.6 Designs allowing for curvature, 260
 - a. Volume approximation, 260
 - b. An example, 264
 - c. Conclusions, 269

6 Autocorrelated Errors

271

- 6.1 Introduction, 271
- 6.2 AR(1) Errors, 275
 - 6.2.1 Preliminaries, 275
 - 6.2.2 Maximum-likelihood estimation, 277

- 6.2.3 Two-stage estimation, 279
- 6.2.4 Iterated two-stage estimation, 280
- 6.2.5 Conditional least squares, 281
- 6.2.6 Choosing between the estimators, 282
- 6.2.7 Unequally spaced time intervals, 285
- 6.3 AR(2) Errors, 286
- 6.4 AR(q_1) Errors, 289
 - 6.4.1 Introduction, 289
 - 6.4.2 Preliminary results, 290
 - 6.4.3 Maximum-likelihood estimation and approximations, 294
 - a. Ignore the determinant, 295
 - b. Approximate the derivative of the determinant, 295
 - c. Asymptotic variances, 296
 - 6.4.4 Two-stage estimation, 301
 - 6.4.5 Choosing a method, 303
 - 6.4.6 Computational considerations, 304
- 6.5 MA(q_2) Errors, 305
 - 6.5.1 Introduction, 305
 - 6.5.2 Two-stage estimation, 306
 - 6.5.3 Maximum-likelihood estimation, 306
- 6.6 ARMA(q_1, q_2) Errors, 307
 - 6.6.1 Introduction, 307
 - 6.6.2 Conditional least-squares method, 310
 - 6.6.3 Other estimation procedures, 314
- 6.7 Fitting and Diagnosis of Error Processes, 318
 - 6.7.1 Choosing an error process, 318
 - 6.7.2 Checking the error process, 321
 - a. Overfitting, 321
 - b. Use of noise residuals, 322

7 Growth Models

325

- 7.1 Introduction, 325
- 7.2 Exponential and Monomolecular Growth Curves, 327
- 7.3 Sigmoidal Growth Models, 328
 - 7.3.1 General description, 328
 - 7.3.2 Logistic (autocatalytic) model, 329
 - 7.3.3 Gompertz growth curve, 330
 - 7.3.4 Von Bertalanffy model, 331

- 7.3.5 Richards curve, 332
- 7.3.6 Starting values for fitting Richards models, 335
- 7.3.7 General procedure for sigmoidal curves, 337
- 7.3.8 Weibull model, 338
- 7.3.9 Generalized logistic model, 339
- 7.3.10 Fletcher family, 339
- 7.3.11 Morgan–Mercer–Flodin (MMF) family, 340
- 7.4 Fitting Growth Models: Deterministic Approach, 342
- 7.5 Stochastic Growth-Curve Analysis, 344
 - 7.5.1 Introduction, 344
 - 7.5.2 Rates as functions of time, 346
 - a. Uncorrelated error process, 347
 - b. Autocorrelated error processes, 348
 - 7.5.3 Rates as functions of size, 353
 - a. A tractable differential equation, 354
 - b. Approximating the error process, 356
- 7.6 Yield–Density Curves, 360
 - 7.6.1 Preliminaries, 360
 - 7.6.2 Bleasdale–Nelder model, 363
 - 7.6.3 Holliday model, 364
 - 7.6.4 Choice of model, 364
 - 7.6.5 Starting values, 365

8 Compartmental Models

367

- 8.1 Introduction, 367
- 8.2 Deterministic Linear Models, 370
 - 8.2.1 Linear and nonlinear models, 370
 - 8.2.2 Tracer exchange in steady-state systems, 372
 - 8.2.3 Tracers in nonsteady-state linear systems, 375
- 8.3 Solving the Linear Differential Equations, 376
 - 8.3.1 The solution and its computation, 376
 - 8.3.2 Some compartmental structures, 383
 - 8.3.3 Properties of compartmental systems, 384
- *8.4 Identifiability in Deterministic Linear Models, 386
- 8.5 Linear Compartmental Models with Random Error, 393
 - 8.5.1 Introduction, 393
 - 8.5.2 Use of the chain rule, 395
 - 8.5.3 Computer-generated exact derivatives, 396
 - a. Method of Bates et al., 396

- b. Method of Jennrich and Bright, 400
 - c. General algorithm, 401
 - 8.5.4 The use of constraints, 402
 - 8.5.5 Fitting compartmental models without using derivatives, 404
 - 8.5.6 Obtaining initial parameter estimates, 406
 - a. All compartments observed with zero or linear inputs, 406
 - b. Some compartments unobserved, 407
 - c. Exponential peeling, 407
 - 8.5.7 Brunhilda example revisited, 410
 - 8.5.8 Miscellaneous topics, 412
- 8.6 More Complicated Error Structures, 413
- *8.7 Markov-Process Models, 415
 - 8.7.1 Background theory, 415
 - a. No environmental input, 416
 - b. Input from the environment, 420
 - 8.7.2 Computational methods, 423
 - a. Unconditional generalized least squares, 423
 - b. Conditional generalized least squares, 424
 - 8.7.3 Discussion, 429
- 8.8 Further Stochastic Approaches, 431

9 Multiphase and Spline Regressions 433

- 9.1 Introduction, 433
- 9.2 Noncontinuous Change of Phase for Linear Regimes, 438
 - 9.2.1 Two linear regimes, 438
 - 9.2.2 Testing for a two-phase linear regression, 440
 - 9.2.3 Parameter inference, 445
 - 9.2.4 Further extensions, 446
- 9.3 Continuous Case, 447
 - 9.3.1 Models and inference, 447
 - 9.3.2 Computation, 455
 - 9.3.3 Segmented polynomials, 457
 - a. Inference, 460
 - b. Computation, 463
 - 9.3.4 Exact tests for no change of phase in polynomials, 463
- 9.4 Smooth Transitions between Linear Regimes, 465
 - 9.4.1 The sgn formulation, 465
 - 9.4.2 The max-min formulation, 471

- a. Smoothing $\max(0, z)$, 472
 - b. Limiting form for $\max\{z_i\}$, 474
 - c. Extending sgn to a vector of regressors, 476
 - 9.4.3 Examples, 476
 - 9.4.4 Discussion, 480
- 9.5 Spline Regression, 481
 - 9.5.1 Fixed and variable knots, 481
 - a. Fixed knots, 484
 - b. Variable knots, 484
 - 9.5.2 Smoothing splines, 486

***10 Errors-In-Variables Models 491**

- 10.1 Introduction, 491
- 10.2 Functional Relationships: Without Replication, 492
- 10.3 Functional Relationships: With Replication, 496
- 10.4 Implicit Functional Relationships: Without Replication, 501
- 10.5 Implicit Functional Relationships: With Replication, 508
 - 10.5.1 Maximum-likelihood estimation, 508
 - 10.5.2 Bayesian estimation, 510
- 10.6 Implicit Relationships with Some Unobservable Responses, 516
 - 10.6.1 Introduction, 516
 - 10.6.2 Least-squares estimation, 516
 - 10.6.3 The algorithm, 519
- 10.7 Structural and Ultrastructural Models, 523
- 10.8 Controlled Variables, 525

11 Multiresponse Nonlinear Models 529

- 11.1 General Model, 529
- 11.2 Generalized Least Squares, 531
- 11.3 Maximum-Likelihood Inference, 536
 - 11.3.1 Estimation, 536
 - 11.3.2 Hypothesis testing, 538
- 11.4 Bayesian Inference, 539
 - 11.4.1 Estimates from posterior densities, 539
 - 11.4.2 H.P.D. regions, 542
 - 11.4.3 Missing observations, 544
- *11.5 Linear Dependencies, 545
 - 11.5.1 Dependencies in expected values, 545
 - 11.5.2 Dependencies in the data, 546

- 11.5.3 Eigenvalue analysis, 547
- 11.5.4 Estimation procedures, 549
- *11.6 Functional Relationships, 557

***12 Asymptotic Theory 563**

- 12.1 Introduction, 563
- 12.2 Least-Squares Estimation, 563
 - 12.2.1 Existence of least-squares estimate, 563
 - 12.2.2 Consistency, 564
 - 12.2.3 Asymptotic normality, 568
 - 12.2.4 Effects of misspecification, 572
 - 12.2.5 Some extensions, 574
 - 12.2.6 Asymptotics with vanishingly small errors, 575
- 12.3 Maximum-Likelihood Estimation, 576
- 12.4 Hypothesis Testing, 576
- 12.5 Multivariate Estimation, 581

13 Unconstrained Optimization 587

- 13.1 Introduction, 587
- 13.2 Terminology, 588
 - 13.2.1 Local and global minimization, 588
 - 13.2.2 Quadratic functions, 590
 - 13.2.3 Iterative algorithms, 593
 - a. Convergence rates, 593
 - b. Descent methods, 594
 - c. Line searches, 597
- 13.3 Second-Derivative (Modified Newton) Methods, 599
 - 13.3.1 Step-length methods, 599
 - a. Directional discrimination, 600
 - b. Adding to the Hessian, 602
 - 13.3.2 Restricted step methods, 603
- 13.4 First-Derivative Methods, 605
 - 13.4.1 Discrete Newton methods, 605
 - 13.4.2 Quasi-Newton methods, 605
 - *13.4.3 Conjugate-gradient methods, 609
- 13.5 Methods without Derivatives, 611
 - 13.5.1 Nonderivative quasi-Newton methods, 611
 - *13.5.2 Direction-set (conjugate-direction) methods, 612
 - 13.5.3 Direct search methods, 615

- 13.6 Methods for Nonsmooth Functions, 616
- 13.7 Summary, 616

14 Computational Methods for Nonlinear Least Squares 619

- 14.1 Gauss–Newton Algorithm, 619
- 14.2 Methods for Small-Residual Problems, 623
 - 14.2.1 Hartley’s method, 623
 - 14.2.2 Levenberg–Marquardt methods, 624
 - 14.2.3 Powell’s hybrid method, 627
- 14.3 Large-Residual Problems, 627
 - 14.3.1 Preliminaries, 627
 - 14.3.2 Quasi-Newton approximation of $A(\theta)$, 628
 - 14.3.3 The Gill–Murray method, 633
 - a. Explicit second derivatives, 636
 - b. Finite-difference approximation, 636
 - c. Quasi-Newton approximation, 637
 - 14.3.4 Summary, 639
- 14.4 Stopping Rules, 640
 - 14.4.1 Convergence criteria, 640
 - 14.4.2 Relative offset, 641
 - 14.4.3 Comparison of criteria, 644
- 14.5 Derivative-Free Methods, 646
- 14.6 Related Problems, 650
 - 14.6.1 Robust loss functions, 650
 - 14.6.2 L_1 -minimization, 653
- 14.7 Separable Least-Squares Problems, 654
 - 14.7.1 Introduction, 654
 - 14.7.2 Gauss–Newton for the concentrated sum of squares, 655
 - 14.7.3 Intermediate method, 657
 - 14.7.4 The NIPALS method, 659
 - 14.7.5 Discussion, 660

15 Software Considerations 661

- 15.1 Selecting Software, 661
 - 15.1.1 Choosing a method, 661
 - 15.1.2 Some sources of software, 663
 - a. Commercial libraries, 663
 - b. Noncommercial libraries, 663
 - c. ACM algorithms, 663

- 15.2 User-Supplied Constants, 664
 - 15.2.1 Starting values, 665
 - a. Initial parameter value, 665
 - b. Initial approximation to the Hessian, 666
 - 15.2.2 Control constants, 666
 - a. Step-length accuracy, 666
 - b. Maximum step length, 666
 - c. Maximum number of function evaluations, 667
 - 15.2.3 Accuracy, 667
 - a. Precision of function values, 667
 - b. Magnitudes, 668
 - 15.2.4 Termination criteria (stopping rules), 668
 - a. Convergence of function estimates, 669
 - b. Convergence of gradient estimates, 670
 - c. Convergence of parameter estimates, 670
 - d. Discussion, 671
 - e. Checking for a minimum, 672
- 15.3 Some Causes and Modes of Failure, 672
 - 15.3.1 Programming error, 672
 - 15.3.2 Overflow and underflow, 673
 - 15.3.3 Difficulties caused by bad scaling (parametrization), 673
- 15.4 Checking the Solution, 675

APPENDIXES

A. Vectors and Matrices

677

- A1 Rank, 677
- A2 Eigenvalues, 677
- A3 Patterned matrices, 678
- A4 Positive definite and semidefinite matrices, 678
- A5 Generalized inverses, 679
- A6 Ellipsoids, 679
- A7 Optimization, 679
- A8 Matrix factorizations, 680
- A9 Multivariate t -distribution, 681
- A10 Vector differentiation, 682
- A11 Projection matrices, 683
- A12 Quadratic forms, 684

A13	Matrix operators, 684	
A14	Method of scoring, 685	
B.	Differential Geometry	687
B1	Derivatives for curves, 687	
B2	Curvature, 688	
B3	Tangent planes, 690	
B4	Multiplication of 3-dimensional arrays, 691	
B5	Invariance of intrinsic curvature, 692	
C.	Stochastic Differential Equations	695
C1	Rate equations, 695	
C2	Proportional rate equations, 697	
C3	First-order equations, 698	
D.	Multiple Linear Regression	701
D1	Estimation, 701	
D2	Inference, 702	
D3	Parameter subsets, 703	
E.	Minimization Subject to Linear Constraints	707
	References	711
	Author Index	745
	Subject Index	753

Nonlinear Regression