

Condensed Table of Contents for
Introduction to Stochastic Search and Optimization:
Estimation, Simulation, and Control
by J. C. Spall

© John Wiley and Sons, Inc., 2003

Preface	xiii
1. Stochastic Search and Optimization: Motivation and Supporting Results	1
1.1 Introduction	1
1.2 Some Principles of Stochastic Search and Optimization	12
1.3 Gradients, Hessians, and Their Connection to Optimization of “Smooth” Functions	20
1.4 Deterministic Search and Optimization: Steepest Descent and Newton–Raphson Search	22
1.5 Concluding Remarks	30
Exercises.....	31
2. Direct Methods for Stochastic Search	34
2.1 Introduction	34
2.2 Random Search with Noise-Free Loss Measurements	36
2.3 Random Search with Noisy Loss Measurements	50
2.4 Nonlinear Simplex (Nelder–Mead) Algorithm	55
2.5 Concluding Remarks	60
Exercises.....	61
3. Recursive Estimation for Linear Models	65
3.1 Formulation for Estimation with Linear Models.....	65
3.2 Least-Mean-Squares and Recursive-Least-Squares for Static θ	72
3.3 LMS, RLS, and Kalman Filter for Time-Varying θ	83
3.4 Case Study: Analysis of Oboe Reed Data	88
3.5 Concluding Remarks	92
Exercises.....	93
4. Stochastic Approximation for Nonlinear Root-Finding	95
4.1 Introduction	95
4.2 Potpourri of Stochastic Approximation Examples	98
4.3 Convergence of Stochastic Approximation.....	104
4.4 Asymptotic Normality and Choice of Gain Sequence	111
4.5 Extensions to Basic Stochastic Approximation.....	115

4.6	Concluding Remarks	121
	Exercises.....	122
5.	Stochastic Gradient Form of Stochastic Approximation.....	126
5.1	Root-Finding Stochastic Approximation as a Stochastic Gradient Method	126
5.2	Neural Network Training	138
5.3	Discrete-Event Dynamic Systems	142
5.4	Image Restoration.....	144
5.5	Concluding Remarks	147
	Exercises.....	147
6.	Stochastic Approximation and the Finite-Difference Method	150
6.1	Introduction and Contrast of Gradient-Based and Gradient-Free Algorithms	150
6.2	Some Motivating Examples for Gradient-Free Stochastic Approximation	153
6.3	Finite-Difference Algorithm.....	157
6.4	Convergence Theory	158
6.5	Asymptotic Normality	162
6.6	Practical Selection of Gain Sequences	164
6.7	Several Finite-Difference Examples.....	166
6.8	Some Extensions and Enhancements to the Finite-Difference Algorithm	172
6.9	Concluding Remarks	174
	Exercises.....	174
7.	Simultaneous Perturbation Stochastic Approximation	176
7.1	Background.....	177
7.2	Form and Motivation for Standard SPSA Algorithm.....	178
7.3	Basic Assumptions and Supporting Theory for Convergence	182
7.4	Asymptotic Normality and Efficiency Analysis.....	186
7.5	Practical Implementation.....	188
7.6	Numerical Examples	191
7.7	Some Extensions: Optimal Perturbation Distribution; One-Measurement Form; Global, Discrete, and Constrained Optimization.....	193
7.8	Adaptive SPSA.....	196
7.9	Concluding Remarks	203
7.10	Appendix: Conditions for Asymptotic Normality	204
	Exercises.....	204

8.	Annealing-Type Algorithms	208
8.1	Introduction to Simulated Annealing and Motivation from the Physics of Cooling	208
8.2	Simulated Annealing Algorithm.....	211
8.3	Some Examples	217
8.4	Global Optimization via Annealing Algorithms Based on Stochastic Approximation	221
8.5	Concluding Remarks	225
8.6	Appendix: Convergence Theory for Simulated Annealing Based on Stochastic Approximation	226
	Exercises.....	228
9.	Evolutionary Computation I: Genetic Algorithms	231
9.1	Introduction	231
9.2	Some Historical Perspective and Motivating Applications.....	235
9.3	Coding of Elements for Searching	237
9.4	Standard Genetic Algorithm Operations	242
9.5	Overview of Basic GA Search Approach.....	246
9.6	Practical Guidance and Extensions: Coefficient Values, Constraints, Noisy Fitness Evaluations, Local Search, and Parent Selection.....	247
9.7	Examples	250
9.8	Concluding Remarks	255
	Exercises.....	256
10.	Evolutionary Computation II: General Methods and Theory	259
10.1	Introduction	259
10.2	Overview of Evolution Strategy and Evolutionary Programming with Comparisons to Genetic Algorithms.....	260
10.3	Schema Theory	263
10.4	What Makes a Problem Hard?.....	266
10.5	Convergence Theory	268
10.6	No Free Lunch Theorems	273
10.7	Concluding Remarks	275
	Exercises.....	276
11.	Reinforcement Learning via Temporal Differences	278
11.1	Introduction	278
11.2	Delayed Reinforcement and Formulation for Temporal Difference Learning	280
11.3	Basic Temporal Difference Algorithm.....	283
11.4	Batch and Online Implementations of TD Learning	287
11.5	Some Examples	289

11.6	Connections to Stochastic Approximation	295
11.7	Concluding Remarks	297
	Exercises.....	298
12.	Statistical Methods for Optimization in Discrete Problems	300
12.1	Introduction to Multiple Comparisons Over a Finite Set.....	301
12.2	Statistical Comparisons Test Without Prior Information.....	306
12.3	Multiple Comparisons Against One Candidate with Known Noise Variance(s).....	310
12.4	Multiple Comparisons Against One Candidate with Unknown Noise Variance(s).....	319
12.5	Extensions to Bonferroni Inequality; Ranking and Selection Methods in Optimization Over a Finite Set	322
12.6	Concluding Remarks	325
	Exercises.....	326
13.	Model Selection and Statistical Information	329
13.1	Bias–Variance Tradeoff.....	330
13.2	Model Selection: Cross-Validation	340
13.3	The Information Matrix: Applications and Resampling-Based Computation.....	349
13.4	Concluding Remarks	363
	Exercises.....	364
14.	Simulation-Based Optimization I: Regeneration, Common Random Numbers, and Selection Methods	367
14.1	Background.....	367
14.2	Regenerative Systems.....	372
14.3	Optimization with Finite-Difference and Simultaneous Perturbation Gradient Estimators.....	382
14.4	Common Random Numbers	385
14.5	Selection Methods for Optimization with Discrete-Valued θ	398
14.6	Concluding Remarks	405
	Exercises.....	406
15.	Simulation-Based Optimization II: Stochastic Gradient and Sample Path Methods	409
15.1	Framework for Gradient Estimation.....	409
15.2	Pure Likelihood Ratio/Score Function and Pure Infinitesimal Perturbation Analysis.....	417
15.3	Gradient Estimation Methods in Root-Finding Stochastic Approximation: The Hybrid LR/SF and IPA Setting.....	420
15.4	Sample Path Optimization.....	425

15.5 Concluding Remarks	432
Exercises.....	433
16. Markov Chain Monte Carlo	436
16.1 Background.....	436
16.2 Metropolis–Hastings Algorithm.....	440
16.3 Gibbs Sampling	445
16.4 Sketch of Theoretical Foundation for Gibbs Sampling.....	450
16.5 Some Examples of Gibbs Sampling	453
16.6 Applications in Bayesian Analysis.....	457
16.7 Concluding Remarks	461
Exercises.....	462
17. Optimal Design for Experimental Inputs.....	464
17.1 Introduction	464
17.2 Linear Models.....	473
17.3 Response Surface Methodology	486
17.4 Nonlinear Models	489
17.5 Concluding Remarks	500
17.6 Appendix: Optimal Design in Dynamic Models	501
Exercises.....	502
Appendix A. Selected Results from Multivariate Analysis.....	505
A.1 Multivariate Calculus and Analysis	505
A.2 Some Useful Results in Matrix Theory	511
Exercises.....	514
Appendix B. Some Basic Tests in Statistics.....	515
B.1 Standard One-Sample Test.....	515
B.2 Some Basic Two-Sample Tests	518
B.3 Comments on Other Aspects of Statistical Testing.....	524
Exercises.....	525
Appendix C. Probability Theory and Convergence	526
C.1 Basic Properties	526
C.2 Convergence Theory	529
Exercises.....	537
Appendix D. Random Number Generation	538
D.1 Background and Introduction to Linear Congruential Generators	538

D.2 Transformation of Uniform Random Numbers to Other Distributions	542
Exercises.....	545
Appendix E. Markov Processes.....	547
E.1 Background on Markov Processes	547
E.2 Discrete Markov Chains	548
Exercises.....	551
Answers to Selected Exercises	552
References	558
Frequently Used Notation.....	580
Index	583