

Contents

Preface	xi
1 Introduction	1
1.1 Overview of Topics	1
1.2 Examples of Problems in Biological Modeling	2
1.2.1 Optimization	2
1.2.2 Simulation and Sampling	4
1.2.3 Parameter-Tuning	8
I MODELS FOR OPTIMIZATION	13
2 Classic Discrete Optimization Problems	15
2.1 Graph Problems	16
2.1.1 Minimum Spanning Trees	16
2.1.2 Shortest Path Problems	19
2.1.3 Max Flow/Min Cut	21
2.1.4 Matching	23
2.2 String and Sequence Problems	24
2.2.1 Longest Common Subsequence	25
2.2.2 Longest Common Substring	26
2.2.3 Exact Set Matching	27
2.3 Mini Case Study: Intraspecies Phylogenetics	28
3 Hard Discrete Optimization Problems	35
3.1 Graph Problems	36
3.1.1 Traveling Salesman Problems	36
3.1.2 Hard Cut Problems	37
3.1.3 Vertex Cover, Independent Set, and k-Clique	38
3.1.4 Graph Coloring	39
3.1.5 Steiner Trees	40
3.1.6 Maximum Subgraph or Induced Subgraph with Property II	42
3.2 String and Sequence Problems	42
3.2.1 Longest Common Subsequence	42
3.2.2 Shortest Common Supersequence/Superstring	43

3.3	Set Problems	44
3.3.1	Minimum Test Set	44
3.3.2	Minimum Set Cover	45
3.4	Hardness Reductions	45
3.5	What to Do with Hard Problems	46
4	Case Study: Sequence Assembly	57
4.1	Sequencing Technologies	57
4.1.1	Maxam–Gilbert	57
4.1.2	Sanger Dideoxy	59
4.1.3	Automated Sequencing	61
4.1.4	What About Bigger Sequences?	63
4.2	Computational Approaches	64
4.2.1	Sequencing by Hybridization	64
4.2.2	Eulerian Path Method	66
4.2.3	Shotgun Sequencing	67
4.2.4	Double-Barreled Shotgun	69
4.3	The Future?	71
4.3.1	SBH Revisited	71
4.3.2	New Sequencing Technologies	72
5	General Continuous Optimization	75
5.1	Bisection Method	76
5.2	Secant Method	78
5.3	Newton–Raphson	80
5.4	Newton–Raphson with Black-Box Functions	84
5.5	Multivariate Functions	85
5.6	Direct Methods for Optimization	89
5.6.1	Steepest Descent	89
5.6.2	The Levenberg–Marquardt Method	90
5.6.3	Conjugate Gradient	91
6	Constrained Optimization	95
6.1	Linear Programming	96
6.1.1	The Simplex Method	97
6.1.2	Interior Point Methods	104
6.2	Primals and Duals	107
6.3	Solving Linear Programs in Practice	107
6.4	Nonlinear Programming	108
II	SIMULATION AND SAMPLING	113
7	Sampling from Probability Distributions	115
7.1	Uniform Random Variables	115
7.2	The Transformation Method	116
7.2.1	Transformation Method for Joint Distributions	119
7.3	The Rejection Method	121
7.4	Sampling from Discrete Distributions	124

8	Markov Models	129
8.1	Time Evolution of Markov Models	131
8.2	Stationary Distributions and Eigenvectors	134
8.3	Mixing Times	138
9	Markov Chain Monte Carlo Sampling	141
9.1	Metropolis Method	141
9.1.1	Generalizing the Metropolis Method	146
9.1.2	Metropolis as an Optimization Method	147
9.2	Gibbs Sampling	149
9.2.1	Gibbs Sampling as an Optimization Method	152
9.3	Importance Sampling	154
9.3.1	Umbrella Sampling	155
9.3.2	Generalizing to Other Samplers	156
10	Mixing Times of Markov Models	159
10.1	Formalizing Mixing Time	160
10.2	The Canonical Path Method	161
10.3	The Conductance Method	166
10.4	Final Comments	170
11	Continuous-Time Markov Models	173
11.1	Definitions	173
11.2	Properties of CTMMs	175
11.3	The Kolmogorov Equations	178
12	Case Study: Molecular Evolution	185
12.1	DNA Base Evolution	185
12.1.1	The Jukes–Cantor (One-Parameter) Model	185
12.1.2	Kimura (Two-Parameter) Model	188
12.2	Simulating a Strand of DNA	191
12.3	Sampling from Whole Populations	192
12.4	Extensions of the Coalescent	195
12.4.1	Variable Population Sizes	196
12.4.2	Population Substructure	197
12.4.3	Diploid Organisms	198
12.4.4	Recombination	198
13	Discrete Event Simulation	201
13.1	Generalized Discrete Event Modeling	203
13.2	Improving Efficiency	204
13.3	Real-World Example: Hard-Sphere Model of Molecular Collision Dynamics	206
13.4	Supplementary Material: Calendar Queues	209
14	Numerical Integration 1: Ordinary Differential Equations	211
14.1	Finite Difference Schemes	213
14.2	Forward Euler	214
14.3	Backward Euler	217

14.4	Higher-Order Single-Step Methods	219
14.5	Multistep Methods	221
14.6	Step Size Selection	223
15	Numerical Integration 2: Partial Differential Equations	227
15.1	Problems of One Spatial Dimension	228
15.2	Initial Conditions and Boundary Conditions	230
15.3	An Aside on Step Sizes	233
15.4	Multiple Spatial Dimensions	233
15.5	Reaction–Diffusion Equations	234
15.6	Convection	237
16	Numerical Integration 3: Stochastic Differential Equations	241
16.1	Modeling Brownian Motion	241
16.2	Stochastic Integrals and Differential Equations	242
16.3	Integrating SDEs	245
16.4	Accuracy of Stochastic Integration Methods	248
16.5	Stability of Stochastic Integration Methods	249
17	Case Study: Simulating Cellular Biochemistry	253
17.1	Differential Equation Models	253
17.2	Markov Models Methods	256
17.3	Hybrid Models	259
17.4	Handling Very Large Reaction Networks	260
17.5	The Future of Whole-Cell Models	262
17.6	An Aside on Standards and Interfaces	263
III	PARAMETER-TUNING	265
18	Parameter-Tuning as Optimization	267
18.1	General Optimization	268
18.2	Constrained Optimization	269
18.3	Evaluating an Implicitly Specified Function	271
19	Expectation Maximization	275
19.1	The “Expectation Maximization Algorithm”	277
19.2	EM Theory	278
19.3	Examples	280
20	Hidden Markov Models	291
20.1	Applications of HMMs	292
20.2	Algorithms for HMMs	295
20.2.1	Problem 1: Optimizing State Assignments	295
20.2.2	Problem 2: Evaluating Output Probability	297
20.2.3	Problem 3: Training the Model	299
20.3	Parameter-Tuning Example: Motif-Finding by HMM	303
21	Linear System-Solving	309
21.1	Gaussian Elimination	310
21.1.1	Pivoting	312

21.2	Iterative Methods	316
21.3	Krylov Subspace Methods	317
21.3.1	Preconditioners	319
21.4	Overdetermined and Underdetermined Systems	320
22	Interpolation and Extrapolation	323
22.1	Polynomial Interpolation	326
22.1.1	Neville's Algorithm	326
22.2	Fitting to Lower-Order Polynomials	329
22.3	Rational Function Interpolation	330
22.4	Splines	331
22.5	Multidimensional Interpolation	334
22.6	Interpolation with Arbitrary Families of Curves	334
22.7	Extrapolation	337
22.7.1	Richardson Extrapolation	337
22.7.2	Aitken's δ^2 Process	338
23	Case Study: Inferring Gene Regulatory Networks	341
23.1	Coexpression Models	342
23.1.1	Measures of Similarity	342
23.1.2	Finding a Union-of-Cliques Graph	344
23.2	Bayesian Graphical Models	347
23.2.1	Defining a Probability Function	347
23.2.2	Finding the Network	349
23.3	Kinetic Models	351
24	Model Validation	355
24.1	Measures of Goodness	355
24.2	Accuracy, Sensitivity, and Specificity	358
24.3	Cross-Validation	361
24.4	Sensitivity Analysis	362
24.5	Modeling and the Scientific Method	363
	References	367
	Index	377