Acceleration, 211
Accuracy. See also Errors
   in Adams-Bashforth methods, 223
   adaptive methods, 224
   centered difference, 84–85, 229–232, 235
   first and second order, 157, 215, 218, 233, 238
   and forward/backward Euler, 218, 222
   in model validation, 358
   of Neville’s algorithm, 327
   and Newton-Raphson, 84–85
   and partial differential equations, 233
   and Runge-Kutta methods, 221, 225, 237
   of secant versus bisection, 79
   and stability, 221, 223, 251
   and stochastic differential equations, 248–249
   and time, 219, 233
Adams-Bashforth methods, 221–223, 225
Adams-Moulton scheme, 225
Adaptive methods, 224–225
Affine method, 104–107, 110
Aitken’s $\delta^2$ process, 338–340
Alleles, 28–33, 280–289
All-pairs shortest path, 21
Amino acids. See also Protein folding
   contact energies, 5–7
   and HMMs, 297–299
   and Metropolis method, 145
   proline cis-trains isomerization, 180–182
   and proteases, 9
Animation, 333
Annealing. See also Simulated annealing
Antibiotics, 45
Approximation
   centered difference, 84–85, 89, 229–232, 235
   and extrapolation, 337–340
   forward difference, 84
   and forward Euler, 214, 224
   and interpolation, 327–337
   and reaction-diffusion equations, 237
   for step size, 224
   with Taylor series, 80–81, 85, 232
Approximation algorithms. See also Traveling salesman
   and branch-and-bound algorithm, 50
   description, 47–59
   and intractability, 47–49, 50, 55
   reference, 55
   traveling salesman, 36, 48, 54
   and vertex cover, 47, 50–51
Approximation schemes, 48
Automated sequencing, 61–63
Backward algorithm, 298–299
Backward error, 77
Backward Euler, 217–219, 222
Bacteria
   antibiotic sensitivity, 45
   bacterial artificial chromosome (BAC), 64
Barrier methods, 104
Baum-Welch algorithm, 300–307
Bayesian models, 347–350, 353, 356
   additional sources, 353
Bellman-Ford algorithm, 20–21
   background, 33
Best-fit
   in interpolation, 336
   and least-squares, 356
   in parameter-tuning, 275
Bias. See also Model validation
   and gene network, 349
   and HMMs, 360
   and importance sampling, 154
   and parameter choices, 362
   unintended, 365
Biconjugate gradient, 319
Bilinear interpolation, 334
Billiard ball model, 206–209
Binary search, 338
Biochemical processes. See also Evolution; Reaction networks
   decaying exponentials, 335
   parameters, 267
   whole-cell models, 253–264
Bipartiteness, 23, 42
Bisection, 76–78, 338
Black-box, 75, 84, 237, 336
Block diagonals, 333
Boltzmann distribution, 7, 142–144, 146
Boltzmann’s constant, 7, 142
Boundary conditions
Dirichlet, 230–231
for multiple dimensions, 234
Neumann, 231
and PDEs, 230–233
and solute diffusion, 230–233
Box-Muller method, 120
background, 127
Branch-and-bound methods, 49–52
Branching process, 199
Brownian motion, 167, 241–249, 263
Brownian noise, 157
Brute force, 47, 50, 53
Calcium, 260
Calendar queue, 205t, 209
Canonical path, 161–166, 169
background, 171
Capillary sequencing, 61
Catalysts. See Enzymes
CellML, 264
Cells
and biochemical networks, 253–264
cycle synchronization, 323–325
Cell simulation
and CTMM, 256–259
electrophysiological components, 264
hybrid models, 259, 263
and PDEs, 253–256, 263
protein expression, 268
standards and software, 263
trends, 262
as very large reaction network, 260–262
Centered difference, 84–85, 229–232, 235
Chain rule, 220
Channel protein, 201–203
Chapman-Kolmogorov equations, 133
Chebyshev polynomials, 340
Chemical reaction. See also Reaction networks
and interpolation, 336
and law of mass action, 211
with noise, 246–248
and stability, 215–217
Chemical solutions. See Solutions
Chromatic number, 40
Chromosomes
diploid, 198
haploid, 192
haplotypes, 280–286
tagging SNP selection, 44, 47
Chromosome walking, 63
cis isomer, proline, 180–182
Cliques, 39, 342
union of, 344–345
Clone-by-clone strategy, 63
Clustering, 342–347, 351
additional sources, 353
Coalescent
background, 200
coaalescent simulation, 195
definition, 193f, 194
and migration, 198
and recombinations, 198
separate populations, 197
variable population sizes, 196
Coexpression models, 342–347, 351
Collisions, 141, 206–209
Coloring
in automated sequencing, 61
in graph problems, 39, 49–50
Compartments, 253–256
Complexity, computational, 55, 260–262, 361. See also Intractability; NP-completeness
Computer graphics, 333
Concave functions, 108
Conditional probability, 295
Condition number, of matrix, 319
Conductance method, 166
background, 171
bounded random walk, 167–170
Conjugate gradient, 91, 92, 318, 319
Consensus sequence, 32
Constraint satisfaction
linear program, 96–108
nonlinear program, 108–110
parameter-tuning, 269–271
primal-dual methods, 107
Contact potentials, 5, 267
Continuous distributions
and importance/umbrella sampling, 154–156
joint distributions, 119–121, 151–152
rejection method, 121–124
transformation method, 116–121, 124f
Continuous optimization. See also Newton-Raphson method
bisection, 76–78, 338
description, 75
local versus global optima, 76
multivariate functions, 85–88
secant method, 78–80
Continuous systems
applications, 211–213
backward Euler, 217–219
definition, 211
differential equations, 212
with discrete event tracking, 206–209, 263
from discrete points, 323–326 (see also Extrapolation; Interpolation)
fine difference, 213, 226
forward Euler (see Forward Euler)
leapfrog, 221–223, 225, 236
single-step methods, 219–221, 223–225
Continuous time Markov models (CTMMs)
additional reading, 183
branching process, 199
cell simulation, 256–260
channel protein example, 201–203
and coalescence, 195
description, 173–178
versus discrete event models, 201–204
and DNA base evolution, 187
Kolmogorov equations, 178–182
Moleculizer program, 261
and population dynamics, 212
and protein folding, 180–182
rate inference, 273
and self-transition, 181
waiting time, 173–175
Convection, 237–239
Convection-diffusion, 238
Convergence, 338
order of, 248–249
Convex functions, 108–110
Cooling schedule, 148
COPASI, 260, 264
Correlation coefficients, 343, 356–358
Cross-validation, 361–362
CTMM. See Continuous time Markov models
Cubic formula polynomials, 76
Cubic formulas, 76, 329, 333
Curve, receiver operating characteristic (ROC), 360
Curve families, 334–337
Curve generation, 333
Curve linearization, 81, 86, 89, 91
Cut problems
k-cut, 38, 54, 344
maximum cut, 37, 344
minimum cut, 21–23

Data. See also Noisy data
ambiguity loss, 30
and Bayesian model, 347
and continuous optimization, 75
fitting, 329–336, 340, 361
gene expression microarray, 341
gene network inference, 352
for HMMs, 299–302
input and output format, 2–3
for intraspecies phylogeny, 29–30
posting time, 205
set relationships, 357
Decision problems, 36
Density
joint, 119
probability, 116–118, 121–122, 154
detailed balance, 143–145, 164, 169
Diagnosics, 358, 360

Differential equations. See Finite difference;
Ordinary differential equations; Partial
differential equations; Stochastic differential
equations
Diffusion
and boundaries, 230–233
and cell simulation, 259
convection-diffusion equation, 238
of particles, in two dimensions, 325
PDE example, 227
reaction-diffusion equations, 234–237, 325
Diffusion term, 234
Dijkstra’s algorithm, 20, 21
background, 33
Diploid organisms, 198
Dirichlet boundary, 230–231
Discrete distributions. See also Transformation method
and continuous models, 323–326
and Metropolis method, 146
rejection method, 124–126
and transformation method, 124
Discrete event models
artificial event, 208
background, 210
and cell simulation, 260
channel protein case, 201–203
and continuous systems, 206–209, 263, 325
versus CTMMs, 201–204
description, 203
efficiency, 204–206, 208–210
event loop, 204, 207
molecular collisions, 206–209
queuing, 205, 209–210
without queue, 208
Discretization
conversions (multigrid), 325
and gene coexpression, 344
of space, 229, 233, 235, 255, 258
of time, 242
Disease, diagnosis of, 358, 360
Distributions. See also Continuous distributions;
Discrete distributions
Boltzmann, 142–144, 146
exponential, 118
gamma, 268
Gaussian, 348
joint, 119–121, 149–152
modified, 156
normal, 120, 123–124
Poisson, 191
prior, 153, 349
probability, 347–349
stationary, 134–138, 149, 153–155, 159, 161
uniform, 115–116
DNA. See also String and sequence problems
diploid and haploid, 198
exact set matching, 27
intraspecies phylogeny, 28–33
DNA (cont.)
random strings, 129–133
repetitive, 63
simulation, 191–195
tagging SNP selection, 44, 47
DNA bases
and CTMMs, 187
evolution, 185–191, 269–271
frequency analysis, 275–277, 280–286
and HMMs, 291–293, 303–307
parameter-tuning, 269–271
DNA microarrays, 64–66, 71, 341
DNA sequencing
big sequences, 61–63
computational methods, 64–72
Eulerian path, 66, 73
hybridization method, 64–66, 71, 73
Maxam-Gilbert, 57–59, 61
n nanopore method, 74
overview, 73–74
Sanger dyeoxy method, 59–61
shotgun methods, 67–69, 73
single molecule, 72, 74
Domain recognition, 294, 297
Double-barrel shotgun, 69
background, 73
Drosophila melanogaster, 74
Duals, 39, 46, 107
Dynafit program, 336
E-Cell system, 260, 264
Edges, graph
and Bayesian model, 349
and bipartiteness, 42
cliques, 39, 342, 344–345
and CTMMs, 174
and gene network, 349–351
in hierarchical clusters, 345
in intraspecies phylogeny, 29–31, 41
in Markov model, 143
and maximum flow, 21
and mixing, 161–166, 170
negative weights, 20
in network structure, 349
in Steiner trees, 41
transition probabilities, 160
in vertex cover, 38, 45, 47, 53
Edit distance, 3–4
Edmonds-Karp algorithm, 22–23, 33
Eigenvalues
definition, 136
of Markov models, 136–139, 159, 186
and matrices, 318, 321, 322
Eigenvectors, 136–139, 186
Einstein, A., 364
Ellipsoid method, 104, 110
Embedded methods, 224
Energy. See also Force field
and amino acids, 5–7
and Metropolis method, 143, 147
potential, 142
and simulated annealing, 52, 148
and umbrella sampling, 157
Entropy, 343–344, 358
Enzymatic reactions, 253–256, 324f
Enzymes
concentration, 325
and ODEs, 212
protease, 8–11
Expectation maximization, 345–347
Equilibrium
and Boltzmann distribution, 142
in chemical diffusion, 325
Hardy-Weinberg, 282
Ergodicity
and canonical path, 164
definition, 136
and Markov models, 136, 148, 159, 164, 167, 169
and Metropolis method, 143
Errors. See also Accuracy
in differential equation types, 248
and expectation maximization, 286–287
and extrapolation, 337–339
false positives/negatives, 359, 360
forward and backward, 77, 90
and intraspecies phylogeny, 30
in leapfrog method, 222
Newton-Raphson algorithm, 83–85
in noisy data, 287
and physical conservation laws, 226
and sensitivity analysis, 363
and steepest descent, 89–90
and step size, 223
Euclidian distance, 343, 345
Euclidian traveling salesman, 48–49, 54
Eukaryotic genomes
assembly, 69–70, 73
DNA sequencing, 63, 67, 73
gene prediction, 307
sequence problems, 26
Eulerian path, 66, 73. Euler-Maruyama method, 246, 249, 250
Event loop, 204, 207
Evolution. See also Continuous time Markov models; Molecular evolution
coalescent model, 193–199
and data ambiguity, 30
description, 2–4
DNA base evolution, 185–191, 269–271
DNA strand simulation, 191
genetic algorithms, 52–53
graph problems, 16–18
intraspecies phylogeny, 28–33, 41
Jukes-Cantor model, 185–188
Graph problems (cont.)
  independent set, 38, 42
  matching, 23
  maximum clique, 39
  maximum cut, 37, 344
  maximum flow/minimum cut, 21–23
  minimum spanning trees, 16–18, 20, 29–31
  multigraphs, 16
  NP-completeness, 4, 36–42, 47, 344
  phylogeny example, 28–33
  and set problems, 44
  shortest path, 19–21
  Steiner trees, 40–41
  subgraphs, 42, 54
  traveling salesman, 36, 48, 54
  and union-of-cliques, 344
  vertex cover, 38, 45, 47, 53, 54
Graph properties, 42
Green’s function reaction dynamics (GFRD), 263
Grid box, 334
Growth factor, 223
Guilt by association method, 344

Haemophilus influenzae, 73
Hamiltonian path, 36–37, 65
Hamming distance, 41
Haploidy, 198
Haplotypes
  frequency estimation, 280–286
  inference from noisy data, 286–289
Hard sphere model, 206–209
Hardy-Weinberg equilibrium, 282
Hastings-Metropolis method, 160. See also
  Metropolis method
  Heat equation, 227
  background, 239
  Hessian, 86–89, 109
Heuristic methods. See also Simulated annealing
  background, 53, 158
  clustering methods, 344–347
  definition, 52
  and gene (co)expression, 344–347
  genetic algorithms, 52
  and Gibbs sampling, 152–154
  and intractability, 52
  kitchen sink approach, 53
  and Metropolis model, 52, 147
  and network inference, 349
Hexamers, 260–262
Hidden Markov models (HMMs)
  and amino acids, 297–299
  background and sources, 289, 307
  and DNA bases, 291–293, 303–307
  and expectation maximization, 300–307
  gene structure, 292, 299–302
  motif-finding, 303–307, 359–362
  and Newton-Raphson method, 302
  and output probability, 297–299
and protein domain, 294
and protein folding, 308
special features, 291
state assignment, 295–297
training, 299–302
transcription factor binding, 293
Hierarchical clustering, 345
HIV, 10
HMM. See Hidden Markov models
Huen’s method, 224
Hungarian method, 24, 33
Hybridization, sequencing by, 64–66, 71
  background, 73
Hydrogen bonds, 158
Hyperplanes, 97
Identity matrix, 310–312, 320
Image analysis, 325, 340
Imino acid, 180
Implicitly specified functions, 271–273
Importance sampling, 154–156, 170
  umbrella sampling, 155, 158
Infeasible points, 97
Infinite series, 337–340
Infinite sites model, 191–192
Information, mutual, 344
Information theory, 343, 358
Inheritable properties, 42
Integer linear programs, 51
Intermediate point methods, 104–107, 108
Interpolation
  best-fit, 336
  bilinear, 334
  in biochemical reactions, 335
  curve families, 334–337
  definition, 325
  examples, 323–326
  Fourier interpolants, 340
  Levenberg-Marquardt method, 336
  linear, 272
  multidimensional, 334
  and Newton-Raphson method, 81, 336
  and optimization, 335–337
  polynomial type, 326–330
  rational function, 330
  and secant method, 79
  splines, 331–334
  and steepest descent, 90
Intractability, See also NP-completeness
  approximation algorithms, 47–49, 50, 55
  branch-and-bound methods, 49–52
  brute force approach, 47, 53
  coping with, 30–32, 35, 46, 49, 53
  definition, 24–26, 35
heuristic approaches, 52
trade-offs, 30–32, 46, 49
Isomerization, 180–182
Iterative methods
finite difference, 338
Gauss-Seidel method, 317
Jacobi method, 317
Krylov subspace, 317–320
and Newton-Raphson, 82, 88
Itô integral, 244. See also Stochastic integrals;
Stochastic differential equations
Itô-Taylor series, 249

Jacobian, 86–89, 92
Jacobi method, 317
Johnson’s algorithm, 21
background, 33
Joint distributions, 119–121, 149–152
Joint entropy, 344
Jukes-Cantor model, 185–189, 191
background, 200
Karmarkar’s method, 104, 108, 110
k-coloring, 40
k-cut problems, 38, 54, 344
k-fold cross validation, 361
Kimura model, 188–191
background, 200
Kinetic models, 351–353
Kolmogorov criterion, 160, 164, 168
Kolmogorov equations
Chapman-Kolmogorov, 133
and CTMMs, 178–182
and discrete event simulation, 201
and evolutionary processes, 187, 190
and implicitly specified functions, 273
Kruskal’s algorithm, 17, 31
background, 33
Krylov subspace, 91, 317–320, 333
kth-order Markov model, 130–131

Laplacian, 227
Latent variables, 277, 284, 288–289, 300, 345
Lattice models
background, 10
description, 5–7
and discretized states, 324f, 325
and heuristics, 52
in Markov example, 145
move sets, 10
parameters, 267
and protein folding, 5–6, 145–146
for spatial discretization of PDEs, 258–259
Law of mass action, 211
Lazy queuing, 205
Leapfrog method, 221–223, 225, 236
Least-squares, 320, 336, 349, 356
Leave-one-out cross validation, 361
Levenberg-Marquardt method, 90, 273, 336
background, 93
Likelihood, maximum. See Maximum likelihood
Linear congruential generators, 116
Linear interpolation, 272
Linearization, of curve, 81, 86, 89, 91
Linear programming
barrier methods, 104
cost factors, 108
definition, 96
ellipsoid method, 104, 110
primal and dual, 107
relaxation, 51
simplex method, 97–103, 108, 110
software, 107, 111
standard form, 98–99
Linear recurrence, 222
Linear regression, 310
Linear systems
definition, 309
and differential equations, 213
Gaussian elimination, 310–316, 318
and gene networks, 352
and interpolation, 330–334
iterative methods, 316–321
Krylov subspace methods, 317–319
linear regression, 310
and multivariate functions, 87
optimization in, 92
over- and under determined, 320
pivoting, 312–316
preconditioners, 319–320
pseudo-inverse, 321
references, 93
and Taylor expansions, 85
Line-by-line method, 256
Local linearizing, 81, 86, 89, 91
Local optimum, 52
LU decomposition, 315

Macromolecular complexes, 260–262, 264
Markov chain Monte Carlo (MCMC), 141–158, 350
Markov chains
background, 139
definition, 129
and gene network, 350
irreducibility, 136
and mixing times, 163, 166–170
in molecular evolution, 185–188
Markov models
background, 139
branching process, 199
components, 129, 291
conductance, 166–170
continuous time (see Continuous time Markov models)
and DNA bases, 185–188, 269–271, 291–293
Markov models (cont.)
and DNA motifs, 153
eigenvectors, 136–139, 186
ergodicity, 136, 148, 159, 164, 169
gene sequence types, 276
and Gibbs sampling, 149–156
hidden, 291 (see also Hidden Markov models)
and Metropolis method, 142–148 (see also Metropolis method)
mixing time, 138, 159–160, 166, 170
and molecular evolution, 185–191
nonergodic, 137
order, 130–131
and prior distribution, 153
with random walk, 167
and spatial effects, 258
stationary distribution, 134–138, 149, 153–155, 159, 161
and waiting time (see Continuous time Markov models)
Mass action, law of, 211
Matching problems
exact set, 27
unweighted, 23
weighted, 24
Mating, 53
Matrices. See also Transition matrix
condition number, 319
inversion, 87
over/underdetermined, 310, 320, 330, 333
permutations, 314
positive (semi)definite, 92, 318, 319
Maxam-Gilbert method, 57–59, 61
Maximal matching, 47
Maximum a posteriori probability (MAP), 275
Maximum clique problems, 39
Maximum cut problems, 37, 344
Maximum edge loading, 161–166, 170
Maximum flow problems, 21
Maximum likelihood
background, 289
and clustering, 345–346
description, 268
and expectation maximization, 275, 277–280 (see Expectation maximization)
in haplotype error correction, 286–287
in haplotype frequency estimation, 282–283
and Hardy-Weinberg equilibrium, 282
and latent variables, 284
and network inference, 347–351
and parameter-tuning, 8–10, 268, 275–277, 283
MCell, 258–259, 264
Metropolis criterion, 6–7, 10
Metropolis method
background, 158
caveats on use, 146
efficiency, 154–156
generalized, 146–147
and mixing time, 146, 154, 170
for optimization, 147, 350
and protein folding, 142, 145, 154
and simulated annealing, 52, 148
and thermodynamics, 141–143, 146
and traveling salesman, 147
Michaelis-Menten reaction, 253–256
Microarrays, 64, 71, 341
Microreversibility, 143–145, 164, 169. See also Detailed balance
Midpoint method, 219–222
Migration, 198
Milstein’s method, 249, 251
Minimum cut, 21–23
Minimum description length (MDL), 361
Minimum set cover, 45. See also Vertex cover
Minimum spanning network, 31
Minimum spanning tree, 16–18, 20, 29–31
Minimum test set, 44
Mixing time
canonical path method, 161–166, 169, 171
conductance method, 166–170, 171
definition, 138, 159–160
and eigenvalues, 138–139
and importance sampling, 170
and Metropolis method, 146, 154
monomer-dimer systems, 171
Model space, reduction, 351
Model validation
accuracy, 358 (see also Accuracy)
cross-validation, 362
goodness measures, 355–358
overfitting avoidance, 361
receiver operating characteristic (ROC) curve, 360
scientific method, 363–366
sensitivity, 359, 360, 362
specificity, 359–361
Mode-of-action by network identification (MNI), 351
Modified distribution, 156
Molecular evolution
coalescent model, 192–198
DNA strand, 191
Jukes-Cantor model, 185–188
Kimura model, 188–191
and Kolmogorov equations, 187, 190
one-parameter, 185–188
and self-transition, 163–164
two-parameter, 185–188
Molecular modeling
and continuous optimization, 75
lattice models, 5–7, 145–146
macromolecular complexes, 260–262, 264
and numerical integration, 211
and stochastic differential equations, 245
and umbrella sampling, 156–158
Moleculizer program, 261
Monomer-dimer systems, 170–171
Monte Carlo samplers, 350
Motifs
  alignment of, 152
  detection of, 152, 303–307, 347, 359–362
  transcription factor binding, 293
Move sets, for lattice models, 6
Multicommodity flows, 23
Multidimensional curve, 336–337
Multigraphs, 16
Multigrid methods, 324f, 325
Multiple independent variables, 356 (see also Partial differential equations)
Multiple regression, 351
Multivariate functions, 85–88
Mutations
  in genetic algorithm, 53
  infinite sites model, 191–192
  and Jukes-Cantor model, 187, 191
  and Kimura model, 186f, 188–189
  random, 163–166
  simulation, 4–7, 191
  transitions/transversions, 189
  and Wright-Fisher neutral model, 192
Mutual information, 344
Needleman-Wunsch algorithm, 33
Network identification by multiple regression (NIR), 351
Networks
  gene regulatory, 341–353
  inference of, 349–353, 363
  minimum spanning, 31
  reaction networks, 260–264, 323–325, 340
  reduced median, 33
  Neumann boundary condition, 231–232
  Neville’s algorithm, 326–329
Newton-Raphson method
  background, 93
  black-box functions, 84
  and HMMs, 302
  and implicitly specified function, 273
  and interpolation, 336
  and Levenberg-Marquardt method, 90
  multidimensional, 85–88
  and parameter-tuning, 80–84, 269
  and steepest descent, 90
  Newton’s second law, 211
Next reaction method, 257
Noisy data, 286–289, 323–325, 329, 347
Nonlinear programming, 108–110
Nonlinear systems, 91–92
Nontrivial graph properties, 42
Normal distributions, 120, 123–124
NP-completeness
  background, 53–55
  coping with, 35, 46–53
  and DNA sequencing, 65
  linear programming relaxation, 51
  in Steiner tree, 41
  and union-of-cliques graph, 344
NP-hardness. See NP-completeness
Numerical integration. See also Partial differential equations; Stochastic differential equations
  additional readings, 225
  backward Euler, 217–219
  and black box functions, 336
  definition, 213
  and extrapolation, 337
  finite difference method, defined, 213
  forward Euler, 214–217, 222, 223, 224
  implicit, 316
  and interpolation, 336
  and Kolmogorov equations, 273
  leapfrog method, 221–223, 225, 236
  line-by-line method, 256
  midpoint method, 219–221
  multistep methods, 221–223
  and parameter-tuning, 272
  single-step methods, 214–221
  spectral methods, 226
  speed and efficiency, 223–225
  step size selection, 223–225, 233–234
  and transformation method, 118
Objective function, 96, 268–271, 336
ODEs. See Ordinary differential equations
Optimization. See also Continuous optimization;
  Gibbs sampling; Metropolis method; Parameter-tuning
  background, 92
  in bootstrapping, 350, 363
  conjugate gradient, 91, 318
  constrained (see Constraint satisfaction)
  and decision problems, 36
  description, 1–4
  discrete, 15
  and gene networks, 349
  and Gibbs sampling, 152–154, 350
  and interpolation, 335–337
  lattice models, 5–7
  Levenberg-Marquardt method, 90, 93
  and Metropolis method, 147–148, 350
  and model goodness, 356
  (non)linear systems, 91–92, 318
  and parameter-tuning, 8, 267–271
  of state assignments, in HMM, 295–297
  steepest descent, 89–90
  without zero-finding, 89–92
Order of convergence, 248
Ordinary differential equations (ODEs)
  backward Euler, 217–219
  and curve fitting, 335
  and errors, 248
  examples, 211–213
  forward Euler, 214–217, 222, 223, 224
Ordinary differential equations (cont.)
and gene networks, 351–353
leapfrog method, 221–223, 225, 236
line-by-line method, 256
living cell simulation, 253–256
midpoint method, 219–221
and reaction network, 323, 352
step size selection, 223–225
Overdetermined systems, 310, 320, 330
Overfitting, 361
Parameter selection, 267, 362
Parameter-tuning. See also Expectation
maximization; Hidden Markov models;
Optimization
and biochemical reactions, 267
description, 8–10, 267, 275
DNA base evolution, 269–271
and gene sequences, 276
haplotype frequency, 280–286
haplotype inference, 286–289
implicitly specified functions, 271–273
and linear systems, 309 (see also Linear systems)
maximum likelihood, 8–10, 268, 275–277, 283
motif-finding, 303–307
and Newton-Raphson method, 80–84, 269
and noisy data, 286–289
protease example, 8–10
and protein expression, 268
protein folding example, 267
and sensitivity, 363
Parsimony, 3, 29–33
background, 10
Partial differential equations (PDEs). See also
Reaction-diffusion equations.
additional information, 239
boundary conditions, 230–233
convection, 237–239
coupled one-dimension, 228–230
diffusion example, 227
initial conditions, 230
line-by-line method, 256
cell simulation, 253–256, 263
multiple spatial dimensions, 233–234
one spatial dimension, 228–230
step size, 233
Particle collisions, 141, 206–209
Particle diffusion, 325
Particle interactions, 177
PDEs. See Partial differential equations
Pearson correlation coefficient, 343
Permutation matrix, 314
Pfam protein database, 295
Philosophy of science, 363–366
Phylogeny, intraspecies, 28–33
Pivoting, 312–316
Poisson process, 191
Poisson random variable, 191
Polymerization, 61
Polynomial reduction, 46
Polynomials
Chebyshev, 340
cubic formula, 76, 329, 333
fitting to lower order, 329–331, 340
Neville’s algorithm, 326–329
quadratic formula, 76
quartic formula, 76
splines, 331–334
Polytope, 97
Popper, Karl, 364
Population dynamics, 29, 212
Population genetics, 280–286
Posting time, 205
Prediction
cut site, in proteases, 8–11
gene expression, 307, 309
protein expression, 323
Predictor-corrector schemes, 225
Primals and duals, 107
Prim’s algorithm, 18, 20
background, 33
Prior distribution. See Prior probability
Prior estimate, 303
Priority queue, 18, 205, 209–210
Prior probability, 153, 298, 349
Probability. See also Sampling
of best-fit, 275
conditional, and transitioning, 295
distribution, 347–349
fundamental transformation law, 117
maximum a posteriori (MAP), 275
maximum likelihood, 8–10, 268, 275–277, 283, 356
of migration, 198
prior, 298
Proline, 180–182
Proteases
cut site prediction, 8–11
and HIV, 10
and parameter-tuning, 8–10
Proteasomes, 11
Protein expression, 268, 323
Protein folding
and CTMMs, 180–182
and HMMs, 308
importance sampling, 154–156
lattice models, 5–7, 10 (see also Lattice models)
Markov model example, 145
Metropolis model, 142, 145, 154
parameters, 267
umbrella sampling, 155–158
Proteins
and Brownian motion, 157
channel protein, 201–203
coiled-coil, 293–295
complexes, 177, 260–262
database, 295
domain recognition, 294, 297
exact set matching, 27
growth rate example, 95
hydrogen bonds, 158
ligand binding, 75
longest common subsequence, 25, 42–43
longest common substring, 26
sampling programs, 261
string and sequence problems, 24–27
structure simulation, 4–7
translation, 268
Pseudoinverse, 321
Pseudorandom numbers, 115
P-value calculators, 343
Quadratic formula, 76
Quadratic programming, 109
Quartic polynomials, 76
Queues, 18, 205, 207–210. See also Priority queues
Random DNA strings, 129–133
Random mutations, 163–166
Random number generation
  pseudorandom numbers, 115
  rejection method, 121–124
  transformation method, 115–121
Random variables. See Distributions
Random walk, 167–170, 324f
Rational function, 330
Rational interpolation, 330
Reaction-diffusion equations, 234–237, 325
  background, 239
Reaction networks, 211, 217, 260–264, 264, 271, 323–325, 335, 340
  cell simulation, 260–262
  data-fitting, 340
  Reaction term, 234
Receptor operating characteristic (ROC) curve, 360
Recombination, 198
Reduced median network, 33
Rejection method, 121–126
  background, 127
Relaxation, 51
Reversibility, 143–145
Reweighting, 21
Richardson extrapolation, 225, 337
RNAi, 352
Runge-Kutta methods
  and accuracy, 221, 225, 237
  with black box, 237
  and cell simulation, 260
  embedded, 225
  fourth order, 221
  midpoint method, 219–221
  and stability, 221
Run time. See also Optimization; Simulation
  and accuracy, 219, 233
  and boundary conditions, 231
  coalescent, 195–197
  and CTMMs, 173–175, 273
  and discrete event models, 204–206, 208–210
  and importance sampling, 155
  and intraspecies phylogeny, 29
  and Krylov subspace methods, 319
  and Metropolis method, 146, 154
  and numerical integration, 225
  and stability, 215
  and step size selection, 217, 233
  and umbrella sampling, 156–158
Sampling. See also Gibbs sampling; Importance sampling; Markov models; Metropolis method; Umbrella sampling
  continuous distributions, 116–124, 156
  discrete distributions, 124–126, 146
  efficiency, 154
  exponential random variable, 118–119
  geometric random variable, 125–126
  joint distributions, 119–121, 149–152
  modified distribution, 156
  and network inference, 350, 363
  normal distributions, 120
  with optimization, 350
  at point in time, 182
  (pseudo)random numbers, 115
  rejection method, 121–124
  and simulation, 7, 115
  transformation method, 116–121
  uniform random variable, 116
Sanger dideoxy method, 59–61
Scaled variables, 105
Science, philosophy of, 363–366
Scientific method, 363–366
Secant method, 78–80
Selfing, 198
Self-transitions
  conversion to, 168
  and CTMMs, 181
  and mixing time bounds, 159
  and molecular evolution, 163–164
Semidefinite programming, 108–110
Sensitivity, 359, 360, 362
Sequences. See DNA sequencing; String and sequence problems
Set problems
  independent set, 38, 42, 46, 54
  minimum set cover, 45
  minimum test set, 44
  Shortest common supersequence, 43
  Shortest common superstring, 44
  Shortest path, 19–21
Shotgun methods, 67–71
  background, 73
Signal processing, 340
Similarity measures, 342–344
Simplex method, 97–103, 108, 110
Simulated annealing
background, 54
and Bayesian models, 349–350
description, 52
and Metropolis method, 52, 148
Simulation
Brownian motion, 241–249
chemical, in inhomogeneous solution, 234–237
continuous systems, 211–213 (see also Continuous systems)
of CTMM (pseudocode), 175f
of discrete events (see Discrete event models)
DNA, haploid, 198
DNA random string, 129–133
DNA strand, 191
DNA whole population, 192–195
implicit functions, 271–273
of macromolecular reactions, 260–262
of mutation, 4–7, 191
parameter-tuning, 267–271
of particle collisions, 141, 206–209
protein structure example, 4–7
reaction networks, 253–264
of recombination, 198
and sampling, 7, 115
Single-molecule sequencing, 72, 74
Single-pair shortest path, 19–21
Single-step methods, 219–221, 223–225
Smith-Waterman algorithm, 33
SNP selection, 44, 47
Social constructivism, 365
Solutions
convection, 237–239
diffusion, 227, 230–237, 259, 325
inhomogeneous, 234
Sparse candidate algorithm, 351, 352
Sparse graphs, 18, 21
Sparse matrices, 315, 316, 322
Spatial models
discretization, 229, 233, 235, 255, 258
multidimensional, 85–89, 233, 325
one dimension, 228–230
reaction-diffusion equations, 234–236
three-dimensional, 234
and time, 233
two-dimensional, 325
Spearman correlation coefficient, 343
Species tree, 28–33
Specificity, 359–361
Spectral methods. See also Eigenvalues, Fourier transforms
interpolation, 340
numerical integration, 226, 239
Splines, 331–334
Stability
and accuracy, 221, 223, 251
of Adams-Bashforth methods, 223
additional information, 239
of backward Euler, 218
classifications, 215
disadvantages, 217
of forward Euler, 215–216
of leapfrog method, 222
and mutations, 4–7
references, 239
and Runge-Kutta methods, 221
and step size, 217
and stochastic differential equations, 249–251
unconditional, 219
von Neumann analysis, 215–217
Standards, 264
Standard Weiner process, 241
Stationary distribution, 134–138, 149, 153–155, 159, 161
Steeped descent, 89
Steiner nodes, 32, 41
Steiner trees, 31–32, 40–41
Step sizes, 233, 337
adaptive methods, 223–225
predictor-corrector schemes, 225
and stability, 217
Stochastic differential equations
accuracy, 248
additional information, 252
for Brownian motion, 241–248
and cell simulation, 256
Euler-Maruyama method, 246, 249, 250
and implicit function, 273
for protein-folding, 157
stability, 249–251
Stochastic integrals, 244
Stochastic simulation algorithm (SSA), 256–260, 263
StochSim, 256–259
Stratonovich integral, 244
String and sequence problems
applications, 24
exact set matching, 27
haplotype frequency, 280–286
haplotype inference, 286–289
HMM, 292
hybridization, 64–66, 71, 73
longest common subsequence, 25, 42–43
longest common substring, 26
Markov model example, 276
noisy data, 286–289
NP completeness, 42–44, 47
random DNA strings, 129–133
sequence alignment, 33
shortest common supersequence, 43
shortest common superstring, 44
suffix trees, 26, 27, 33
Index

Subgraphs, 42, 54
Subsequences, 25, 42–43
Subspace. See Krylov subspace
Substrings, 26
Successive squaring, 133
Suffix trees, 26, 27, 33
Sum-of-squares. See Least-squares
Supersequences, 43
Superstrings, 44
Systems Biology Markup Language (SBML), 264

Tagging SNP selection, 44, 47
Tau leap algorithm, 259
Taylor series
  approximation with, 80–82, 85, 232
  and backward Euler, 218
  and finite difference approximations, 229, 232
  and forward Euler, 215
  and midpoint method, 220
  and multistep methods, 222, 225
  and Newton-Raphson method, 80–82, 84–85
  and Richardson extrapolation, 337
  stochastic. See Itô-Taylor series
Temperature. See Simulated annealing
Terminal nodes. See Steiner trees
Terminator base, 59–61
Thermodynamics
  and CTMMs, 180–182
  and Metropolis method, 141–143, 146
Time. See Evolution; Mixing time; Run time
  Tractability, 24–26, 35. See also Intractability
  Transcription factor binding, 293
  Transformation method, 116–121, 124
    background, 127
  trans isomer, proline, 180–182
  Transition, Markov model, 130
  Transition matrix
    for CTMMs, 173
    in Jukes-Cantor model, 186
    in Kimura model, 189
    of Markov models, 132, 134–137
Traveling salesman problem (TSP), 36, 48, 54, 147
Trees
  minimum spanning, 16–18, 20, 29–31
  and optimization, 2–4
  Steiner, 31–32, 40–41
  suffix, 26, 27, 33
  and traveling salesman, 48
Triangle traveling salesman, 48–49, 54
True negatives, 359
True positives, 359, 360
Truth, 365
Twofold cross-validation, 361

Umbrella sampling
  background, 158
  and Gibbs sampler, 156–158
  and Metropolis sampler, 155

Unconditional stability, 219
Underdetermined system, 310, 321, 333
Union-of-cliques, 344–347
Variation distance, 160
Vertex cover
  approximation algorithms, 47, 50–51
  description, 38
  and genetic algorithm, 53
  hardness testing, 46
  and independent set, 39
  and minimum set cover, 45
  reference, 54
Virtual Cell, 255, 261, 264
Viterbi algorithm, 296, 299
von Neumann analysis, 215, 219, 220, 250

Waiting time, 173–175
  and coalescence, 198
  and CTMMs, 201–204
  and Poisson process, 191
  and recombination, 199
Wave equation, 238
Wavelets, 223, 226, 340
Weiner process, 241
Whole population sampling, 192–195. See also Coalescent
Wikipedia, 93, 110
Wright-Fisher neutral model, 192

Zero, avoiding, 105–107
Zero-finding
  alternative approaches, 89–92
  bisection method, 76–78
  multivariate functions, 85–88
  Newton-Raphson methods, 80–88, 90, 269
  secant method, 78–80
  0–1 integer programming, 51