

INDEX

- A* algorithm, 114
- actions, 147
- actor-critic, 285
- affine function, 53
- aggregation, 226
 - modeling, 229
 - multiple levels, 233
- algorithm
 - ADP for asset acquisition, 389
 - ADP for infinite horizon, 304
 - ADP for policy iteration, 305
 - ADP using post-decision state, 105
 - ADP with exact expectation, 97
 - ADP with pre-decision state, 276
 - approximate expectation, 98
 - approximate hybrid value/policy iteration, 284
 - approximate policy iteration with VFA, 282
 - asynchronous dynamic programming, 114
 - backward dynamic programming, 54
 - bias-adjusted Kalman filter stepsize, 204
 - CUPPS algorithm, 373
 - double-pass ADP, 273
 - Gauss-Seidel variation, 58
 - generic ADP, 110
 - hybrid value/policy iteration, 63
 - infinite horizon generic ADP, 305
 - policy iteration, 62
 - Q-learning
 - finite horizon, 278
 - infinite-horizon, 310
 - real-time dynamic programming, 115
 - relative value iteration, 58
 - roll-out policy, 293
 - SHAPE algorithm, 362
 - shortest path, 18
 - single-pass ADP, 272
 - SPAR, 355
 - stochastic decomposition, 372
 - synchronous ADP, 291
 - synchronous dynamic programming, 114
 - temporal-difference learning for infinite horizon,
 - 309
 - tree-search, 292
 - value iteration, 57
- aliasing, 235
- American option, 239
- apparent convergence, 210
- asset acquisition, 28–29
 - ADP algorithm, 389
 - variations, 391
 - lagged, 30
- asset pricing, 26
- asynchronous dynamic programming, 114
- attribute transition function, 164
- backpropagation through time, 273
- backward dynamic programming, 54
- bandit problem, 37
- bandit problems, 332
- basis functions, 127, 237, 362
 - approximate linear programming, 311

- geometric view, 244
- Longstaff and Schwartz, 241
- neural network, 253
- recursive time-series, 251
- tic-tac-toe, 243
- batch process, 257
- batch replenishment, 31, 65
- Bellman's equation, 3, 28, 48
 - deterministic, 49
 - operator form, 53
 - standard form, 49
 - expectation form, 49
 - vector form, 51
- Bellman error, 98
- Bellman
 - functional equation, 4
 - Hamilton-Jacobi, 3
 - optimality equation, 4
 - recurrence equation, 4
- Benders' decomposition, 370
 - CUPPS algorithm, 373
 - stochastic decomposition, 372
- bias, 195
 - due to value iteration, 286
 - statistical error in max operator, 287
- blood management
 - ADP algorithm, 397
 - model, 393
- Boltzmann exploration, 328
- budgeting problem
 - continuous, 21
 - discrete, 19
- contribution function, 40, 166
- controls, 147
- cost function, 166
- CUPPS algorithm, 372
- curse of dimensionality, 92
 - action space, 5
 - outcome space, 5
 - state space, 5
- cutting planes, 365
- decision node, 23
- decision tree, 23
- decisions, 147
- double-pass algorithm, 273
- dynamic assignment problem, 34
- error measures, 315
- exogenous information, 29, 40, 151
 - lagged, 155
 - outcomes, 153
 - scenarios, 153
- experimental issues
 - convergence, 295
 - starting, 294
- exploitation, 327
- exploration, 326–327
- exploration vs. exploitation, 116, 323
- exponential smoothing, 99
- factored representation of a state, 146
- finite horizon
 - for infinite horizon models, 317
 - flat representation of a state, 146
- forward dynamic programming, 93
- gambling problem, 25
- Gittins exploration, 344
- Gittins indices, 332
 - basic theory, 334
 - foundations, 332
 - normally distributed rewards, 336
- gradients, 352
- greedy strategy, 95
- infinite horizon, 55, 304
 - finite-horizon approximations, 317
 - policy iteration, 305
 - Q-learning, 310
 - temporal-difference learning, 307
 - value iteration, 305
- information acquisition, 36
 - illustration, 330
- initialization, 112
- interval estimation, 337
- knowledge gradient, 339
- L-shaped decomposition, 372
- lagged information, 155
- lattice, 67
- learning rate, 181
- learning rate schedules, 183
- learning strategies
 - Boltzmann exploration, 328
 - epsilon-greedy exploration, 329
 - exploitation, 327
 - exploration, 326
 - Gittins exploration, 344
 - Gittins indices, 332
 - interval estimation, 337
 - knowledge gradient, 339
 - mixed, 327
 - upper confidence bound, 338
- leveling algorithm, 355
- linear filter, 99
- linear operator, 53
- linear programming method
 - approximate, 311
 - exact, 64
- linear regression, 238
 - Longstaff and Schwartz, 239
 - recursive estimation
 - derivation, 263
 - multiple observations, 250
 - time-series, 251
 - recursive least squares
 - nonstationary data, 249
 - stationary data, 248
 - recursive methods, 246
 - stochastic gradient algorithm, 247
- Longstaff and Schwartz, 239
- lookup-table, 99
- Markov decision processes, 47
- max operator, 53
- measure-theoretic view of information, 170
- min operator, 53
- model-free dynamic programming, 118

- model
 - contribution function, 166
 - decisions, 147
 - elements of a dynamic program
 - contribution function, 130
 - decision variable, 130
 - exogenous information, 130
 - objective function, 130
 - state, 130
 - transition function, 130
 - policies, 149
 - transition function, 159
- modeling dynamic programs, 40
- models
 - contribution function, 119
 - elements of a dynamic program, 130
 - state variable, 130
 - exogenous information, 119
 - resources, 135
 - multiple, 137
 - single discrete, 136
 - state, 139
 - time, 132
 - transition function, 118
- monotone policies, 64–65
 - proof of optimality, 81
- Monte-Carlo sampling, 100
- myopic policy, 150
- neural networks, 253
- nomadic trucker, 137
 - learning, 323
- objective function, 40, 48, 169
- on-line applications, 317
- optimality equation, 48
- optimality equations
 - post-decision state, 104
 - proof, 70
- outcome node, 23
- outcomes, 153
- partially observable states, 145
- policies, 149, 159
 - randomized, 151
- policy iteration, 62, 282
 - hybrid, 63
 - infinite horizon, 305
 - with look-up table representation, 282
 - with myopic rules, 283
 - with neural networks, 283
 - with regression function, 283
 - with value function approximation, 282
- post-decision state, 142
 - optimality equations, 104
 - perspective, 107
- Q -learning, 276
 - infinite horizon, 310
- randomized policies, 80, 151
- real-time dynamic programming, 114
- reinforcement learning, 119
- resource allocation
 - asset acquisition, 388
 - blood management, 392
 - fleet management, 416
 - general model, 404
 - portfolio optimization, 401
 - trucking application, 421
- resources
 - multiple, 137
 - nomadic trucker, 137
 - single discrete, 136
- reward function, 166
- RTDP, 114
- sample path, 95
- scenarios, 153
- SHAPE algorithm, 359
 - proof of convergence, 377
- Sherman-Morrison, 264
- shortest path
 - deterministic, 2, 18
 - information collecting, 39
 - stochastic, 24
- single-pass algorithm, 272
- smoothing factor, 181
- SPAR algorithm
 - projection operation, 382
- SPAR
 - projection, 357
 - weighted projection, 359
- state sampling
 - all states, 290
 - roll-out heuristic, 293
 - tree search, 291
- state variable
 - definition, 139
- state
 - alias, 235
 - definition, 40, 139
 - factored, 146
 - flat, 146
 - partially observable, 145
 - post-decision, 101, 103, 142
 - pre-decision, 103
 - sampling strategies, 114
 - asset acquisition I, 28
 - asset acquisition II, 30
 - asset pricing, 27
 - bandit problem, 38
 - budgeting problem, 20
 - dynamic assignment problem, 36
 - gambling problem, 25
 - shortest path, 19
 - state of knowledge, 38
 - transformer replacement, 33
- states of a system
 - hyperstate, 141
 - information state, 141
 - resource state, 141
 - single resource, 141
- stepsize, 99, 181
- stepsize rule, 183
- stepsize
 - apparent convergence, 210
 - bias-adjusted Kalman filter, 204

- bias and variance, 195
- convergence conditions, 184
- deterministic, 183
 - $1/n$, 187
 - constant, 186
 - harmonic, 187
 - McClain, 188
 - polynomial learning rate, 187
 - search-then-converge, 189
- infinite horizon, 313
 - bounds, 314
- optimal
 - nonstationary I, 200
 - nonstationary II, 201
 - stationary, 198,
- stochastic, 190
 - Belgacem's rule, 194
 - convergence conditions, 191
 - Gaivoronski's rule, 193
 - Godfrey's rule, 194
 - Kesten's rule, 192
 - Mirozahmedov's rule, 193
 - stochastic gradient adaptive stepsize, 193
 - Trigg, 194
- stochastic approximation procedure, 181
- stochastic approximation
 - Martingale proof, 215
 - older proof, 212
- stochastic decomposition, 372
- stochastic gradient algorithm, 181
- stochastic programming, 365
 - Benders, 370
- submodular, 67
- superadditive, 68
- supermodular, 67–68
- supervisor, 244
- supervisory learning, 244
- supervisory processes, 158
- synchronous dynamic programming, 114
- system model, 30
- taxonomy of ADP algorithms, 296
- temporal-difference learning, 279
 - infinite horizon, 307
- tic-tac-toe, 242
- time, 132
- transformer replacement, 32
- transition function, 3, 20, 30–31, 40, 159
 - attribute transition, 164
 - resource transition function, 162
 - special cases, 165
- batch, 31
- transition matrix, 47, 52
- two-stage stochastic program, 366
- uncertainty bonus, 344
- upper confidence bound sampling algorithm, 338
- value function approximation, 94, 107
 - aggregation, 226
 - batch process, 257
 - cutting planes, 365
 - error measures, 315
 - leveling, 355
 - mixed strategies, 252
 - neural networks, 253
 - recursive methods, 246
 - regression, 237
 - regression methods, 362
 - SPAR, 357
 - tic-tac-toe, 242
- value function approximations
 - gradients, 352
 - linear approximation, 353
 - piecewise linear, 355
 - SHAPE algorithm, 359
- value iteration, 57
 - bound, 60
 - error bound, 79
 - Gauss-Seidel variation, 58
 - infinite horizon, 305
 - monotonic behavior, 59
 - pre-decision state, 276
 - relative value iteration, 58
 - stopping rule, 57
 - proof of convergence, 74
 - proof of monotonicity, 77
- variance of estimates, 195