
Index

- Action space, 12
- Actor-critic methods
 - Compatible features, 380
 - Fisher information matrix, 384
 - Score function, 377
- Admissible input, 246
- Advantage function, 322, 382
 - For optimal policy, 337
- ARMA model, 12
- Asymptotic covariance, 281
- Asymptotic stability, 18
 - Exponential, 19
 - Global, 18
- Average-cost optimal control, 402
- Average-cost optimality equation, 245, 403
- Basis, 68
 - Linearly independent, 177, 325
 - Tabular, 163, 358
- Belief state, 410
- Bellman equation, 53
 - Fixed-policy, 56
- Bellman error, 63
- Central Limit Theorem (CLT), 222
 - Batch means method, 223
- Coercive, 20
- Compact, 395
- Comparison theorem, 21
- Compatible features, 380
- Condition number, 396
- Conditional expectation, 237, 323
- Continuously differentiable, 23
- Control variate, 230
- Controlled transition matrix, 245
- Convex, 91
 - Strict, 92
 - Strong, 92
- Cost to go, 52
 - Continuous time, 69
- Discount factor, 59
- Discounted cost, 402
- Disturbance rejection, 13
- Drift condition, 23
- Drift inequality, 19
- Dynamic programming, 54
- Dynamic programming (DP) equation, 53
- Eligibility vector, 163, 178, 327
- Empirical
 - Distribution, 162
 - Mean, 162
 - Pmf, 162
- Empirical risk, 161
- Equilibrium, 17
 - Region of attraction, 19
- Ergodic, 209, 213
 - Geometric, 235
- Euler approximation, 87
- Examples
 - Acrobot, 39
 - Bandits, 266
 - CartPole, 38
 - Frictionless pendulum, 28
 - Linear state space model, 208
 - M/M/1 queue, 211
 - M/M/1 value functions, 252, 322, 358
 - MagBall, 36
 - Mountain car, 34, 70, 118
 - Queueing game, 261
 - Queueing network, 73
 - Rover with partial information, 264
 - Rowing, 41
 - Speed scaling, 253
- Exercises
 - CartPole, 38
 - Criteria for instability, 239
 - Introduction to conditional expectation, 237
 - Inventory model, 275
 - MagBall, 36, 153
 - Pendulum, 47, 153
 - Perron–Frobenius theory, 241
 - Rover with full observations, 271
 - Rover with partial information, 277
 - Simulation theory and practice, 239

434

Exercises (Cont.)
 Speed scaling, 272
 Tabular basis, 358
 Zap zero, 358
 Exogenous, 11
 Expectation
 Conditional, 237, 323
 Experience replay buffer, 162
 Exploration, 4, 31, 266
 Extremum seeking control, 114

 Feedback law, 9
 Feedforward control, 9
 Fixed-point equation, 53
 Fluid model, 248
 Fundamental matrix, 58, 241

 Galerkin, 163, 178
 Gradient descent
 Acceleration, 155
 Stochastic, 114
 Gradient flow, 90
 Matrix gain, 150
 Graveyard state, 60, 401
 Grönwall inequality, 86

 Hamilton–Jacobi–Bellman equation, 69
 History state, 13
 HJB equation, 69
 Hurwitz, 27

 Indicator function, 395
 Inf-compact, 20
 Infimum, 396
 Information state, 410
 Input, 12
 Admissible, 246
 Input space U , 12
 Integral control, 14
 Internal Model Principle, 14
 Inverse dynamic programming (IDP), 63
 Irreducible, 233

 Kac’s theorem, 234

 L -smooth, 91
 Law of Large Numbers (LLN), 222
 Linear independence, 177, 325
 Linear quadratic regulator, 52
 Continuous time, 77
 Linear state space model, 15, 208
 Continuous time, 16
 Linearization, 29
 Lipschitz continuous, 85
 Load, 211
 LSTD Learning, 329
 LTI system, 12
 Gain matrix, 15

Index

Lyapunov equation, 26
 Continuous time, 27
 Lyapunov function, 19
 Control, 63, 81

 Markov chain
 ψ -irreducible, 233, 234
 x^* -irreducible, 234
 Aperiodic, 233
 Communication diagram, 214
 Ergodic, 213
 Memoryless property, 205
 Shift operator, 398
 Spectral gap, 214
 Transition kernel, 206
 Transition matrix, 206
 Uni-chain, 233
 Markov decision process, 245
 Markov property, 399
 Strong, 400
 Martingale difference, 307
 Mean-field dynamics, 248
 Model predictive control, 62
 Monte Carlo, 222
 Control variate, 230

 Neighborhood, 395
 Newton–Raphson algorithm, 284
 Newton–Raphson flow, 88
 Regularized, 89, 301
 Nonlinear state space model, 12
 Continuous time, 16
 Markov, 205

 Observations, 9
 Occupancy pmf ω , 194
 ODE method, 84, 281
 Oja’s algorithm, 151
 Optimal control
 Hamiltonian, 70
 Minimum principle, 70
 Risk sensitive, 276
 Optimality equation
 Average cost, 245, 403
 Optimization, 90
 Stationary point, 47
 Ordinary differential equation (ODE), 84
 ODE method, 84, 281
 Vector field, 23

 Poisson’s equation, 216
 Poisson’s inequality, 20, 218
 Continuous time, 24
 Policy, 9
 H -greedy, 321
 Linear optimal, 66
 Markov, 13, 247
 Stationary Markov, 247

- Policy improvement
 - Approximate, 104
 - Average cost, 404
 - Discounted cost, 320
- Polyak–Juditsky–Ruppert Averaging, 108, 282
- Positive definite, 396
- Principle of optimality, 53
 - Continuous time, 69
- Probability mass function (pmf), 206
- Probing signal, 98
- Q-function
 - Average cost, 246
 - Fixed policy, 320
 - Total cost, 54
- Q-learning
 - Asynchronous, 339
 - GQ, 183, 349
 - $Q(0)$, 338
 - Relative, 344
 - Synchronous, 339
 - Watkins, 339
- Quasistochastic approximation, 98
- Queue
 - CRW, 251
 - M/M/1, 211
 - MaxWeight policy, 75
- Random walk, 211
 - Reflected r.w., 211
- Reference signal, 9
- Relative value function, 216, 403
- Representer theorem, 166
- Reproducing kernel Hilbert space, 166
- Riccati equation, 66, 77
- Risk-sensitive optimal control, 276
- Sample complexity, 290
- SARSA, 172, 321
- Score function, 231, 377
- Shift operator, 398
- Shortest path problem (SPP), 60, 402
- Simplex, 410
- Simultaneous perturbation stochastic approximation, 157
- Singular value, 396
- Span seminorm, 395
- Spectral gap, 214
- Split sampling, 229, 333
- Stable
 - Asymptotically stable, 18
 - Exponentially asymptotically stable, 19
 - Globally asymptotically stable, 18
 - in the sense of Lyapunov, 18
 - Ultimately bounded, 109, 125
- State, 12, 205
- State feedback, 13, 247
 - linear, 66
- State space X , 12
- State space model, 12
- Stochastic approximation
 - Algorithm, 280
 - Averaging, 282
 - Projection, 286
 - Restart, 286
 - SNR, 302
 - Zap, 301
- Stochastic gradient descent, 231, 294
 - Quasi, 114
- Stochastic Newton–Raphson, 302
- Sublevel set, 19
- Successive approximation, 54
 - Picard iteration, 85
- Sufficient statistic, 12, 205
- Supremum, 396
- Tabular basis, 163, 358
- TD-learning, 172, 327, 332
 - Advantage, 337
 - LSTD, 329
 - Off-policy, 172, 332
 - On-policy, 172, 332
 - Regenerative, 370, 374
 - Regenerative for a.c., 374
 - Relative, 335
 - State weighting, 366
- Temporal difference, 30, 68, 162, 328
- Total cost, 17
- Tracking, 9, 13
 - Regulation, 15
- Transition matrix, 206
 - Controlled, 245
- Uni-chain, 233
- Value function, 18
 - Continuous time, 69
 - Discounted cost, 59, 402
 - Finite horizon, 61
 - Optimal, 51
- Value Iteration, 54, 403
 - Average cost, 403
 - Boundary condition, 61
- Vector field, 23
 - Newton–Raphson, 88
- Workload
 - Virtual station, 76
 - Virtual w. process, 76
- Zap
 - Q-learning, 351
 - QSA, 112
 - SA, 301
 - Zero, 301