CAMBRIDGE

Cambridge University Press
978-1-107-13460-7 — Partially Observed Markov Decision Processes
Vikram Krishnamurthy
Table of Contents
More Information

Contents

Preface *page* xi

1 Introduction 1
1.1 Part I: Stochastic models and Bayesian filtering 1
1.2 Part II: POMDPs: models and algorithms 2
1.3 Part III: POMDPs: structural results 4
1.4 Part IV: Stochastic approximation and reinforcement learning 5
1.5 Examples of controlled (active) sensing 6

Part I Stochastic models and Bayesian filtering 9

2 Stochastic state space models 11
2.1 Stochastic state space model 11
2.2 Optimal prediction: Chapman–Kolmogorov equation 15
2.3 Example 1: Linear Gaussian state space model 16
2.4 Example 2: Finite-state hidden Markov model (HMM) 19
2.5 Example 3: Jump Markov linear systems (JMLS) 22
2.6 Modeling moving and maneuvering targets 24
2.7 Geometric ergodicity of HMM predictor: Dobrushin's coefficient 26
2.8 Complements and sources 31

Appendix to Chapter 2 32
2.A Proof of Theorem 2.7.2 and Lemma 2.7.3 32

3 Optimal filtering 34
3.1 Optimal state estimation 35
3.2 Conditional expectation and Bregman loss functions 36
3.3 Optimal filtering, prediction and smoothing formulas 40
3.4 The Kalman filter 44
3.5 Hidden Markov model (HMM) filter 50
3.6 Examples: Markov modulated time series, out of sequence measurements and reciprocal processes 53
3.7 Geometric ergodicity of HMM filter 57

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-13460-7 — Partially Observed Markov Decision Processes
Vikram Krishnamurthy
Table of Contents
More Information

3.8 Suboptimal filters 59
3.9 Particle filter 61
3.10 Example: Jump Markov linear systems (JMLS) 67
3.11 Complements and sources 69

Appendices to Chapter 3 70
3.A Proof of Lemma 3.7.2 70
3.B Proof of Lemma 3.7.3 71
3.C Proof of Theorem 3.7.5 71

4 Algorithms for maximum likelihood parameter estimation 73
4.1 Maximum likelihood estimation criterion 73
4.2 MLE of partially observed models 75
4.3 Expectation Maximization (EM) algorithm 79
4.4 Forward-only filter-based EM algorithms 85
4.5 Method of moments estimator for HMMs 88
4.6 Complements and sources 89

5 Multi-agent sensing: social learning and data incest 93
5.1 Social sensing 93
5.2 Multi-agent social learning 96
5.3 Information cascades and constrained social learning 98
5.4 Data incest in online reputation systems 101
5.5 Fair online reputation system 104
5.6 Belief-based expectation polling 110
5.7 Complements and sources 114

Appendix to Chapter 5 115
5.A Proofs 115

Part II Partially observed Markov decision processes: models and applications 119

6 Fully observed Markov decision processes 121
6.1 Finite state finite horizon MDP 121
6.2 Bellman's stochastic dynamic programming algorithm 124
6.3 Continuous-state MDP 127
6.4 Infinite horizon discounted cost 128
6.5 Infinite horizon average cost 132
6.6 Average cost constrained Markov decision process 136
6.7 Inverse optimal control and revealed preferences 139
6.8 Complements and sources 142

Appendices to Chapter 6 143
6.A Proof of Theorem 6.2.2 143
6.B Proof of Theorems 6.5.3 and 6.5.4 144

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-13460-7 — Partially Observed Markov Decision Processes
Vikram Krishnamurthy
Table of Contents
More Information

7 Partially observed Markov decision processes (POMDPs) 147
7.1 Finite horizon POMDP 147
7.2 Belief state formulation and dynamic programming 150
7.3 Machine replacement POMDP: toy example 153
7.4 Finite dimensional controller for finite horizon POMDP 154
7.5 Algorithms for finite horizon POMDPs with finite observation space 157
7.6 Discounted infinite horizon POMDPs 163
7.7 Example: Optimal search for a Markovian moving target 169
7.8 Complements and sources 177

8 POMDPs in controlled sensing and sensor scheduling 179
8.1 Introduction 179
8.2 State and sensor control for state space models 180
8.3 Example 1: Linear Gaussian control and controlled radars 182
8.4 Example 2: POMDPs in controlled sensing 185
8.5 Example 3: Multi-agent controlled sensing with social learning 191
8.6 Risk-averse MDPs and POMDPs 195
8.7 Notes and sources 199

Appendix to Chapter 8 199
8.A Proof of theorems 199

Part III Partially observed Markov decision processes: structural results 201

9 Structural results for Markov decision processes 203
9.1 Submodularity and supermodularity 204
9.2 First-order stochastic dominance 207
9.3 Monotone optimal policies for MDPs 208
9.4 How does the optimal cost depend on the transition matrix? 210
9.5 Algorithms for monotone policies – exploiting sparsity 211
9.6 Example: Transmission scheduling over wireless channel 213
9.7 Complements and sources 216

Appendix to Chapter 9 216
9.A Proofs of theorems 216

10 Structural results for optimal filters 219
10.1 Monotone likelihood ratio (MLR) stochastic order 220
10.2 Total positivity and copositivity 223
10.3 Monotone properties of optimal filter 224
10.4 Illustrative example 226
10.5 Discussion and examples of Assumptions (F1)–(F4) 227

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-13460-7 — Partially Observed Markov Decision Processes
Vikram Krishnamurthy
Table of Contents
More Information

10.6 Example: Reduced complexity HMM filtering with stochastic dominance bounds 229
10.7 Complements and sources 238

Appendix to Chapter 10 238
10.A Proofs 238

11 Monotonicity of value function for POMDPs 241
11.1 Model and assumptions 242
11.2 Main result: monotone value function 244
11.3 Example 1: Monotone policies for 2-state POMDPs 245
11.4 Example 2: POMDP multi-armed bandits structural results 247
11.5 Complements and sources 251

Appendix to Chapter 11 251
11.A Proof of Theorem 11.3.1 251

12 Structural results for stopping time POMDPs 255
12.1 Introduction 255
12.2 Stopping time POMDP and convexity of stopping set 257
12.3 Monotone optimal policy for stopping time POMDP 261
12.4 Characterization of optimal linear decision threshold for stopping time POMDP 267
12.5 Example: Machine replacement POMDP 271
12.6 Multivariate stopping time POMDPs 271
12.7 Radar scheduling with mutual information cost 273
12.8 Complements and sources 278

Appendices to Chapter 12 279
12.A Lattices and submodularity 279
12.B MLR dominance and submodularity on lines 279
12.C Proof of Theorem 12.3.4 280

13 Stopping time POMDPs for quickest change detection 284
13.1 Example 1: Quickest detection with phase-distributed change time and variance penalty 285
13.2 Example 2: Quickest transient detection 290
13.3 Example 3: Risk-sensitive quickest detection with exponential delay penalty 291
13.4 Examples 4, 5 and 6: Stopping time POMDPs in multi-agent social learning 294
13.5 Example 7: Quickest detection with controlled sampling 304
13.6 Complements and sources 309

Appendix to Chapter 13 310
13.A Proof of Theorem 13.4.1 310

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-13460-7 — Partially Observed Markov Decision Processes
Vikram Krishnamurthy
Table of Contents
More Information

14 Myopic policy bounds for POMDPs and sensitivity to model parameters 312
14.1 The partially observed Markov decision process 313
14.2 Myopic policies using copositive dominance: insight 314
14.3 Constructing myopic policy bounds for optimal policy using copositive dominance 316
14.4 Optimizing the myopic policy bounds to match the optimal policy 318
14.5 Controlled sensing POMDPs with quadratic costs 320
14.6 Numerical examples 321
14.7 Blackwell dominance of observation distributions and optimality of myopic policies 324
14.8 Ordinal sensitivity: how does optimal POMDP cost vary with state and observation dynamics? 329
14.9 Cardinal sensitivity of POMDP 331
14.10 Complements and sources 332

Appendices to Chapter 14 332
14.A POMDP numerical examples 332
14.B Proof of Theorem 14.8.1 336
14.C Proof of Theorem 14.9.1 337

Part IV Stochastic approximation and reinforcement learning 341

15 Stochastic optimization and gradient estimation 343
15.1 Stochastic gradient algorithm 344
15.2 How long to simulate a Markov chain? 347
15.3 Gradient estimators for Markov processes 348
15.4 Finite difference gradient estimators and SPSA 349
15.5 Score function gradient estimator 350
15.6 Weak derivative gradient estimator 352
15.7 Bias and variance of gradient estimators 355
15.8 Complements and sources 357

Appendices to Chapter 15 358
15.A Proof of Theorem 15.2.1 358
15.B Proof of Theorem 15.7.1 360

16 Reinforcement learning 364
16.1 Q-learning algorithm 365
16.2 Policy gradient reinforcement learning for MDP 368
16.3 Score function policy gradient algorithm for MDP 371
16.4 Weak derivative gradient estimator for MDP 372
16.5 Numerical comparison of gradient estimators 374
16.6 Policy gradient reinforcement learning for constrained MDP (CMDP) 375

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-1-107-13460-7 — Partially Observed Markov Decision Processes
Vikram Krishnamurthy
Table of Contents
More Information

16.7 Policy gradient algorithm for POMDPs 377
16.8 Complements and sources 379

17 Stochastic approximation algorithms: examples 380
17.1 A primer on stochastic approximation algorithms 381
17.2 Example 1: Recursive maximum likelihood parameter estimation of HMMs 386
17.3 Example 2: HMM state estimation via LMS algorithm 389
17.4 Example 3: Discrete stochastic optimization for policy search 396
17.5 Example 4: Mean field population dynamics models for social sensing 407
17.6 Complements and sources 413

Appendices to Chapter 17 417
17.A Proof of Theorem 17.3.1 417
17.B Proof of Theorem 17.4.1 420
17.C Proof of Theorem 17.4.2 421
17.D Proof of Theorem 17.5.2 422

18 Summary of algorithms for solving POMDPs 425

Appendix A Short primer on stochastic simulation 428

Appendix B Continuous-time HMM filters 442

Appendix C Markov processes 449

Appendix D Some limit theorems 451

References 455
Index 471

© in this web service Cambridge University Press www.cambridge.org