

Contents

Preface	VII
1 Introduction	1
1.1 An Overview of Learning and Optimization	1
1.1.1 Problem Description	1
1.1.2 Optimal Policies	5
1.1.3 Fundamental Limitations of Learning and Optimization	12
1.1.4 A Sensitivity-Based View of Learning and Optimization	17
1.2 Problem Formulations in Different Disciplines	19
1.2.1 Perturbation Analysis (PA)	21
1.2.2 Markov Decision Processes (MDPs)	26
1.2.3 Reinforcement Learning (RL)	31
1.2.4 Identification and Adaptive Control (I&AC)	34
1.2.5 Event-Based Optimization and Potential Aggregation ..	37
1.3 A Map of the Learning and Optimization World	41
1.4 Terminology and Notation	42
Problems	44

Part I Four Disciplines in Learning and Optimization

2 Perturbation Analysis	51
2.1 Perturbation Analysis of Markov Chains	52
2.1.1 Constructing a Perturbed Sample Path	53
2.1.2 Perturbation Realization Factors and Performance Potentials	57
2.1.3 Performance Derivative Formulas	64
2.1.4 Gradients with Discounted Reward Criteria	68
2.1.5 Higher-Order Derivatives and the MacLaurin Series	74
2.2 Performance Sensitivities of Markov Processes	83
2.3 Performance Sensitivities of Semi-Markov Processes*	90
2.3.1 Fundamentals for Semi-Markov Processes*	90
2.3.2 Performance Sensitivity Formulas*	95
2.4 Perturbation Analysis of Queuing Systems	102
2.4.1 Constructing a Perturbed Sample Path	105

2.4.2	Perturbation Realization	115
2.4.3	Performance Derivatives	121
2.4.4	Remarks on Theoretical Issues*	125
2.5	Other Methods*	132
	Problems	137
3	Learning and Optimization with Perturbation Analysis	147
3.1	The Potentials	148
3.1.1	Numerical Methods	148
3.1.2	Learning Potentials from Sample Paths	151
3.1.3	Coupling*	156
3.2	Performance Derivatives	161
3.2.1	Estimating through Potentials	161
3.2.2	Learning Directly	162
3.3	Optimization with PA	172
3.3.1	Gradient Methods and Stochastic Approximation	172
3.3.2	Optimization with Long Sample Paths	174
3.3.3	Applications	177
	Problems	177
4	Markov Decision Processes	183
4.1	Ergodic Chains	185
4.1.1	Policy Iteration	186
4.1.2	Bias Optimality	192
4.1.3	MDPs with Discounted Rewards	201
4.2	Multi-Chains	203
4.2.1	Policy Iteration	205
4.2.2	Bias Optimality	216
4.2.3	MDPs with Discounted Rewards	226
4.3	The n th-Bias Optimization*	228
4.3.1	n th-Bias Difference Formulas*	229
4.3.2	Optimality Equations*	232
4.3.3	Policy Iteration*	240
4.3.4	n th-Bias Optimal Policy Spaces*	244
	Problems	246
5	Sample-Path-Based Policy Iteration	253
5.1	Motivation	254
5.2	Convergence Properties	258
5.2.1	Convergence of Potential Estimates	259
5.2.2	Sample Paths with a Fixed Number of Regenerative Periods	260
5.2.3	Sample Paths with Increasing Lengths	267
5.3	“Fast” Algorithms*	277

5.3.1	The Algorithm That Stops in a Finite Number of Periods*	278
5.3.2	With Stochastic Approximation*	282
	Problems	284
6	Reinforcement Learning	289
6.1	Stochastic Approximation	290
6.1.1	Finding the Zeros of a Function Recursively	291
6.1.2	Estimating Mean Values	297
6.2	Temporal Difference Methods	298
6.2.1	TD Methods for Potentials	298
6.2.2	Q-Factors and Other Extensions	308
6.2.3	TD Methods for Performance Derivatives	313
6.3	TD Methods and Performance Optimization	318
6.3.1	PA-Based Optimization	318
6.3.2	Q-Learning	321
6.3.3	Optimistic On-Line Policy Iteration	325
6.3.4	Value Iteration	327
6.4	Summary of the Learning and Optimization Methods	330
	Problems	333
7	Adaptive Control Problems as MDPs	341
7.1	Control Problems and MDPs	342
7.1.1	Control Systems Modelled as MDPs	342
7.1.2	A Comparison of the Two Approaches	345
7.2	MDPs with Continuous State Spaces	353
7.2.1	Operators on Continuous Spaces	354
7.2.2	Potentials and Policy Iteration	359
7.3	Linear Control Systems and the Riccati Equation	363
7.3.1	The LQ Problem	363
7.3.2	The JLQ Problem*	368
7.4	On-Line Optimization and Adaptive Control	373
7.4.1	Discretization and Estimation	374
7.4.2	Discussion	379
	Problems	381

Part II The Event-Based Optimization - A New Approach

8	Event-Based Optimization of Markov Systems	387
8.1	An Overview	388
8.1.1	Summary of Previous Chapters	388
8.1.2	An Overview of the Event-Based Approach	390
8.2	Events Associated with Markov Chains	398
8.2.1	The Event and Event Space	400

XVIII Contents

8.2.2	The Probabilities of Events	403
8.2.3	The Basic Ideas Illustrated by Examples	407
8.2.4	Classification of Three Types of Events	410
8.3	Event-Based Optimization	414
8.3.1	The Problem Formulation	414
8.3.2	Performance Difference Formulas	417
8.3.3	Performance Derivative Formulas	420
8.3.4	Optimization	425
8.4	Learning: Estimating Aggregated Potentials	429
8.4.1	Aggregated Potentials	429
8.4.2	Aggregated Potentials in the Event-Based Optimization	432
8.5	Applications and Examples	434
8.5.1	Manufacturing	434
8.5.2	Service Rate Control	438
8.5.3	General Applications	444
	Problems	446
9	Constructing Sensitivity Formulas	455
9.1	Motivation	455
9.2	Markov Chains on the Same State Space	456
9.3	Event-Based Systems	464
9.3.1	Sample-Path Construction*	464
9.3.2	Parameterized Systems: An Example	467
9.4	Markov Chains with Different State Spaces*	470
9.4.1	One Is a Subspace of the Other*	470
9.4.2	A More General Case	478
9.5	Summary	482
	Problems	483

Part III Appendices: Mathematical Background

A	Probability and Markov Processes	491
A.1	Probability	491
A.2	Markov Processes	498
	Problems	504
B	Stochastic Matrices	507
B.1	Canonical Form	507
B.2	Eigenvalues	508
B.3	The Limiting Matrix	511
	Problems	516

C Queueing Theory	519
C.1 Single-Server Queues	519
C.2 Queueing Networks	524
C.3 Some Useful Techniques	536
Problems	538
Notation and Abbreviations	543
References	547
Index	563