

---

# Contents

<b>Selected Notation and Abbreviations</b>	xv
<b>1 Markov Decision Processes</b>	1
1.1 Optimality Equations	3
1.2 Policy Iteration and Value Iteration	5
1.3 Rolling-horizon Control	7
1.4 Survey of Previous Work on Computational Methods	8
1.5 Simulation	11
1.6 Preview of Coming Attractions	14
1.7 Notes	15
<b>2 Multi-stage Adaptive Sampling Algorithms</b>	17
2.1 Upper Confidence Bound Sampling	19
2.1.1 Regret Analysis in Multi-armed Bandits	19
2.1.2 Algorithm Description	20
2.1.3 Alternative Estimators	21
2.1.4 Convergence Analysis	24
2.1.5 Numerical Example	31
2.2 Pursuit Learning Automata Sampling	40
2.2.1 Algorithm Description	41
2.2.2 Convergence Analysis	42
2.2.3 Application to POMDPs	51
2.2.4 Numerical Example	53
2.3 Notes	59
<b>3 Population-based Evolutionary Approaches</b>	61
3.1 Evolutionary Policy Iteration	63
3.1.1 Policy Switching	63
3.1.2 Policy Mutation and Population Generation	65
3.1.3 Stopping Rule	65
3.1.4 Convergence Analysis	66

3.1.5	Parallelization	67
3.2	Evolutionary Random Policy Search	67
3.2.1	Policy Improvement with Reward Swapping	68
3.2.2	Exploration	71
3.2.3	Convergence Analysis	73
3.3	Numerical Examples	76
3.3.1	A One-dimensional Queueing Example	76
3.3.2	A Two-dimensional Queueing Example	85
3.4	Extension to Simulation-based Setting	87
3.5	Notes	87
<b>4</b>	<b>Model Reference Adaptive Search</b>	<b>89</b>
4.1	The Model Reference Adaptive Search Method	91
4.1.1	The MRAS <sub>0</sub> Algorithm (Idealized Version)	93
4.1.2	The MRAS <sub>1</sub> Algorithm (Adaptive Monte Carlo Version)	96
4.1.3	The MRAS <sub>2</sub> Algorithm (Stochastic Optimization)	98
4.2	Convergence Analysis	101
4.2.1	MRAS <sub>0</sub> Convergence	101
4.2.2	MRAS <sub>1</sub> Convergence	107
4.2.3	MRAS <sub>2</sub> Convergence	116
4.3	Application to MDPs via Direct Policy Learning	129
4.3.1	Finite-horizon MDPs	130
4.3.2	Infinite-horizon MDPs	130
4.3.3	MDPs with Large State Spaces	132
4.3.4	Numerical Examples	132
4.4	Application to Infinite-horizon MDPs in Population-based Evolutionary Approaches	141
4.4.1	Algorithm Description	141
4.4.2	Numerical Examples	143
4.5	Application to Finite-horizon MDPs Using Adaptive Sampling	146
4.6	Notes	148
<b>5</b>	<b>On-line Control Methods via Simulation</b>	<b>149</b>
5.1	Simulated Annealing Multiplicative Weights Algorithm	153
5.1.1	Basic Algorithm Description	154
5.1.2	Convergence Analysis	155
5.1.3	Convergence of the Sampling Version of the Algorithm	158
5.1.4	Numerical Example	160
5.1.5	Simulated Policy Switching	164
5.2	Rollout	165
5.2.1	Parallel Rollout	166
5.3	Hindsight Optimization	168
5.3.1	Numerical Example	169
5.4	Notes	174

<b>References</b> .....	177
<b>Index</b> .....	187



Simulation-based Algorithms for Markov Decision  
Processes

Chang, H.S.; Fu, M.C.; Hu, J.; Marcus, S.I.

2007, XVIII, 189 p. 38 illus., Hardcover

ISBN: 978-1-84628-689-6