

# Contents

<b>1</b>	<b>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	10
1.6	Preview of Coming Attractions	13
1.7	Notes	14
<b>2</b>	<b>Multi-stage Adaptive Sampling Algorithms</b>	19
2.1	Upper Confidence Bound Sampling	21
2.1.1	Regret Analysis in Multi-armed Bandits	21
2.1.2	Algorithm Description	22
2.1.3	Alternative Estimators	25
2.1.4	Convergence Analysis	25
2.1.5	Numerical Example	33
2.2	Pursuit Learning Automata Sampling	37
2.2.1	Algorithm Description	42
2.2.2	Convergence Analysis	44
2.2.3	Application to POMDPs	52
2.2.4	Numerical Example	54
2.3	Notes	57
<b>3</b>	<b>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 . . . . .	83
3.4	Extension to Simulation-Based Setting . . . . .	86
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 $\text{MRAS}_0$ Algorithm (Idealized Version) . . . . .	92
4.1.2	The $\text{MRAS}_1$ Algorithm (Adaptive Monte Carlo Version) . . . . .	96
4.1.3	The $\text{MRAS}_2$ Algorithm (Stochastic Optimization) . . . . .	98
4.2	Convergence Analysis of MRAS . . . . .	101
4.2.1	$\text{MRAS}_0$ Convergence . . . . .	101
4.2.2	$\text{MRAS}_1$ Convergence . . . . .	107
4.2.3	$\text{MRAS}_2$ Convergence . . . . .	117
4.3	Application of MRAS to MDPs via Direct Policy Learning . . . . .	131
4.3.1	Finite-Horizon MDPs . . . . .	131
4.3.2	Infinite-Horizon MDPs . . . . .	132
4.3.3	MDPs with Large State Spaces . . . . .	132
4.3.4	Numerical Examples . . . . .	135
4.4	Application of MRAS to Infinite-Horizon MDPs in Population- Based Evolutionary Approaches . . . . .	141
4.4.1	Algorithm Description . . . . .	142
4.4.2	Numerical Examples . . . . .	143
4.5	Application of MRAS to Finite-Horizon MDPs Using Adaptive Sampling . . . . .	144
4.6	A Stochastic Approximation Framework . . . . .	148
4.6.1	Model-Based Annealing Random Search . . . . .	149
4.6.2	Application of MARS to Finite-Horizon MDPs . . . . .	166
4.7	Notes . . . . .	177
<b>5</b>	<b>On-Line Control Methods via Simulation . . . . .</b>	<b>179</b>
5.1	Simulated Annealing Multiplicative Weights Algorithm . . . . .	183
5.1.1	Basic Algorithm Description . . . . .	184
5.1.2	Convergence Analysis . . . . .	185
5.1.3	Convergence of the Sampling Version of the Algorithm . . . . .	189
5.1.4	Numerical Example . . . . .	191
5.1.5	Simulated Policy Switching . . . . .	194
5.2	Rollout . . . . .	195
5.2.1	Parallel Rollout . . . . .	197
5.3	Hindsight Optimization . . . . .	199
5.3.1	Numerical Example . . . . .	200
5.4	Approximate Stochastic Annealing . . . . .	204

5.4.1	Convergence Analysis . . . . .	207
5.4.2	Numerical Example . . . . .	215
5.5	Notes . . . . .	216
<b>References</b>	. . . . .	<b>219</b>
<b>Index</b>	. . . . .	<b>227</b>

Simulation-Based Algorithms for Markov Decision  
Processes

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

2013, XVII, 229 p. 49 illus., 1 illus. in color., Hardcover

ISBN: 978-1-4471-5021-3