

# Contents

Preface	vii
Acknowledgments	xi
List of Figures	xxi
List of Tables	xxvi
1. BACKGROUND	1
1 Motivation	1
1.1 Main Branches	2
1.2 Difficulties with Classical Optimization	3
1.3 Recent Advances in Simulation Optimization	3
2 Goals and Limitations	5
2.1 Goals	5
2.2 Limitations	6
3 Notation	6
3.1 Some Basic Conventions	6
3.2 Vector Notation	7
3.3 Notation for Matrices	8
3.4 Notation for $n$ -tuples	8
3.5 Notation for Sets	8
3.6 Notation for Sequences	9
3.7 Notation for Transformations	9
3.8 Max, Min, and Arg Max	10
4 Organization	10
2. SIMULATION BASICS	13
1 Chapter Overview	13
2 Introduction	13
3 Models	14
4 Simulation Modeling	16
4.1 Random Number Generation	16

4.2	Event Generation	20
4.3	Independence of Samples Collected	25
5	Concluding Remarks	27
3.	SIMULATION-BASED OPTIMIZATION: AN OVERVIEW	29
1	Chapter Overview	29
2	Parametric Optimization	29
3	Control Optimization	32
4	Concluding Remarks	35
4.	PARAMETRIC OPTIMIZATION: RESPONSE SURFACES AND NEURAL NETWORKS	37
1	Chapter Overview	37
2	RSM: An Overview	37
3	RSM: Details	39
3.1	Sampling	39
3.2	Function Fitting	40
3.3	How Good Is the Guessed Metamodel?	46
3.4	Optimization with a Metamodel	46
4	Neuro-Response Surface Methods	47
4.1	Linear Neural Networks	48
4.2	Non-linear Neural Networks	53
5	Concluding Remarks	69
5.	PARAMETRIC OPTIMIZATION: STOCHASTIC GRADIENTS AND ADAPTIVE SEARCH	71
1	Chapter Overview	71
2	Continuous Optimization	72
2.1	Steepest Descent	72
2.2	Non-derivative Methods	83
3	Discrete Optimization	85
3.1	Ranking and Selection	86
3.2	Meta-heuristics	89
3.3	Stochastic Adaptive Search	98
4	Concluding Remarks	120
6.	CONTROL OPTIMIZATION WITH STOCHASTIC DYNAMIC PROGRAMMING	123
1	Chapter Overview	123
2	Stochastic Processes	123
3	Markov, Semi-Markov, and Decision Processes	125
3.1	Markov Chains	129

3.2	Semi-Markov Processes	135
3.3	Markov Decision Problems	137
4	Average Reward MDPs and DP	150
4.1	Bellman Policy Equation	151
4.2	Policy Iteration	152
4.3	Value Iteration and Its Variants	154
5	Discounted Reward MDPs and DP	159
5.1	Discounted Reward	160
5.2	Discounted Reward MDPs	161
5.3	Bellman Policy Equation	162
5.4	Policy Iteration	163
5.5	Value Iteration	164
5.6	Gauss-Siedel Value Iteration	166
6	Bellman Equation Revisited	167
7	Semi-Markov Decision Problems	169
7.1	Natural and Decision-Making Processes	171
7.2	Average Reward SMDPs	173
7.3	Discounted Reward SMDPs	180
8	Modified Policy Iteration	184
8.1	Steps for Discounted Reward MDPs	185
8.2	Steps for Average Reward MDPs	186
9	The MDP and Mathematical Programming	187
10	Finite Horizon MDPs	189
11	Conclusions	193
7.	CONTROL OPTIMIZATION WITH REINFORCEMENT LEARNING	197
1	Chapter Overview	197
2	The Twin Curses of DP	198
2.1	Breaking the Curses	199
2.2	MLE and Small MDPs	201
3	Reinforcement Learning: Fundamentals	203
3.1	$Q$ -Factors	204
3.2	$Q$ -Factor Value Iteration	206
3.3	Robbins-Monro Algorithm	207
3.4	Robbins-Monro and $Q$ -Factors	208
3.5	Asynchronous Updating and Step Sizes	209
4	MDPs	211
4.1	Discounted Reward	211
4.2	Average Reward	231
4.3	R-SMART and Other Algorithms	233
5	SMDPs	234
5.1	Discounted Reward	234
5.2	Average Reward	235

6	Model-Building Algorithms . . . . .	244
6.1	RTDP . . . . .	245
6.2	Model-Building $Q$ -Learning . . . . .	246
6.3	Indirect Model-Building . . . . .	247
7	Finite Horizon Problems . . . . .	248
8	Function Approximation . . . . .	249
8.1	State Aggregation . . . . .	250
8.2	Function Fitting . . . . .	255
9	Conclusions . . . . .	265
8.	CONTROL OPTIMIZATION WITH STOCHASTIC SEARCH . . . . .	269
1	Chapter Overview . . . . .	269
2	The MCAT Framework . . . . .	270
2.1	Step-by-Step Details of an MCAT Algorithm . . . . .	272
2.2	An Illustrative 3-State Example . . . . .	274
2.3	Multiple Actions . . . . .	276
3	Actor Critics . . . . .	276
3.1	Discounted Reward MDPs . . . . .	277
3.2	Average Reward MDPs . . . . .	278
3.3	Average Reward SMDPs . . . . .	279
4	Concluding Remarks . . . . .	280
9.	CONVERGENCE: BACKGROUND MATERIAL . . . . .	281
1	Chapter Overview . . . . .	281
2	Vectors and Vector Spaces . . . . .	282
3	Norms . . . . .	284
3.1	Properties of Norms . . . . .	284
4	Normed Vector Spaces . . . . .	285
5	Functions and Mappings . . . . .	285
5.1	Domain and Range . . . . .	285
5.2	The Notation for Transformations . . . . .	287
6	Mathematical Induction . . . . .	287
7	Sequences . . . . .	290
7.1	Convergent Sequences . . . . .	292
7.2	Increasing and Decreasing Sequences . . . . .	293
7.3	Boundedness . . . . .	293
7.4	Limit Theorems and Squeeze Theorem . . . . .	299
8	Sequences in $\mathbb{R}^n$ . . . . .	301
9	Cauchy Sequences in $\mathbb{R}^n$ . . . . .	301
10	Contraction Mappings in $\mathbb{R}^n$ . . . . .	303

11	Stochastic Approximation	309
11.1	Convergence with Probability 1	309
11.2	Ordinary Differential Equations	310
11.3	Stochastic Approximation and ODEs	313
10.	CONVERGENCE ANALYSIS OF PARAMETRIC OPTIMIZATION METHODS	319
1	Chapter Overview	319
2	Preliminaries	319
2.1	Continuous Functions	320
2.2	Partial Derivatives	320
2.3	A Continuously Differentiable Function	320
2.4	Stationary Points and Local and Global Optima	321
2.5	Taylor's Theorem	322
3	Steepest Descent	324
4	Finite Differences Perturbation Estimates	327
5	Simultaneous Perturbation	328
5.1	Stochastic Gradient	329
5.2	ODE Approach	334
5.3	Spall's Conditions	336
6	Stochastic Adaptive Search	336
6.1	Pure Random Search	338
6.2	Learning Automata Search Technique	340
6.3	Backtracking Adaptive Search	341
6.4	Simulated Annealing	343
6.5	Modified Stochastic Ruler	349
7	Concluding Remarks	350
11.	CONVERGENCE ANALYSIS OF CONTROL OPTIMIZATION METHODS	351
1	Chapter Overview	351
2	Dynamic Programming: Background	351
2.1	Special Notation	353
2.2	Monotonicity of $T, T_{\hat{\mu}}, L$ , and $L_{\hat{\mu}}$	354
2.3	Key Results for Average and Discounted MDPs	355
3	Discounted Reward DP: MDPs	358
3.1	Bellman Equation for Discounted Reward	358
3.2	Policy Iteration	367
3.3	Value Iteration	369
4	Average Reward DP: MDPs	371
4.1	Bellman Equation for Average Reward	371
4.2	Policy Iteration	376
4.3	Value Iteration	380
5	DP: SMDPs	389

6	Asynchronous Stochastic Approximation . . . . .	390
6.1	Asynchronous Convergence . . . . .	390
6.2	Two-Time-Scale Convergence . . . . .	396
7	Reinforcement Learning: Convergence Background . . . . .	400
7.1	Discounted Reward MDPs . . . . .	401
7.2	Average Reward MDPs . . . . .	402
8	Reinforcement Learning for MDPs: Convergence . . . . .	404
8.1	$Q$ -Learning: Discounted Reward MDPs . . . . .	404
8.2	Relative $Q$ -Learning: Average Reward MDPs . . . . .	415
8.3	CAP-I: Discounted Reward MDPs . . . . .	417
8.4	$Q$ - $P$ -Learning: Discounted Reward MDPs . . . . .	422
9	Reinforcement Learning for SMDPs: Convergence . . . . .	424
9.1	$Q$ -Learning: Discounted Reward SMDPs . . . . .	424
9.2	Average Reward SMDPs . . . . .	425
10	Reinforcement Learning for Finite Horizon: Convergence . . . . .	447
11	Conclusions . . . . .	449
12.	CASE STUDIES . . . . .	451
1	Chapter Overview . . . . .	451
2	Airline Revenue Management . . . . .	451
3	Preventive Maintenance . . . . .	459
4	Production Line Buffer Optimization . . . . .	463
5	Other Case Studies . . . . .	468
6	Conclusions . . . . .	470
	APPENDIX . . . . .	473
	Bibliography . . . . .	477
	Index . . . . .	504

Simulation-Based Optimization  
Parametric Optimization Techniques and Reinforcement  
Learning

Gosavi, A.

2015, XXVI, 508 p. 42 illus., Hardcover

ISBN: 978-1-4899-7490-7