

# CONTENTS

---

Preface	xi
Acknowledgments	xv
<b>1 The challenges of dynamic programming</b>	<b>1</b>
1.1 A dynamic programming example: a shortest path problem	2
1.2 The three curses of dimensionality	3
1.3 Some real applications	6
1.4 Problem classes	10
1.5 The many dialects of dynamic programming	12
1.6 What is new in this book?	14
1.7 Bibliographic notes	16
<b>2 Some illustrative models</b>	<b>17</b>
2.1 Deterministic problems	18
2.2 Stochastic problems	23
2.3 Information acquisition problems	36
2.4 A simple modeling framework for dynamic programs	40
2.5 Bibliographic notes	43
Problems	43
	<b>v</b>

<b>3</b>	<b>Introduction to Markov decision processes</b>	<b>47</b>
3.1	The optimality equations	48
3.2	Finite horizon problems	53
3.3	Infinite horizon problems	55
3.4	Value iteration	57
3.5	Policy iteration	61
3.6	Hybrid value-policy iteration	63
3.7	The linear programming method for dynamic programs	63
3.8	Monotone policies*	64
3.9	Why does it work?***	70
3.10	Bibliographic notes	85
	Problems	86
<b>4</b>	<b>Introduction to approximate dynamic programming</b>	<b>91</b>
4.1	The three curses of dimensionality (revisited)	92
4.2	The basic idea	93
4.3	Sampling random variables	100
4.4	ADP using the post-decision state variable	101
4.5	Low-dimensional representations of value functions	107
4.6	So just what is approximate dynamic programming?	110
4.7	Experimental issues	112
4.8	Dynamic programming with missing or incomplete models	118
4.9	Relationship to reinforcement learning	119
4.10	But does it work?	120
4.11	Bibliographic notes	122
	Problems	123
<b>5</b>	<b>Modeling dynamic programs</b>	<b>129</b>
5.1	Notational style	131
5.2	Modeling time	132
5.3	Modeling resources	135
5.4	The states of our system	139
5.5	Modeling decisions	147
5.6	The exogenous information process	151
5.7	The transition function	159
5.8	The contribution function	166
5.9	The objective function	169
5.10	A measure-theoretic view of information***	170
5.11	Bibliographic notes	173
	Problems	173

<b>6</b>	<b>Stochastic approximation methods</b>	<b>179</b>
6.1	A stochastic gradient algorithm	181
6.2	Deterministic stepsize recipes	183
6.3	Stochastic stepsizes	190
6.4	Computing bias and variance	195
6.5	Optimal stepsizes	197
6.6	Some experimental comparisons of stepsize formulas	204
6.7	Convergence	208
6.8	Why does it work?*	210
6.9	Bibliographic notes	220
	Problems	221
<b>7</b>	<b>Approximating value functions</b>	<b>225</b>
7.1	Approximation using aggregation	226
7.2	Approximation methods using regression models	237
7.3	Recursive methods for regression models	246
7.4	Neural networks	253
7.5	Value function approximation for batch processes	257
7.6	Why does it work?*	263
7.7	Bibliographic notes	265
	Problems	267
<b>8</b>	<b>ADP for finite horizon problems</b>	<b>271</b>
8.1	Strategies for finite horizon problems	272
8.2	$Q$ -learning	276
8.3	Temporal difference learning	279
8.4	Policy iteration	282
8.5	Monte Carlo value and policy iteration	284
8.6	The actor-critic paradigm	285
8.7	Bias in value function estimation	286
8.8	State sampling strategies	290
8.9	Starting and stopping	294
8.10	A taxonomy of approximate dynamic programming strategies	296
8.11	Why does it work?*	298
8.12	Bibliographic notes	298
	Problems	299
<b>9</b>	<b>Infinite horizon problems</b>	<b>303</b>
9.1	From finite to infinite horizon	304
9.2	Algorithmic strategies	304
9.3	Stepsizes for infinite horizon problems	313

9.4	Error measures	315
9.5	Direct ADP for on-line applications	317
9.6	Finite horizon models for steady-state applications	317
9.7	Why does it work?*	319
9.8	Bibliographic notes	319
	Problems	320
<b>10</b>	<b>Exploration vs. exploitation</b>	<b>323</b>
10.1	A learning exercise: the nomadic trucker	323
10.2	Learning strategies	326
10.3	A simple information acquisition problem	330
10.4	Gittins indices and the information acquisition problem	332
10.5	Variations	337
10.6	The knowledge gradient algorithm	339
10.7	Information acquisition in dynamic programming	342
10.8	Bibliographic notes	346
	Problems	346
<b>11</b>	<b>Value function approximations for special functions</b>	<b>351</b>
11.1	Value functions versus gradients	352
11.2	Linear approximations	353
11.3	Piecewise linear approximations	355
11.4	The SHAPE algorithm	359
11.5	Regression methods	362
11.6	Cutting planes*	365
11.7	Why does it work?*	377
11.8	Bibliographic notes	383
	Problems	384
<b>12</b>	<b>Dynamic resource allocation problems</b>	<b>387</b>
12.1	An asset acquisition problem	388
12.2	The blood management problem	392
12.3	A portfolio optimization problem	401
12.4	A general resource allocation problem	404
12.5	A fleet management problem	416
12.6	A driver management problem	421
12.7	Bibliographic references	427
	Problems	427
<b>13</b>	<b>Implementation challenges</b>	<b>433</b>
13.1	Will ADP work for your problem?	433

13.2	Designing an ADP algorithm for complex problems	434
13.3	Debugging an ADP algorithm	436
13.4	Convergence issues	437
13.5	Modeling your problem	438
13.6	On-line vs. off-line models	440
13.7	If it works, patent it!	441
	Index	457

