Contents

Preface to the Second Edition xi
Preface to the First Edition xv
Acknowledgments xvii

1 The Challenges of Dynamic Programming 1

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, 11
1.5 The Many Dialects of Dynamic Programming, 15
1.6 What Is New in This Book?, 17
1.7 Pedagogy, 19
1.8 Bibliographic Notes, 22

2 Some Illustrative Models 25

2.1 Deterministic Problems, 26
2.2 Stochastic Problems, 31
2.3 Information Acquisition Problems, 47
2.4 A Simple Modeling Framework for Dynamic Programs, 50
2.5 Bibliographic Notes, 54

3 Introduction to Markov Decision Processes 57

3.1 The Optimality Equations, 58
3.2 Finite Horizon Problems, 65
3.3 Infinite Horizon Problems, 66
3.4 Value Iteration, 68
3.5 Policy Iteration, 74
3.6 Hybrid Value-Policy Iteration, 75
3.7 Average Reward Dynamic Programming, 76
3.8 The Linear Programming Method for Dynamic Programs, 77
3.9 Monotone Policies*, 78
3.10 Why Does It Work??, 84
3.11 Bibliographic Notes, 103
   Problems, 103

4 Introduction to Approximate Dynamic Programming 111
   4.1 The Three Curses of Dimensionality (Revisited), 112
   4.2 The Basic Idea, 114
   4.3 Q-Learning and SARSA, 122
   4.4 Real-Time Dynamic Programming, 126
   4.5 Approximate Value Iteration, 127
   4.6 The Post-Decision State Variable, 129
   4.7 Low-Dimensional Representations of Value Functions, 144
   4.8 So Just What Is Approximate Dynamic Programming?, 146
   4.9 Experimental Issues, 149
   4.10 But Does It Work?, 155
   4.11 Bibliographic Notes, 156
      Problems, 158

5 Modeling Dynamic Programs 167
   5.1 Notational Style, 169
   5.2 Modeling Time, 170
   5.3 Modeling Resources, 174
   5.4 The States of Our System, 178
   5.5 Modeling Decisions, 187
   5.6 The Exogenous Information Process, 189
   5.7 The Transition Function, 198
   5.8 The Objective Function, 206
   5.9 A Measure-Theoretic View of Information**, 211
   5.10 Bibliographic Notes, 213
      Problems, 214
6 Policies

6.1 Myopic Policies, 224
6.2 Lookahead Policies, 224
6.3 Policy Function Approximations, 232
6.4 Value Function Approximations, 235
6.5 Hybrid Strategies, 239
6.6 Randomized Policies, 242
6.7 How to Choose a Policy?, 244
6.8 Bibliographic Notes, 247
Problems, 247

7 Policy Search

7.1 Background, 250
7.2 Gradient Search, 253
7.3 Direct Policy Search for Finite Alternatives, 256
7.4 The Knowledge Gradient Algorithm for Discrete Alternatives, 262
7.5 Simulation Optimization, 270
7.6 Why Does It Work?**, 274
7.7 Bibliographic Notes, 285
Problems, 286

8 Approximating Value Functions

8.1 Lookup Tables and Aggregation, 290
8.2 Parametric Models, 304
8.3 Regression Variations, 314
8.4 Nonparametric Models, 316
8.5 Approximations and the Curse of Dimensionality, 325
8.6 Why Does It Work?**, 328
8.7 Bibliographic Notes, 333
Problems, 334

9 Learning Value Function Approximations

9.1 Sampling the Value of a Policy, 337
9.2 Stochastic Approximation Methods, 347
9.3 Recursive Least Squares for Linear Models, 349
9.4 Temporal Difference Learning with a Linear Model, 356
9.5 Bellman’s Equation Using a Linear Model, 358
9.6 Analysis of TD(0), LSTD, and LSPE Using a Single State, 364
9.7 Gradient-Based Methods for Approximate Value Iteration*, 366
9.8 Least Squares Temporal Differencing with Kernel Regression*, 371
9.9 Value Function Approximations Based on Bayesian Learning*, 373
9.10 Why Does It Work*, 376
9.11 Bibliographic Notes, 379
Problems, 381

10 Optimizing While Learning 383
10.1 Overview of Algorithmic Strategies, 385
10.2 Approximate Value Iteration and \(Q\)-Learning Using Lookup Tables, 386
10.3 Statistical Bias in the Max Operator, 397
10.4 Approximate Value Iteration and \(Q\)-Learning Using Linear Models, 400
10.5 Approximate Policy Iteration, 402
10.6 The Actor–Critic Paradigm, 408
10.7 Policy Gradient Methods, 410
10.8 The Linear Programming Method Using Basis Functions, 411
10.9 Approximate Policy Iteration Using Kernel Regression*, 413
10.10 Finite Horizon Approximations for Steady-State Applications, 415
10.11 Bibliographic Notes, 416
Problems, 418

11 Adaptive Estimation and Stepsizes 419
11.1 Learning Algorithms and Stepsizes, 420
11.2 Deterministic Step size Recipes, 425
11.3 Stochastic Stepsizes, 433
11.4 Optimal Stepsizes for Nonstationary Time Series, 437
11.5 Optimal Stepsizes for Approximate Value Iteration, 447
11.6 Convergence, 449
11.7 Guidelines for Choosing Step Size Formulas, 451
11.8 Bibliographic Notes, 452
Problems, 453
CONTENTS

12 Exploration Versus Exploitation 457
  12.1 A Learning Exercise: The Nomadic Trucker, 457
  12.2 An Introduction to Learning, 460
  12.3 Heuristic Learning Policies, 464
  12.4 Gittins Indexes for Online Learning, 470
  12.5 The Knowledge Gradient Policy, 477
  12.6 Learning with a Physical State, 482
  12.7 Bibliographic Notes, 492
    Problems, 493

13 Value Function Approximations for Resource Allocation Problems 497
  13.1 Value Functions versus Gradients, 498
  13.2 Linear Approximations, 499
  13.3 Piecewise-Linear Approximations, 501
  13.4 Solving a Resource Allocation Problem Using Piecewise-Linear Functions, 505
  13.5 The SHAPE Algorithm, 509
  13.6 Regression Methods, 513
  13.7 Cutting Planes*, 516
  13.8 Why Does It Work?**, 528
  13.9 Bibliographic Notes, 535
    Problems, 536

14 Dynamic Resource Allocation Problems 541
  14.1 An Asset Acquisition Problem, 541
  14.2 The Blood Management Problem, 547
  14.3 A Portfolio Optimization Problem, 557
  14.4 A General Resource Allocation Problem, 560
  14.5 A Fleet Management Problem, 573
  14.6 A Driver Management Problem, 580
  14.7 Bibliographic Notes, 585
    Problems, 586

15 Implementation Challenges 593
  15.1 Will ADP Work for Your Problem?, 593
  15.2 Designing an ADP Algorithm for Complex Problems, 594
  15.3 Debugging an ADP Algorithm, 596