# Contents

**Foreword**

*Shankar Sastry*

1. **ADP: Goals, Opportunities and Principles**
   *Paul Werbos*
   1.1 Goals of This Book 3
   1.2 Funding Issues, Opportunities and the Larger Context 5
   1.3 Unifying Mathematical Principles and Roadmap of the Field 17

**Part I Overview**

2. **Reinforcement Learning and Its Relationship to Supervised Learning**
   *Andrew G. Barto and Thomas G. Dietterich*
   2.1 Introduction 47
   2.2 Supervised Learning 48
   2.3 Reinforcement Learning 50
   2.4 Sequential Decision Tasks 54
   2.5 Supervised Learning for Sequential Decision Tasks 58
   2.6 Concluding Remarks 60

3. **Model-Based Adaptive Critic Designs**
   *Silvia Ferrari and Robert F. Stengel*
   3.1 Introduction 65
   3.2 Mathematical Background and Foundations 67
   3.3 Adaptive Critic Design and Implementation 74
   3.4 Discussion 88
   3.5 Summary 89
## CONTENTS

### 4 Guidance in the Use of Adaptive Critics for Control  
*George G. Lendaris and James C. Neidhoefer*

4.1 Introduction  
4.2 Reinforcement Learning  
4.3 Dynamic Programming  
4.4 Adaptive Critics: “Approximate Dynamic Programming”  
4.5 Some Current Research on Adaptive Critic Technology  
4.6 Application Issues  
4.7 Items for Future ADP Research

### 5 Direct Neural Dynamic Programming  
*Jennie Si, Lei Yang and Derong Liu*

5.1 Introduction  
5.2 Problem Formulation  
5.3 Implementation of Direct NDP  
5.4 Comparisons  
5.5 Continuous State Control Problem  
5.6 Call Admission Control for CDMA Cellular Networks  
5.7 Conclusions and Discussions

### 6 The Linear Programming Approach to Approximate Dynamic Programming  
*Daniela Pucci de Farias*

6.1 Introduction  
6.2 Markov Decision Processes  
6.3 Approximate Linear Programming  
6.4 State-Relevance Weights and the Performance of ALP Policies  
6.5 Approximation Error Bounds  
6.6 Application to Queueing Networks  
6.7 Efficient Constraint Sampling Scheme  
6.8 Discussion

### 7 Reinforcement Learning in Large, High-Dimensional State Spaces  
*Greg Grudic and Lyle Ungar*

7.1 Introduction  
7.2 Theoretical Results and Algorithm Specifications  
7.3 Experimental Results  
7.4 Conclusion
8 Hierarchical Decision Making
Malcolm Ryan

8.1 Introduction 203
8.2 Reinforcement Learning and the Curse of Dimensionality 204
8.3 Hierarchical Reinforcement Learning in Theory 209
8.4 Hierarchical Reinforcement Learning in Practice 217
8.5 Termination Improvement 221
8.6 Intra-Behavior Learning 223
8.7 Creating Behaviors and Building Hierarchies 225
8.8 Model-based Reinforcement Learning 225
8.9 Topics for Future Research 226
8.10 Conclusion 227

Part II Technical Advances 233

9 Improved Temporal Difference Methods with Linear Function Approximation
Dimitri P. Bertsekas, Vivek S. Borkar, and Angelia Nedich

9.1 Introduction 235
9.2 Preliminary Analysis 241
9.3 Convergence Analysis 243
9.4 Relations Between λ-LSPE and Value Iteration 245
9.5 Relation Between λ-LSPE and LSTD 252
9.6 Computational Comparison of λ-LSPE and TD(λ) 253

10 Approximate Dynamic Programming for High-Dimensional Resource Allocation Problems
Warren B. Powell and Benjamin Van Roy

10.1 Introduction 261
10.2 Dynamic Resource Allocation 262
10.3 Curses of Dimensionality 265
10.4 Algorithms for Dynamic Resource Allocation 266
10.5 Mathematical programming 271
10.6 Approximate Dynamic Programming 275
10.7 Experimental Comparisons 277
10.8 Conclusion 279
## 11 Hierarchical Approaches to Concurrency, Multiagency, and Partial Observability

Sridhar Mahadevan, Mohammad Ghavamzadeh, Khashayar Rohanimanesh, and Georgios Theocharous

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.1 Introduction</td>
<td>285</td>
</tr>
<tr>
<td>11.2 Background</td>
<td>287</td>
</tr>
<tr>
<td>11.3 Spatiotemporal Abstraction of Markov Processes</td>
<td>289</td>
</tr>
<tr>
<td>11.4 Concurrency, Multiagency, and Partial Observability</td>
<td>294</td>
</tr>
<tr>
<td>11.5 Summary and Conclusions</td>
<td>306</td>
</tr>
</tbody>
</table>

## 12 Learning and Optimization — From a System Theoretic Perspective

Xi-Ren Cao

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.1 Introduction</td>
<td>311</td>
</tr>
<tr>
<td>12.2 General View of Optimization</td>
<td>313</td>
</tr>
<tr>
<td>12.3 Estimation of Potentials and Performance Derivatives</td>
<td>316</td>
</tr>
<tr>
<td>12.4 Gradient-Based Optimization</td>
<td>323</td>
</tr>
<tr>
<td>12.5 Policy Iteration</td>
<td>324</td>
</tr>
<tr>
<td>12.6 Constructing Performance Gradients with Potentials as Building Blocks</td>
<td>328</td>
</tr>
<tr>
<td>12.7 Conclusion</td>
<td>330</td>
</tr>
</tbody>
</table>

## 13 Robust Reinforcement Learning Using Integral-Quadratic Constraints

Charles W. Anderson, Matt Kretchmar, Peter Young, and Douglas Hittle

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.1 Introduction</td>
<td>337</td>
</tr>
<tr>
<td>13.2 Integral-Quadratic Constraints and Stability Analysis</td>
<td>338</td>
</tr>
<tr>
<td>13.3 Reinforcement Learning in the Robust Control Framework</td>
<td>340</td>
</tr>
<tr>
<td>13.4 Demonstrations of Robust Reinforcement Learning</td>
<td>346</td>
</tr>
<tr>
<td>13.5 Conclusions</td>
<td>354</td>
</tr>
</tbody>
</table>

## 14 Supervised Actor-Critic Reinforcement Learning

Michael T. Rosenstein and Andrew G. Barto

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.1 Introduction</td>
<td>359</td>
</tr>
<tr>
<td>14.2 Supervised Actor-Critic Architecture</td>
<td>361</td>
</tr>
<tr>
<td>14.3 Examples</td>
<td>366</td>
</tr>
<tr>
<td>14.4 Conclusions</td>
<td>375</td>
</tr>
</tbody>
</table>
15 BPTT and DAC — A Common Framework for Comparison

Danil V. Prokhorov

15.1 Introduction
15.2 Relationship between BPTT and DAC
15.3 Critic representation
15.4 Hybrid of BPTT and DAC
15.5 Computational complexity and other issues
15.6 Two classes of challenging problems
15.7 Conclusion

Part III Applications

16 Near-Optimal Control Via Reinforcement Learning

Augustine O. Esogbue and Warren E. Hearnes II

16.1 Introduction
16.2 Terminal Control Processes
16.3 A Hybridization: The GCS-Δ Controller
16.4 Experimental Investigation of the GCS-Δ Algorithm
16.5 Dynamic Allocation of Controller Resources
16.6 Conclusions and Future Research

17 Multiobjective Control Problems by Reinforcement Learning

Dong-Oh Kang and Zeungnam Bien

17.1 Introduction
17.2 Preliminary
17.3 Policy Improvement Algorithm with Vector-Valued Reward
17.4 Multi-Reward Reinforcement Learning for Fuzzy Control
17.5 Summary

18 Adaptive Critic Based Neural Network for Control-Constrained Agile Missile

S. N. Balakrishnan and Dongchen Han

18.1 Introduction
18.2 Problem Formulation and Solution Development
18.3 Minimum Time Heading Reversal Problem in a Vertical Plane
18.4 Use of Networks in Real-Time as Feedback Control
18.5 Numerical Results
18.6 Conclusions
## 19 Applications of Approximate Dynamic Programming in Power Systems Control
*Ganesh K Venayagamoorthy, Ronald G Harley, and Donald C Wunsch*

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.1 Introduction</td>
<td>479</td>
</tr>
<tr>
<td>19.2 Adaptive Critic Designs and Approximate Dynamic Programming</td>
<td>483</td>
</tr>
<tr>
<td>19.3 General Training Procedure for Critic and Action Networks</td>
<td>493</td>
</tr>
<tr>
<td>19.4 Power System</td>
<td>494</td>
</tr>
<tr>
<td>19.5 Simulation and Hardware Implementation Results</td>
<td>496</td>
</tr>
<tr>
<td>19.6 Summary</td>
<td>510</td>
</tr>
</tbody>
</table>

## 20 Robust Reinforcement Learning for Heating, Ventilation, and Air Conditioning Control of Buildings
*Charles W. Anderson, Douglas Hittle, Matt Kretchmar, and Peter Young*

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.1 Introduction</td>
<td>517</td>
</tr>
<tr>
<td>20.2 Heating Coil Model and PI Control</td>
<td>521</td>
</tr>
<tr>
<td>20.3 Combined PI and Reinforcement Learning Control</td>
<td>522</td>
</tr>
<tr>
<td>20.4 Robust Control Framework for Combined PI and RL Control</td>
<td>525</td>
</tr>
<tr>
<td>20.5 Conclusions</td>
<td>529</td>
</tr>
</tbody>
</table>

## 21 Helicopter Flight Control Using Direct Neural Dynamic Programming
*Russell Enns and Jennie Si*

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.1 Introduction</td>
<td>535</td>
</tr>
<tr>
<td>21.2 Helicopter Model</td>
<td>538</td>
</tr>
<tr>
<td>21.3 Direct NDP Mechanism Applied to Helicopter Stability Control</td>
<td>540</td>
</tr>
<tr>
<td>21.4 Direct NDP Mechanism Applied to Helicopter Tracking Control</td>
<td>548</td>
</tr>
<tr>
<td>21.5 Reconfigurable Flight Control</td>
<td>553</td>
</tr>
<tr>
<td>21.6 Conclusions</td>
<td>556</td>
</tr>
</tbody>
</table>

## 22 Toward Dynamic Stochastic Optimal Power Flow
*James A. Momoh*

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.1 Grand Overview of the Plan for the Future Optimal Power Flow</td>
<td>561</td>
</tr>
<tr>
<td>22.2 Generalized Formulation of the OPF Problem</td>
<td>567</td>
</tr>
<tr>
<td>22.3 General Optimization Techniques Used in Solving the OPF Problem</td>
<td>571</td>
</tr>
<tr>
<td>22.4 State-of-the-Art Technology in OPF Programs: The Quadratic Interior Point (QIP) Method</td>
<td>575</td>
</tr>
<tr>
<td>22.5 Strategy for Future OPF Development</td>
<td>576</td>
</tr>
<tr>
<td>22.6 Conclusion</td>
<td>596</td>
</tr>
</tbody>
</table>
23 Control, Optimization, Security, and Self-healing of Benchmark Power Systems

James A. Momoh and Edwin Zivi

23.1 Introduction 599
23.2 Description of the Benchmark Systems 601
23.3 Illustrative Terrestrial Power System Challenge Problems 604
23.4 Illustrative Navy Power System Challenge Problems 614
23.5 Summary of Power System Challenges and Topics 629
23.6 Summary 633