REINFORCEMENT LEARNING AND APPROXIMATE DYNAMIC PROGRAMMING FOR FEEDBACK CONTROL
REINFORCEMENT LEARNING AND APPROXIMATE DYNAMIC PROGRAMMING FOR FEEDBACK CONTROL

Edited by

Frank L. Lewis
UTA Automation and Robotics Research Institute
Fort Worth, TX

Derong Liu
University of Illinois
Chicago, IL
PART I FEEDBACK CONTROL USING RL AND ADP

1. Reinforcement Learning and Approximate Dynamic Programming (RLADP)—Foundations, Common Misconceptions, and the Challenges Ahead 3
Paul J. Werbos
1.1 Introduction 3
1.2 What is RLADP? 4
  1.2.1 Definition of RLADP and the Task it Addresses 4
  1.2.2 Basic Tools—Bellman Equation, and Value and Policy Functions 9
  1.2.3 Optimization Over Time Without Value Functions 13
1.3 Some Basic Challenges in Implementing ADP 14
  1.3.1 Accounting for Unseen Variables 15
  1.3.2 Offline Controller Design Versus Real-Time Learning 17
  1.3.3 “Model-Based” Versus “Model Free” Designs 18
  1.3.4 How to Approximate the Value Function Better 19
  1.3.5 How to Choose \( u(t) \) Based on a Value Function 22
  1.3.6 How to Build Cooperative Multiagent Systems with RLADP 25
References 26

2. Stable Adaptive Neural Control of Partially Observable Dynamic Systems 31
J. Nate Knight and Charles W. Anderson
2.1 Introduction 31
2.2 Background 32
2.3 Stability Bias 35
2.4 Example Application 38
  2.4.1 The Simulated System 38
  2.4.2 An Uncertain Linear Plant Model 40
## CONTENTS

2.4.3 The Closed Loop Control System 41  
2.4.4 Determining RNN Weight Updates by Reinforcement Learning 44  
2.4.5 Results 46  
2.4.6 Conclusions 50  
References 50

3. **Optimal Control of Unknown Nonlinear Discrete-Time Systems Using the Iterative Globalized Dual Heuristic Programming Algorithm** 52  
Derong Liu and Ding Wang  
3.1 Background Material 53  
3.2 Neuro-Optimal Control Scheme Based on the Iterative ADP Algorithm 55  
3.2.1 Identification of the Unknown Nonlinear System 55  
3.2.2 Derivation of the Iterative ADP Algorithm 59  
3.2.3 Convergence Analysis of the Iterative ADP Algorithm 59  
3.2.4 Design Procedure of the Iterative ADP Algorithm 64  
3.2.5 NN Implementation of the Iterative ADP Algorithm Using GDHP Technique 64  
3.3 Generalization 67  
3.4 Simulation Studies 68  
3.5 Summary 74  
References 74

4. **Learning and Optimization in Hierarchical Adaptive Critic Design** 78  
Haibo He, Zhen Ni, and Dongbin Zhao  
4.1 Introduction 78  
4.2 Hierarchical ADP Architecture with Multiple-Goal Representation 80  
4.2.1 System Level Structure 80  
4.2.2 Architecture Design and Implementation 81  
4.2.3 Learning and Adaptation in Hierarchical ADP 83  
4.3 Case Study: The Ball-and-Beam System 87  
4.3.1 Problem Formulation 88  
4.3.2 Experiment Configuration and Parameters Setup 89  
4.3.3 Simulation Results and Analysis 90  
4.4 Conclusions and Future Work 94  
References 95

5. **Single Network Adaptive Critics Networks—Development, Analysis, and Applications** 98  
Jie Ding, Ali Heydari, and S.N. Balakrishnan  
5.1 Introduction 98  
5.2 Approximate Dynamic Programing 100
5.3  SNAC
   5.3.1  State Generation for Neural Network Training 103
   5.3.2  Neural Network Training 103
   5.3.3  Convergence Condition 104

5.4  J-SNAC
   5.4.1  Neural Network Training 105
   5.4.2  Numerical Analysis 105

5.5  Finite-SNAC
   5.5.1  Neural Network Training 109
   5.5.2  Convergence Theorems 111
   5.5.3  Numerical Analysis 112

5.6  Conclusions 116

References 116

6.  Linearly Solvable Optimal Control 119
K. Dvijotham and E. Todorov

6.1  Introduction 119
   6.1.1  Notation 121
   6.1.2  Markov Decision Processes 122

6.2  Linearly Solvable Optimal Control Problems 123
   6.2.1  Probability Shift: An Alternate View of Control 123
   6.2.2  Linearly Solvable Markov Decision Processes (LMDPs) 124
   6.2.3  An Alternate View of LMDPs 124
   6.2.4  Other Problem Formulations 126
   6.2.5  Applications 126
   6.2.6  Linearly Solvable Controlled Diffusions (LDs) 127
   6.2.7  Relationship Between Discrete and Continuous-Time Problems 128
   6.2.8  Historical Perspective 129

6.3  Extension to Risk-Sensitive Control and Game Theory 130
   6.3.1  Game Theoretic Control: Competitive Games 130
   6.3.2  Rényi Divergence 130
   6.3.3  Linearly Solvable Markov Games 130
   6.3.4  Linearly Solvable Differential Games 133
   6.3.5  Relationships Among the Different Formulations 134

6.4  Properties and Algorithms 134
   6.4.1  Sampling Approximations and Path-Integral Control 134
   6.4.2  Residual Minimization via Function Approximation 135
   6.4.3  Natural Policy Gradient 136
   6.4.4  Compositionality of Optimal Control Laws 136
   6.4.5  Stochastic Maximum Principle 137
   6.4.6  Inverse Optimal Control 138

6.5  Conclusions and Future Work 139

References 139
7. **Approximating Optimal Control with Value Gradient Learning**  
Michael Fairbank, Danil Prokhorov, and Eduardo Alonso  
7.1 Introduction  
7.2 Value Gradient Learning and BPTT Algorithms  
7.2.1 Preliminary Definitions  
7.2.2 VGL(\(\lambda\)) Algorithm  
7.2.3 BPTT Algorithm  
7.3 A Convergence Proof for VGL(1) for Control with Function Approximation  
7.3.1 Using a Greedy Policy with a Critic Function  
7.3.2 The Equivalence of VGL(1) to BPTT  
7.3.3 Convergence Conditions  
7.3.4 Notes on the \(\Omega_1\) Matrix  
7.4 Vertical Lander Experiment  
7.4.1 Problem Definition  
7.4.2 Efficient Evaluation of the Greedy Policy  
7.4.3 Observations on the Purpose of \(\Omega_1\)  
7.4.4 Experimental Results for Vertical Lander Problem  
7.5 Conclusions  
References  

8. **A Constrained Backpropagation Approach to Function Approximation and Approximate Dynamic Programming**  
Silvia Ferrari, Keith Rudd, and Gianluca Di Muro  
8.1 Background  
8.2 Constrained Backpropagation (CPROP) Approach  
8.2.1 Neural Network Architecture and Procedural Memories  
8.2.2 Derivation of LTM Equality Constraints and Adjoined Error Gradient  
8.2.3 Example: Incremental Function Approximation  
8.3 Solution of Partial Differential Equations in Nonstationary Environments  
8.3.1 CPROP Solution of Boundary Value Problems  
8.3.2 Example: PDE Solution on a Unit Circle  
8.3.3 CPROP Solution to Parabolic PDEs  
8.4 Preserving Prior Knowledge in Exploratory Adaptive Critic Designs  
8.4.1 Derivation of LTM Constraints for Feedback Control  
8.4.2 Constrained Adaptive Critic Design  
8.5 Summary  
Appendix: Algebraic ANN Control Matrices  
References
9. Toward Design of Nonlinear ADP Learning Controllers with Performance Assurance
Jennie Si, Lei Yang, Chao Lu, Kostas S. Tsakalis, and Armando A. Rodriguez
9.1 Introduction 183
9.2 Direct Heuristic Dynamic Programming 184
9.3 A Control Theoretic View on the Direct HDP 186
9.3.1 Problem Setup 187
9.3.2 Frequency Domain Analysis of Direct HDP 189
9.3.3 Insight from Comparing Direct HDP to LQR 192
9.4 Direct HDP Design with Improved Performance Case 1—Design Guided by a Priori LQR Information 193
9.4.1 Direct HDP Design Guided by a Priori LQR Information 193
9.4.2 Performance of the Direct HDP Beyond Linearization 195
9.5 Direct HDP Design with Improved Performance Case 2—Direct HDP for Coorindated Damping Control of Low-Frequency Oscillation 198
9.6 Summary 201
References 202

10. Reinforcement Learning Control with Time-Dependent Agent Dynamics
Kenton Kirkpatrick and John Valasek
10.1 Introduction 203
10.2 Q-Learning 205
10.2.1 Q-Learning Algorithm 205
10.2.2 ε-Greedy 207
10.2.3 Function Approximation 208
10.3 Sampled Data Q-Learning 209
10.3.1 Sampled Data Q-Learning Algorithm 209
10.3.2 Example 210
10.4 System Dynamics Approximation 213
10.4.1 First-Order Dynamics Learning 214
10.4.2 Multiagent System Thought Experiment 216
10.5 Closing Remarks 218
References 219

Hassan Zargarzadeh, Qinmin Yang, and S. Jagannathan
11.1 Introduction 221
11.2 Background 224
11.3 Reinforcement Learning Based Control 225
11.3.1 Affine-Like Dynamics 225
11.3.2 Online Reinforcement Learning Controller Design 229
11.3.3 The Action NN Design 229
11.3.4 The Critic NN Design 230
11.3.5 Weight Updating Laws for the NNs 231
11.3.6 Main Theoretic Results 232

11.4 Time-Based Adaptive Dynamic Programming-Based Optimal Control 234
11.4.1 Online NN-Based Identifier 235
11.4.2 Neural Network-Based Optimal Controller Design 237
11.4.3 Cost Function Approximation for Optimal Regulator Design 238
11.4.4 Estimation of the Optimal Feedback Control Signal 240
11.4.5 Convergence Proof 242
11.4.6 Robustness 244

11.5 Simulation Result 247
11.5.1 Reinforcement-Learning-Based Control of a Nonlinear System 247
11.5.2 The Drawback of HDP Policy Iteration Approach 250
11.5.3 OLA-Based Optimal Control Applied to HCCI Engine 251

References 255

12.1 Introduction 259
12.2 Actor–Critic–Identifier Architecture for HJB Approximation 260
12.3 Actor–Critic Design 263
12.4 Identifier Design 264
12.5 Convergence and Stability Analysis 270
12.6 Simulation 274
12.7 Conclusion 275

References 278

13. Robust Adaptive Dynamic Programming 281
Yu Jiang and Zhong-Ping Jiang
13.1 Introduction 281
13.2 Optimality Versus Robustness 283
13.2.1 Systems with Matched Disturbance Input 283
13.2.2 Adding One Integrator 284
13.2.3 Systems in Lower-Triangular Form 286
13.3 Robust-ADP Design for Disturbance Attenuation 288
13.3.1 Horizontal Learning 288
13.3.2 Vertical Learning 290
13.3.3 Robust-ADP Algorithm for Disturbance Attenuation 291
13.4 Robust-ADP for Partial-State Feedback Control 292

References 295
PART II LEARNING AND CONTROL IN MULTIAGENT GAMES

Quanyan Zhu, Hamidou Tembine, and Tamer Başar
14.1 Introduction
14.1.1 Related Work
14.1.2 Contribution
14.1.3 Organization of the Chapter
14.2 Two-Person Game
14.3 Learning in NZSGs
14.3.1 Learning Procedures
14.3.2 Learning Schemes
14.4 Main Results
14.4.1 Stochastic Approximation of the Pure Learning Schemes
14.4.2 Stochastic Approximation of the Hybrid Learning Scheme
14.4.3 Connection with Equilibria of the Expected Game
14.5 Security Application
14.6 Conclusions and Future Works
Appendix: Assumptions for Stochastic Approximation
References

15. Integral Reinforcement Learning for Online Computation of Nash Strategies of Nonzero-Sum Differential Games
Draguna Vrabie and F.L. Lewis
15.1 Introduction
15.2 Two-Player Games and Integral Reinforcement Learning
15.2.1 Two-Player Nonzero-Sum Games and Nash Equilibrium
15.2.2 Integral Reinforcement Learning for Two-Player Nonzero-Sum Games
15.3 Continuous-Time Value Iteration to Solve the Riccati Equation
15.4 Online Algorithm to Solve Nonzero-Sum Games
### 15. Online Learning Algorithms for Solving Nonzero-Sum Games

15.1 Finding Stabilizing Gains to Initialize the Online Algorithm
15.2 Online Partially Model-Free Algorithm for Solving the Nonzero-Sum Differential Game
15.3 Adaptive Critic Structure for Solving the Two-Player Nash Differential Game

### 16. Online Learning Algorithms for Optimal Control and Dynamic Games

16.1 Introduction
16.2 Optimal Control and the Continuous Time Hamilton-Jacobi-Bellman Equation
16.2.1 Optimal Control and Hamilton-Jacobi-Bellman Equation
16.2.2 Policy Iteration for Optimal Control
16.2.3 Online Synchronous Policy Iteration
16.2.4 Simulation

16.3 Online Solution of Nonlinear Two-Player Zero-Sum Games and Hamilton-Jacobi-Isaacs Equation
16.3.1 Zero-Sum Games and Hamilton-Jacobi-Isaacs Equation
16.3.2 Policy Iteration for Two-Player Zero-Sum Differential Games
16.3.3 Online Solution for Two-Player Zero-Sum Differential Games
16.3.4 Simulation

16.4 Online Solution of Nonlinear Nonzero-Sum Games and Coupled Hamilton-Jacobi Equations
16.4.1 Nonzero Sum Games and Coupled Hamilton-Jacobi-Equations
16.4.2 Policy Iteration for Nonzero Sum Differential Games
16.4.3 Online Solution for Two-Player Nonzero Sum Differential Games
16.4.4 Simulation

References
18.5.4 The Knowledge Gradient for a Parametric Belief Model 425
18.5.5 Discussion 426
18.6 Learning with a Physical State 427
18.6.1 Heuristic Policies 428
18.6.2 The Knowledge Gradient with a Physical State 428
References 429

19. An Introduction to Event-Based Optimization: Theory and Applications 432
Xi-Ren Cao, Yanjia Zhao, Qing-Shan Jia, and Qianchuan Zhao
19.1 Introduction 432
19.2 Literature Review 433
19.3 Problem Formulation 434
19.4 Policy Iteration for EBO 435
19.4.1 Performance Difference and Derivative Formulas 435
19.4.2 Policy Iteration for EBO 440
19.5 Example: Material Handling Problem 441
19.5.1 Problem Formulation 441
19.5.2 Event-Based Optimization for the Material Handling Problem 444
19.5.3 Numerical Results 446
19.6 Conclusions 448
References 449

20. Bounds for Markov Decision Processes 452
Vijay V. Desai, Vivek F. Farias, and Ciamac C. Moallemi
20.1 Introduction 452
20.1.1 Related Literature 454
20.2 Problem Formulation 455
20.3 The Linear Programming Approach 456
20.3.1 The Exact Linear Program 456
20.3.2 Cost-to-Go Function Approximation 457
20.3.3 The Approximate Linear Program 457
20.4 The Martingale Duality Approach 458
20.5 The Pathwise Optimization Method 461
20.6 Applications 463
20.6.1 Optimal Stopping 464
20.6.2 Linear Convex Control 467
20.7 Conclusion 470
References 471
CONTENTS

23.6 Conclusions 532
References 533

Dayu Huang, W. Chen, P. Mehta, S. Meyn, and A. Surana
24.1 Introduction 535
24.2 Optimality Equations 536
  24.2.1 Deterministic Model 537
  24.2.2 Diffusion Model 538
  24.2.3 Models in Discrete Time 539
  24.2.4 Approximations 539
24.3 Neuro-Dynamic Algorithms 542
  24.3.1 MDP Model 542
  24.3.2 TD-Learning 543
  24.3.3 SARSA 546
  24.3.4 Q-Learning 547
  24.3.5 Architecture 550
24.4 Fluid Models 551
  24.4.1 The CRW Queue 551
  24.4.2 Speed-Scaling Model 552
24.5 Diffusion Models 554
  24.5.1 The CRW Queue 555
  24.5.2 Speed-Scaling Model 556
24.6 Mean Field Games 556
24.7 Conclusions 557
References 558

25. Approximate Dynamic Programming for Optimizing Oil Production 560
Zheng Wen, Louis J. Durlofsky, Benjamin Van Roy, and Khalid Aziz
25.1 Introduction 560
25.2 Petroleum Reservoir Production Optimization Problem 562
25.3 Review of Dynamic Programming and Approximate Dynamic Programming 564
25.4 Approximate Dynamic Programming Algorithm for Reservoir Production Optimization 566
  25.4.1 Basis Function Construction 566
  25.4.2 Computation of Coefficients 568
  25.4.3 Solving Subproblems 570
  25.4.4 Adaptive Basis Function Selection and Bootstrapping 571
  25.4.5 Computational Requirements 572
25.5 Simulation Results 573
25.6 Concluding Remarks 578
References 580
26. A Learning Strategy for Source Tracking in Unstructured Environments 582
Titus Appel, Rafael Fierro, Brandon Rohrer, Ron Lumia, and John Wood
26.1 Introduction 582
26.2 Reinforcement Learning 583
   26.2.1 Q-Learning 584
   26.2.2 Q-Learning and Robotics 589
26.3 Light-Following Robot 589
26.4 Simulation Results 592
26.5 Experimental Results 595
   26.5.1 Hardware 596
   26.5.2 Problems in Hardware Implementation 597
   26.5.3 Results 598
26.6 Conclusions and Future Work 599
References 599

INDEX 601
Modern day society relies on the operation of complex systems including aircraft, automobiles, electric power systems, economic entities, business organizations, banking and finance systems, computer networks, manufacturing systems, and industrial processes. Decision and control are responsible for ensuring that these systems perform properly and meet prescribed performance objectives. The safe, reliable, and efficient control of these systems is essential for our society. Therefore, automatic decision and control systems are ubiquitous in human engineered systems and have had an enormous impact on our lives. As modern systems become more complex and performance requirements more stringent, improved methods of decision and control are required that deliver guaranteed performance and the satisfaction of prescribed goals.

Feedback control works on the principle of observing the actual outputs of a system, comparing them to desired trajectories, and computing a control signal based on that error, which is used to modify the performance of the system to make the actual output follow the desired trajectory. The optimization of sequential decisions or controls that are repeated over time arises in many fields, including artificial intelligence, automatic control systems, power systems, economics, medicine, operations research, resource allocation, collaboration and coalitions, business and finance, and games including chess and backgammon. Optimal control theory provides methods for computing feedback control systems that deliver optimal performance. Optimal controllers optimize user-prescribed performance functions and are normally designed offline by solving Hamilton–Jacobi–Bellman (HJB) design equations. This requires knowledge of the full system dynamics model. However, it is often difficult to determine an accurate dynamical model of practical systems. Moreover, determining optimal control policies for nonlinear systems requires the offline solution of nonlinear HJB equations, which are often difficult or impossible to solve. Dynamic programming (DP) is a sequential algorithmic method for finding optimal solutions in sequential decision problems. DP was developed beginning in the 1960s with the work of Bellman and Pontryagin. DP is fundamentally a backwards-in-time procedure that does not offer methods for solving optimal decision problems in a forward manner in real time.

The real-time adaptive learning of optimal controllers for complex unknown systems has been solved in nature. Every agent or system is concerned with acting on its environment in such a way as to achieve its goals. Agents seek to learn how to collaborate to improve their chances of survival and increase. The idea that there is
a cause and effect relation between actions and rewards is inherent in animal learning. Most organisms in nature act in an optimal fashion to conserve resources while achieving their goals. It is possible to study natural methods of learning and use them to develop computerized machine learning methods that solve sequential decision problems.

Reinforcement learning (RL) describes a family of machine learning systems that operate based on principles used in animals, social groups, and naturally occurring systems. RL methods were used by Ivan Pavlov in the 1860s to train his dogs. RL refers to an actor or agent that interacts with its environment and modifies its actions, or control policies, based on stimuli received in response to its actions. RL computational methods have been developed by the Computational Intelligence Community that solve optimal decision problems in real time and do not require the availability of analytical system models. The RL algorithms are constructed on the idea that successful control decisions should be remembered, by means of a reinforcement signal, such that they become more likely to be used another time. Successful collaborating groups should be reinforced. Although the idea originates from experimental animal learning, it has also been observed that RL has strong support from neurobiology, where it has been noted that the dopamine neurotransmitter in the basal ganglia acts as a reinforcement informational signal, which favors learning at the level of the neurons in the brain. RL techniques were first developed for Markov decision processes having finite state spaces. They have been extended for the control of dynamical systems with infinite state spaces.

One class of RL methods is based on the actor–critic structure, where an actor component applies an action or a control policy to the environment, whereas a critic component assesses the value of that action. Actor–critic structures are particularly well adapted for solving optimal decision problems in real time through reinforcement learning techniques. Approximate dynamic programming (ADP) refers to a family of practical actor–critic methods for finding optimal solutions in real time. These techniques use computational enhancements such as function approximation to develop practical algorithms for complex systems with disturbances and uncertain dynamics. Now, the ADP approach has become a key direction for future research in understanding brain intelligence and building intelligent systems.

The purpose of this book is to give an exposition of recently developed RL and ADP techniques for decision and control in human engineered systems. Included are both single-player decision and control and multiplayer games. RL is strongly connected from a theoretical point of view with both adaptive learning control and optimal control methods. There has been a great deal of interest in RL and recent work has shown that ideas based on ADP can be used to design a family of adaptive learning algorithms that converge in real-time to optimal control solutions by measuring data along the system trajectories. The study of RL and ADP requires methods from many fields, including computational intelligence, automatic control systems, Markov decision processes, stochastic games, psychology, operations research, cybernetics, neural networks, and neurobiology. Therefore, this book is interested in bringing together ideas from many communities.
This book has three parts. Part I develops methods for feedback control of systems based on RL and ADP. Part II treats learning and control in multiagent games. Part III presents some ideas of fundamental importance in understanding and implementing decision algorithm in Markov processes.

F.L. Lewis
Derong Liu

Fort Worth, TX
Chicago, IL
CONTRIBUTORS

Eduardo Alonso, School of Informatics, City University, London, UK

Charles W. Anderson, Department of Computer Science, Colorado State University, Fort Collins, CO, USA

Titus Appel, MARHES Lab, Department of Electrical & Computer Engineering, University of New Mexico, Albuquerque, NM, USA

Khalid Aziz, Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA

Robert Babuska, Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands

S.N. Balakrishnan, Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, MO, USA

Tamer Başar, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Dimitri Bertsekas, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA

Shubhendu Bhasin, Department of Electrical Engineering, Indian Institute of Technology, Delhi, India

Shalabh Bhatnagar, Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India

V.S. Borkar, Department of Electrical Engineering, Indian Institute of Technology, Powai, Mumbai, India

Lucian Busoniu, Université de Lorraine, CRAN, UMR 7039 and CNRS, CRAN, UMR 7039, Vandœuvre-lès-Nancy, France

Xi-Ren Cao, Shanghai Jiaotong University, Shanghai, China

W. Chen, Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Vijay Desai, Industrial Engineering and Operations Research, Columbia University, New York, NY, USA
Gianluca Di Muro, Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Jie Ding, Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, MO, USA

Warren E. Dixon, Department of Mechanical and Aerospace Engineering, University of Florida, FL, USA

Louis J. Durlofsky, Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA

Krishnamurthy Dvijotham, Computer Science and Engineering, University of Washington, Seattle, WA, USA

Michael Fairbank, School of Informatics, City University, London, UK

Vivek Farias, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA

Silvia Ferrari, Laboratory for Intelligent Systems and Control (LISC), Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

Rafael Fierro, MARHES Lab, Department of Electrical & Computer Engineering, University of New Mexico, Albuquerque, NM, USA

Haibo He, Department of Electrical, Computer and Biomedical Engineering, University of Rhode Island, Kingston, RI, USA

Ali Heydari, Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, MO, USA

Dayu Huang, Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

S. Jagannathan, Electrical and Computer Engineering Department, Missouri University of Science and Technology, Rolla, MI, USA

Qing-Shan Jia, Department of Automation, Tsinghua University, Beijing, China

Yu Jiang, Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn, NY, USA

Marcus Johnson, Department of Mechanical and Aerospace Engineering, University of Florida, FL, USA

Zhong-Ping Jia, Department of Electrical and Computer Engineering, Polytechnic Institute of New York University, Brooklyn, NY, USA

Rushikesh Kamalapurkar, Department of Mechanical and Aerospace Engineering, University of Florida, FL, USA

Kenton Kirkpatrick, Department of Aerospace Engineering, Texas A&M University, College Station, TX, USA
J. Nate Knight, Numerica Corporation, Loveland, CO, USA

F.L. Lewis, UTA Research Institute, University of Texas, Arlington, TX, USA

Derong Liu, State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China

Chao Lu, Department of Electrical Engineering, Tsinghua University, Beijing, P. R. China

Ron Lumia, Department of Mechanical Engineering, University of New Mexico, Albuquerque, NM, USA

P. Mehta, Coordinated Science Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Sean Meyn, Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA

Ciamac Moallemi, Graduate School of Business, Columbia University, New York, NY, USA

Remi Munos, SequeL team, INRIA Lille – Nord Europe, France

Zhen Ni, Department of Electrical, Computer and Biomedical Engineering, University of Rhode Island, Kingston, RI, USA

Warren B. Powell, Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ, USA

L.A. Prashanth, Department of Computer Science and Automation, Indian Institute of Science, Bangalore, India

Danil Prokhorov, Toyota Research Institute North America, Toyota Technical Center, Ann Arbor, MI, USA

Armando A. Rodriguez, School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA

Brandon Rohrer, Sandia National Laboratories, Albuquerque, NM, USA

Keith Rudd, Laboratory for Intelligent Systems and Control (LISC), Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC, USA

I.O. Ryzhov, Department of Decision, Operations and Information Technologies, Robert H. Smith School of Business, University of Maryland, College Park, MD, USA

John Seiffertt, Department of Electrical and Computer Engineering, Missouri University of Science & Technology, Rolla, MO, USA

Jennie Si, School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA

A. Surana, United Technologies Research Center, East Hartford, CT, USA
Hamidou Tembine, Telecommunication Department, Supelec, Gif sur Yvette, France

Emanuel Todorov, Applied Mathematics, Computer Science and Engineering, University of Washington, Seattle, WA, USA

Kostas S. Tsakalis, School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA

John Valasek, Department of Aerospace Engineering, Texas A&M University, College Station, TX, USA

K. Vamvoudaki, Center for Control, Dynamical-Systems and Computation, University of California, Santa Barbara, CA, USA

Benjamin Van Roy, Department of Management Science and Engineering and Department of Electrical Engineering, Stanford University, Stanford, CA, USA

Draguna Vrabie, United Technologies Research Center, East Hartford, CT, USA

Ding Wang, State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, P.R. China

Zheng Wen, Department of Electrical Engineering, Stanford University, Stanford, CA, USA

Paul Werbos, National Science Foundation, Arlington, VA, USA

John Wood, Department of Mechanical Engineering, University of New Mexico, Albuquerque, NM, USA

Don Wunsch, Department of Electrical and Computer Engineering, Missouri University of Science & Technology, Rolla, MO, USA

Lei Yang, College of Information and Control Science and Engineering, Zhejiang University, Hangzhou, China

Qinmin Yang, State Key Laboratory of Industrial Control Technology, Department of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang, China

Hassan Zargarzadeh, Embedded Systems and Networking Laboratory, Electrical and Computer Engineering Department, Missouri University of Science and Technology, Rolla, MI, USA

Dongbin Zhao, State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

Qianchuan Zhao, Department of Automation, Tsinghua University, Beijing, China

Yanjia Zhao, Department of Automation, Tsinghua University, Beijing, China

Quanyan Zhu, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, USA
PART I

FEEDBACK CONTROL USING RL AND ADP