Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles

Draguna Vrabie, Kyriakos G. Vamvoudakis and Frank L. Lewis

The Institution of Engineering and Technology
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td></td>
<td>xii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td></td>
<td>xv</td>
</tr>
<tr>
<td>1</td>
<td>Introduction to optimal control, adaptive control and reinforcement learning</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Optimal control</td>
<td>2</td>
</tr>
<tr>
<td>1.1.1</td>
<td>Linear quadratic regulator</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Linear quadratic zero-sum games</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Adaptive control</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Reinforcement learning</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>Optimal adaptive control</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Reinforcement learning and optimal control of discrete-time systems: Using natural decision methods to design optimal adaptive controllers</td>
<td>9</td>
</tr>
<tr>
<td>2.1</td>
<td>Markov decision processes</td>
<td>11</td>
</tr>
<tr>
<td>2.1.1</td>
<td>Optimal sequential decision problems</td>
<td>12</td>
</tr>
<tr>
<td>2.1.2</td>
<td>A backward recursion for the value</td>
<td>14</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Dynamic programming</td>
<td>15</td>
</tr>
<tr>
<td>2.1.4</td>
<td>Bellman equation and Bellman optimality equation</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>Policy evaluation and policy improvement</td>
<td>19</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Policy iteration</td>
<td>21</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Iterative policy iteration</td>
<td>21</td>
</tr>
<tr>
<td>2.2.3</td>
<td>Value iteration</td>
<td>22</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Generalized policy iteration</td>
<td>25</td>
</tr>
<tr>
<td>2.2.5</td>
<td>Q function</td>
<td>26</td>
</tr>
<tr>
<td>2.3</td>
<td>Methods for implementing policy iteration and value iteration</td>
<td>29</td>
</tr>
<tr>
<td>2.4</td>
<td>Temporal difference learning</td>
<td>30</td>
</tr>
<tr>
<td>2.5</td>
<td>Optimal adaptive control for discrete-time systems</td>
<td>32</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Policy iteration and value iteration for discrete-time dynamical systems</td>
<td>34</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Value function approximation</td>
<td>35</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Optimal adaptive control algorithms for discrete-time systems</td>
<td>36</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Introduction of a second ‘Actor’ neural network</td>
<td>38</td>
</tr>
<tr>
<td>2.5.5</td>
<td>Online solution of Lyapunov and Riccati equations</td>
<td>42</td>
</tr>
</tbody>
</table>
2.5.6 Actor-critic implementation of discrete-time optimal adaptive control 43
2.5.7 Q learning for optimal adaptive control 43
2.6 Reinforcement learning for continuous-time systems 46

PART I Optimal adaptive control using reinforcement learning structures 49

3 Optimal adaptive control using integral reinforcement learning for linear systems 51
3.1 Continuous-time adaptive critic solution for the linear quadratic regulator 53
3.1.1 Policy iteration algorithm using integral reinforcement 54
3.1.2 Proof of convergence 55
3.2 Online implementation of IRL adaptive optimal control 58
3.2.1 Adaptive online implementation of IRL algorithm 58
3.2.2 Structure of the adaptive IRL algorithm 61
3.3 Online IRL load-frequency controller design for a power system 64
3.4 Conclusion 69

4 Integral reinforcement learning (IRL) for non-linear continuous-time systems 71
4.1 Non-linear continuous-time optimal control 72
4.2 Integral reinforcement learning policy iterations 74
4.2.1 Integral reinforcement learning policy iteration algorithm 76
4.2.2 Convergence of IRL policy iteration 78
4.3 Implementation of IRL policy iterations using value function approximation 79
4.3.1 Value function approximation and temporal difference error 79
4.3.2 Convergence of approximate value function to solution of the Bellman equation 81
4.3.3 Convergence of approximate IRL policy iteration to solution of the HJB equation 85
4.4 Online IRL actor–critic algorithm for optimal adaptive control 85
4.4.1 Actor–critic structure for online implementation of adaptive optimal control algorithm 85
4.4.2 Relation of adaptive IRL control structure to learning mechanisms in the mammal brain 88
7.4 Action neural network and online synchronous policy iteration 136
7.5 Structure of adaptive controllers and synchronous optimal adaptive control 138
7.6 Simulations 142
7.6.1 Linear system example 142
7.6.2 Non-linear system example 143
7.7 Conclusion 147

8 Synchronous online learning with integral reinforcement 149
8.1 Optimal control and policy iteration using integral reinforcement learning 150
8.2 Critic neural network and Bellman equation solution 153
8.3 Action neural network and adaptive tuning laws 156
8.4 Simulations 158
8.4.1 Linear system 159
8.4.2 Non-linear system 161
8.5 Conclusion 164

PART III Online differential games using reinforcement learning 165

9 Synchronous online learning for zero-sum two-player games and H-infinity control 167
9.1 Two-player differential game and $H_\infty$ control 168
9.1.1 Two-player zero-sum differential games and Nash equilibrium 169
9.1.2 Application of zero-sum games to $H_\infty$ control 172
9.1.3 Linear quadratic zero-sum games 173
9.2 Policy iteration solution of the HJI equation 174
9.3 Actor–critic approximator structure for online policy iteration algorithm 176
9.3.1 Value function approximation and critic neural network 177
9.3.2 Tuning and convergence of the critic neural network 179
9.3.3 Action and disturbance neural networks 182
9.4 Online solution of two-player zero-sum games using neural networks 183
9.5 Simulations 187
9.5.1 Online solution of generalized ARE for linear quadratic ZS games 187
9.5.2 Online solution of HJI equation for non-linear ZS game 189
9.6 Conclusion 194

10 Synchronous online learning for multiplayer non-zero-sum games 195
10.1 $N$-player differential game for non-linear systems 196
10.1.1 Background on non-zero-sum games 196
10.1.2 Cooperation vs. non-cooperation in multiplayer dynamic games 199
10.2 Policy iteration solution for non–zero-sum games 199
10.3 Online solution for two-player non–zero-sum games 200
10.3.1 Value function approximation and critic neural networks for solution of Bellman equations 200
10.3.2 Action neural networks and online learning algorithm 204
10.4 Simulations 211
10.4.1 Non-linear system 211
10.4.2 Linear system 214
10.4.3 Zero-sum game with unstable linear system 215
10.5 Conclusion 218

11 **Integral reinforcement learning for zero-sum two-player games** 221
11.1 Zero-sum games for linear systems 223
11.1.1 Background 223
11.1.2 Offline algorithm to solve the game algebraic Riccati equation 224
11.1.3 Continuous-time HDP algorithm to solve Riccati equation 227
11.2 Online algorithm to solve the zero-sum differential game 229
11.3 Online load–frequency controller design for a power system 232
11.4 Conclusion 235

Appendix A: Proofs 237
References 273
Index 281