Learning-Based Control: A Tutorial and Some Recent Results

Other titles in Foundations and Trends® in Systems and Control

Achieving Ecological Resilience Through Regime Shift Management
M. D. Lemmon
ISBN: 978-1-68083-716-2

Distributed Optimization for Smart Cyber-Physical Networks
Giuseppe Notarstefano, Ivano Notarnicola and Andrea Camisa
ISBN: 978-1-68083-618-9

Resilient Control in Cyber-Physical Systems: Countering Uncertainty, Constraints, and Adversarial Behavior
Sean Weerakkody, Omur Ozel, Yilin Mo and Bruno Sinopoli
ISBN: 978-1-68083-586-1

Learning-Based Control: A Tutorial and Some Recent Results

Zhong-Ping Jiang
Tandon School of Engineering
New York University
New York, USA
zjiang@nyu.edu

Tao Bian
Bank of America
New York, USA
tbian@nyu.edu

Weinan Gao
College of Engineering and Science
Florida Institute of Technology
Florida, USA
wgao@fit.edu

the essence of knowledge
Boston — Delft

Foundations and Trends® in Systems and Control

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

Z.-P. Jiang, T. Bian and W. Gao. *Learning-Based Control: A Tutorial and Some Recent Results*. Foundations and Trends® in Systems and Control, vol. 8, no. 3, pp. 176–284, 2020.

ISBN: 978-1-68083-753-7

Foundations and Trends® in Systems and Control
Volume 8, Issue 3, 2021

Editorial Scope

Topics

Foundations and Trends® in Systems and Control publishes survey and tutorial articles in the following topics:

- Control of:
  - Hybrid and Discrete Event Systems
  - Nonlinear Systems
  - Network Systems
  - Stochastic Systems
  - Multi-agent Systems
  - Distributed Parameter Systems
  - Delay Systems
- Filtering, Estimation, Identification
- Optimal Control
- Systems Theory
- Control Applications

Information for Librarians

Foundations and Trends® in Systems and Control, 2021, Volume 8, 4 issues. ISSN paper version 2325-6818. ISSN online version 2325-6826. Also available as a combined paper and online subscription.

Contents

Learning-Based Control: A Tutorial and Some Recent Results

Zhong-Ping Jiang[1], Tao Bian[2] and Weinan Gao[3]

[1] *Tandon School of Engineering, New York University, New York, USA; zjiang@nyu.edu*

[2] *Bank of America, New York, USA; tbian@nyu.edu*

[3] *College of Engineering and Science, Florida Institute of Technology, Florida, USA; wgao@fit.edu*

ABSTRACT

This monograph presents a new framework for learning-based control synthesis of continuous-time dynamical systems with unknown dynamics. The new design paradigm proposed here is fundamentally different from traditional control theory. In the classical paradigm, controllers are often designed for a given class of dynamical control systems; it is a model-based design. Under the learning-based control framework, controllers are learned online from real-time input–output data collected along the trajectories of the control system in question. An entanglement of techniques from reinforcement learning and model-based control theory is advocated to find a sequence of suboptimal controllers that converge to the optimal solution as learning steps increase. On the one hand, this learning-based design approach attempts to overcome the well-known "curse of dimensionality" and the "curse of modeling" associated with Bellman's Dynamic Programming. On the other hand, rigorous stability and robustness analysis can be derived for

Zhong-Ping Jiang, Tao Bian and Weinan Gao (2020), "Learning-Based Control: A Tutorial and Some Recent Results", Foundations and Trends® in Systems and Control: Vol. 8, No. 3, pp 176–284. DOI: 10.1561/2600000023.

the closed-loop system with real-time learning-based controllers. The effectiveness of the proposed learning-based control framework is demonstrated via its applications to theoretical optimal control problems tied to various important classes of continuous-time dynamical systems and practical problems arising from biological motor control, connected and autonomous vehicles.

1

Introduction

The idea of learning-based control can be traced back at least to the Ph.D. dissertation (Minsky, 1954), where Minsky for the first time introduced the concept of reinforcement learning (RL) motivated by the problem of gaining further insight into the learning, memorizing, and thinking processes in human brain. Borrowing the words from Sutton *et al.* (1992), RL is direct adaptive optimal control. The field of RL is vibrant and is far from being saturated as clearly shown in numerous review articles and books (Bertsekas, 2011, 2013; Schmidhuber, 2015; Silver, 2015; Sutton and Barto, 2018; Szepesvári, 2010). Sixty years later after Minsky's original work, Google DeepMind developed perhaps one of the most advanced artificial intelligence (AI) system based on RL, and defeated the human world champion in the game of Go (Silver *et al.*, 2016, 2017). Indeed, besides Google DeepMind's AI system, RL has demonstrated its advantage in multiple industry applications (Barto *et al.*, 2017; Lorica, 2017). The recent success of RL and related methods can be attributed to several key factors. First, RL is driven by reward signals obtained through the interaction with the environment. Different from other machine learning (ML) techniques, this learning architecture is especially useful when the learning objective is to find the optimal behavior or policy over a time interval. Second, RL is closely related to the human learning behavior. It has been identified in a number of papers that the learning behavior in the frontal cortex and the basal ganglia is driven by the neuron spikes in dopamine neurons. These spikes encode the temporal difference error signal (Dayan and Balleine, 2002; Doya, 2002; Glimcher, 2011; Lo and Wang, 2006; Wang *et al.*, 2018; Wise, 2004), which is a key element in the RL theory (Sutton and Barto, 2018, Chapter 6). Hence, it is not surprising that we can achieve human-level intelligence through RL. Third, RL has a solid mathematical foundation. The main theoretical result behind RL is the dynamic programming (DP) theory (Bellman, 1957), which is a powerful tool for solving sequential decision making problems. The mathematical guarantee from DP theory gives the advantage of RL over other heuristic AI methods. Finally, RL can be incorporated with other ML and optimization methods to build a sophisticated learning system. For example, the learning performance of RL methods can be significantly improved by incorporating the recently developed deep neural network technique (Mnih *et al.*, 2015, 2016; Schmidhuber, 2015; Silver *et al.*, 2016, 2017). Because of these important features, RL and its extensions have become one of the most active research topics in AI and ML communities. Nonetheless, conventional RL theory exhibits some shortcomings. A common feature of most RL algorithms is that they are only applicable for discrete environments described by Markov decision processes (MDP) or discrete-time systems. To overcome this limitation, several researchers Baird, III (1993, 1994), Munos (2000), Doya (2000), Doya *et al.* (2002), van Hasselt and Wiering (2007), Theodorou *et al.* (2010), and van Hasselt (2012), have made significant efforts in adapting RL into the continuous environment, by discretizing and interpolating the time-state-action spaces. Alternatively, Bradtke and Duff (1994), Sutton *et al.* (1999), and Das *et al.* (1999) investigated RL for the semi-Markov process, a continuous-time dynamical system equipped with discrete state space. It should be mentioned that these methods may suffer from high computational burden when performing the discretization and approximation for continuous-time dynamical systems evolving in continuous state and action spaces. More recently, RL-based methods, mostly known under the name of adaptive dynamic programming (ADP), have been developed for continuous learning environments (Russell and Norvig, 2010, Chapter 2) described by ordinary differential equations (ODEs) or stochastic differential equations (SDEs). Another limitation of traditional RL methods is that the stability and robustness of the controlled process is usually not considered. In fact, a common assumption in the convergence analysis of various RL methods is that the underlying MDP always has a steady state distribution (Bhatnagar *et al.*, 2009; Nedić and Bertsekas, 2003; Sutton *et al.*, 2000; Tsitsiklis, 1994; Tsitsiklis and Van Roy, 1997). However, few results have been proposed to guarantee this assumption, especially when there exist policies under which the MDP does not have steady state distribution. In contrast with these limitations, experimental results have demonstrated that biological systems exhibit the ability of learning complicated motor movements in an unstable environment composed with high-dimensional continuous state space (Adams, 1971; Shadmehr and Mussa-Ivaldi, 2012; Wolpert *et al.*, 2011). Traditional RL theory is insufficient in explaining this type of learning process.

The purpose of this tutorial is to present a learning-based approach to control dynamical systems from real-time data and to review some major developments in this relatively young field. Due to space limitation, we will focus on continuous-time dynamical systems described by ODEs and SDEs. With input–output data at hand, we can certainly opt for the indirect route as in model-based control theory, that is, first build a mathematical model and then design controllers for the practical system in question. This indirect method has proven successful for a variety of problems arising in the contexts of engineering and sciences. However, it is widely known that building precise mathematical models that can describe the motion of dynamical systems is time-consuming and costly. For certain classes of optimal control problems, especially when the dynamical systems under consideration are strongly nonlinear, it is very hard, if not impossible, to solve the Bellman equation. This observation has led Bellman (1957) to state: “Turning to the succor of modern computing machines, let us renounce all analytic tools.” In this monograph, we aim to develop a framework for learning-based control theory that shows how to learn directly suboptimal controllers from input–output data. Ultimately, these suboptimal controllers are expected to converge to the (unknown) optimal solution to the Bellman equation. Besides the benefit of direct vs indirect control methods, the learning-based control theory overcomes the curse of modeling tied to the traditional DP. There are three main challenges on the development of learning-based control. First, there is a need to generalize existing recursive methods, known under the names of policy iteration (PI) and value iteration (VI), from model-based to data-driven contexts when the system dynamics are completely unknown. Previous RL-based learning algorithms are not directly extendable to the setting of continuous-time dynamical systems, let alone convergence and sensitivity analyses. Second, as a fundamental difference between learning-based control and RL, stability and robustness are important issues that must be addressed for the safety-critical engineering systems such as self-driving cars. Therefore, there is a need to develop new tools and methods, beyond the present literature of RL, that can provide theoretic guarantees on the stability and robustness of the controller learned from real-time data collected online along the trajectories of the control system under consideration. Third, data efficiency of RL algorithms need be addressed for safety-critical engineering systems. In this monograph, we will address the first two issues and only discuss the third issue from the perspective of numerical and experimental studies by means of some case studies. The learning-based control theory as reviewed in this monograph is closely tied to the literature of safe RL and ADP, and is a new direction in control theory that is still in its infancy and especially so for continuous-time dynamical systems described by differential equations. For prior work of others on ADP-based optimal control, the reader may consult (Jiang and Jiang, 2017; Lewis and Vrabie, 2009; Lewis *et al.*, 2012b; Liu *et al.*, 2017; Luo *et al.*, 2014; Song *et al.*, 2015; Vrabie *et al.*, 2013; Wang *et al.*, 2009; Werbos, 1968) and many references therein. For recent developments in learning-based control for other types of systems and problems, see Antsaklis *et al.* (1991), Antsaklis and Rahnama (2018), Rahnama and Antsaklis (2019), Werbos (2013, 2014, 2018), Kiumarsi *et al.* (2017), He and Zhong (2018), Recht (2019), Bertsekas (2019), Kamalapurkar *et al.* (2018), Chen *et al.* (2019), Pang *et al.* (2020), and references therein.

The rest of the monograph is organized as follows. Section 2 describes the learning-based optimal control of continuous-time linear and nonlinear systems described by (ordinary or stochastic) differential equations. Section 3 is concerned with the learning-based optimal control of a class of large-scale dynamical systems. Section 4 deals with the learning-based adaptive optimal tracking with disturbance rejection, the so-called adaptive optimal output regulation problem, for classes of linear and nonlinear control systems. Applications of the presented learning-based control theory to autonomous vehicles and human motor control are given in Section 5. Finally, some concluding remarks and discussions on future work are provided in Section 6.

References

Acerbi, L., S. Vijayakumar, and D. M. Wolpert (2014). "On the origins of suboptimality in human probabilistic inference". *PLOS Computational Biology*. 10(6): 1–23.

Adams, J. A. (1971). "A closed-loop theory of motor learning". *Journal of Motor Behavior*. 3(2): 111–150.

Anderson, B. D. O. and J. B. Moore (1989). *Optimal Control: Linear Quadratic Methods*. Englewood Cliffs, NJ: Prentice Hall International, Inc.

Antsaklis, P. J., K. M. Passino, and S. J. Wang (1991). "An introduction to autonomous control systems". *IEEE Control Systems Magazine*. 11(4): 5–13.

Antsaklis, P. J. and A. Rahnama (2018). "Control and machine intelligence for system autonomy". *Journal of Intelligent and Robotic Systems, 30th Year Anniversary Special Issue*. 91(1): 23–24.

Arapostathis, A., V. S. Borkar, and M. K. Ghosh (2012). *Ergodic Control of Diffusion Processes*. New York, NY: Cambridge University Press.

Arapostathis, A., V. S. Borkar, and K. Kumar (2014). "Convergence of the relative value iteration for the ergodic control problem of nondegenerate diffusions under near-monotone costs". *SIAM Journal on Control and Optimization*. 52(1): 1–31.

Arem, B. van, C. van Driel, and R. Visser (2006). "The impact of cooperative adaptive cruise control on traffic-flow characteristics". *IEEE Transactions on Intelligent Transportation Systems*. 7(4): 429–436.

Asadi, B. and A. Vahidi (2011). "Predictive cruise control: Utilizing upcoming traffic signal information for improving fuel economy and reducing trip time". *IEEE Transactions on Control Systems Technology*. 19(3): 707–714.

Åström, K. J. and B. Wittenmark (1997). *Adaptive Control*. 2nd Ed. Reading, MA: Addison-Wesley.

Bach, D. R. and R. J. Dolan (2012). "Knowing how much you don't know: A neural organization of uncertainty estimates". *Nature Reviews Neuroscience*. 13(8): 572–586.

Baird, III, L. C. (1993). "Advantage updating". *Tech. rep.* No. WL–TR-93-1146. Washington DC: Wright-Patterson Air Force Base Ohio: Wright Laboratory.

Baird, III, L. C. (1994). "Reinforcement learning in continuous time: Advantage updating". In: *Proceedings of the 1999 International Conference on Neural Networks*. Vol. 4. 2448–2453.

Banks, H. and K. Ito (1991). "A numerical algorithm for optimal feedback gains in high dimensional linear quadratic regulator problems". *SIAM Journal on Control and Optimization*. 29(3): 499–515.

Barto, A. G., P. S. Tomas, and R. S. Sutton (2017). "Some recent applications of reinforcement learning". In: *Proceedings of the Eighteenth Yale Workshop on Adaptive and Learning Systems*.

Beard, R. W. (1995). "Improving the closed-loop performance of nonlinear systems". *PhD thesis*. Rensselaer Polytechnic Institute.

Beard, R. W., G. N. Saridis, and J. T. Wen (1997). "Galerkin approximations of the generalized Hamilton–Jacobi–Bellman equation". *Automatica*. 33(12): 2159–2177.

Beck, J. M., W. J. Ma, X. Pitkow, P. E. Latham, and A. Pouget (2012). "Not noisy, just wrong: The role of suboptimal inference in behavioral variability". *Neuron*. 74(1): 30–39.

Beers, R. J. van, P. Baraduc, and D. M. Wolpert (2002). "Role of uncertainty in sensorimotor control". *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 357(1424): 1137–1145.

Bellman, R. E. (1954). "Dynamic programming and a new formalism in the calculus of variations". *Proceedings of the National Academy of Sciences of the United States of America*. 40(4): 231–235.

Bellman, R. E. (1957). *Dynamic Programming*. Princeton, NJ: Princeton University Press.

Benner, P. and R. Byers (1998). "An exact line search method for solving generalized continuous-time algebraic Riccati equations". *IEEE Transactions on Automatic Control*. 43(1): 101–107.

Benner, P., J.-R. Li, and T. Penzl (2008). "Numerical solution of large-scale Lyapunov equations, Riccati equations, and linear-quadratic optimal control problems". *Numerical Linear Algebra with Applications*. 15(9): 755–777.

Bensoussan, A. and J. Frehse (1992). "On Bellman equations of ergodic control in R^n". In: *Applied Stochastic Analysis*. Ed. by I. Karatzas and D. Ocone. Berlin, Heidelberg: Springer. 21–29.

Bertsekas, D. P. (2005). *Dynamic Programming and Optimal Control*. 3rd Ed. Vol. 1. Belmont, MA: Athena Scientific.

Bertsekas, D. P. (2007). *Dynamic Programming and Optimal Control*. 3rd Ed. Vol. 2. Belmont, MA: Athena Scientific.

Bertsekas, D. P. (2011). "Approximate policy iteration: A survey and some new methods". *Journal of Control Theory and Applications*. 9(3): 310–335.

Bertsekas, D. P. (2013). *Abstract Dynamic Programming*. Belmont, MA: Athena Scientific.

Bertsekas, D. P. (2017). "Value and policy iterations in optimal control and adaptive dynamic programming". *IEEE Transactions on Neural Networks and Learning Systems*. 28(3): 500–509.

Bertsekas, D. P. (2019). *Reinforcement Learning and Optimal Control*. Belmont, MA: Athena Scientific.

Bertsekas, D. P. and S. Ioffe (1996). "Temporal differences-based policy iteration and applications in neuro-dynamic programming". *Tech. rep.* No. Report LIDS-P-2349. Cambridge, MA: Lab. for Info. and Decision Systems, MIT.

Bhatnagar, S., R. S. Sutton, M. Ghavamzadeh, and M. Lee (2009). "Natural actor-critic algorithms". *Automatica*. 45(11): 2471–2482.

Bhushan, N. and R. Shadmehr (1999). "Computational nature of human adaptive control during learning of reaching movements in force fields". *Biological Cybernetics*. 81(1): 39–60.

Bian, T., Y. Jiang, and Z. P. Jiang (2014). "Adaptive dynamic programming and optimal control of nonlinear nonaffine systems". *Automatica*. 50(10): 2624–2632.

Bian, T., Y. Jiang, and Z. P. Jiang (2015). "Decentralized adaptive optimal control of large-scale systems with application to power systems". *IEEE Transactions on Industrial Electronics*. 62(4): 2439–2447.

Bian, T., Y. Jiang, and Z. P. Jiang (2016). "Adaptive dynamic programming for stochastic systems with state and control dependent noise". *IEEE Transactions on Automatic Control*. 61(12): 4170–4175.

Bian, T. and Z. P. Jiang (2016a). "Stochastic adaptive dynamic programming for robust optimal control design". In: *Control of Complex Systems: Theory and Applications*. Ed. by K. G. Vamvoudakis and S. Jagannathan. Cambridge, MA: Butterworth-Heinemann. Chap. 7. 211–245.

Bian, T. and Z. P. Jiang (2016b). "Value iteration and adaptive dynamic programming for data-driven adaptive optimal control design". *Automatica*. 71: 348–360.

Bian, T. and Z. P. Jiang (2016c). "Value iteration, adaptive dynamic programming, and optimal control of nonlinear systems". In: *Proceedings of the 55th IEEE Conference on Decision and Control (CDC)*. Las Vegas, USA. 3375–3380.

Bian, T. and Z. P. Jiang (2017). "A tool for the global stabilization of stochastic nonlinear systems". *IEEE Transactions on Automatic Control*. 62(4): 1946–1951.

Bian, T. and Z. P. Jiang (2018). "Stochastic and adaptive optimal control of uncertain interconnected systems: A data-driven approach". *Systems & Control Letters*. 115(5): 48–54.

Bian, T. and Z. P. Jiang (2019a). "Continuous-time robust dynamic programming". *SIAM Journal on Control and Optimization*. 57(6): 4150–4174.

Bian, T. and Z. P. Jiang (2019b). "Reinforcement learning for linear continuous-time systems: An incremental learning approach". *IEEE/CAA Journal of Automatica Sinica*. 6(2): 433–440.

Bian, T., D. Wolpert, and Z. P. Jiang (2020). "Model-free robust optimal feedback mechanisms of biological motor control". *Neural Computation*. 32(3): 562–595.

Borkar, V. S. (2006). "Ergodic control of diffusion processes". In: *Proceedings of the International Congress of Mathematicians*. 1299–1309.

Bradtke, S. J. and M. O. Duff (1994). "Reinforcement learning methods for continuous-time Markov decision problems". In: *Advances in Neural Information Processing Systems 7*. MIT Press. 393–400.

Bryson, J. A. E. (1994). *Control of Spaceraft and Aircraft*. Princeton, NJ: Princeton University Press.

Burdet, E., R. Osu, D. W. Franklin, T. E. Milner, and M. Kawato (2001). "The central nervous system stabilizes unstable dynamics by learning optimal impedance". *Nature*. 414(6862): 446–449.

Cashaback, J. G. A., H. R. McGregor, and P. L. Gribble (2015). "The human motor system alters its reaching movement plan for task-irrelevant, positional forces". *Journal of Neurophysiology*. 113(7): 2137–2149.

Chen, C., H. Modares, K. Xie, F. Lewis, Y. Wan, and S. Xie (2019). "Reinforcement learning-based adaptive optimal exponential tracking control of linear systems with unknown dynamics". *IEEE Transactions on Automatic Control*.

Christofides, P. D. (2001). *Nonlinear and Robust Control of PDE Systems: Methods and Applications to Transport-Reaction Processes*. New York: Springer Science+Business Media, LLC.

Clarke, F. (2013). *Functional Analysis, Calculus of Variations and Optimal Control*. London: Springer.

Dahlquist, G. and Å. Björck (1973). *Numerical Methods.* Englewood Cliffs, NJ: Prentice Hall.

Damm, T. (2004). *Rational Matrix Equations in Stochastic Control.* Berlin, Heidelberg: Springer.

Damm, T. and D. Hinrichsen (2001). "Newton's method for a rational matrix equation occurring in stochastic control". *Linear Algebra and its Applications.* 332–334: 81–109.

Das, T. K., A. Gosavi, S. Mahadevan, and N. Marchalleck (1999). "Solving semi-Markov decision problems using average reward reinforcement learning". *Management Science.* 45(4): 560–574.

Dayan, P. and B. W. Balleine (2002). "Reward, motivation, and reinforcement learning". *Neuron.* 36(2): 285–298.

Ding, Z. (2006). "Output regulation of uncertain nonlinear systems with nonlinear exosystems". *IEEE Transactions on Automatic Control.* 51(3): 498–503.

Ding, Z. (2013). "Consensus output regulation of a class of heterogeneous nonlinear systems". *IEEE Transactions on Automatic Control.* 58(10): 2648–2653.

Doya, K. (2000). "Reinforcement learning in continuous time and space". *Neural Computation.* 12(1): 219–245.

Doya, K. (2002). "Metalearning and neuromodulation". *Neural Networks.* 15(4–6): 495–506.

Doya, K., K. Samejima, K.-I. Katagiri, and M. Kawato (2002). "Multiple model-based reinforcement learning". *Neural Computation.* 14(6): 1347–1369.

Evans, L. C. (2005). "An introduction to mathematical optimal control theory". Lecture Notes, University of California, Department of Mathematics, Berkeley.

Flash, T. and N. Hogan (1985). "The coordination of arm movements: An experimentally confirmed mathematical model". *The Journal of Neuroscience.* 5(7): 1688–1703.

Fleming, W. H. and R. Rishel (1975). *Deterministic and Stochastic Optimal Control.* New York: Springer.

Francis, B. (1977). "The linear multivariable regulator problem". *SIAM Journal on Control and Optimization.* 15(3): 486–505.

Francis, B. A. and W. M. Wonham (1976). "The internal model principle of control theory". *Automatica*. 12(5): 457–465.

Franklin, D. W., E. Burdet, R. Osu, M. Kawato, and T. Milner (2003). "Functional significance of stiffness in adaptation of multijoint arm movements to stable and unstable dynamics". *Experimental Brain Research*. 151(2): 145–157.

Gao, W. and Z. P. Jiang (2016a). "Adaptive dynamic programming and adaptive optimal output regulation of linear systems". *IEEE Transactions on Automatic Control*. 61(12): 4164–4169.

Gao, W. and Z.-P. Jiang (2016b). "Nonlinear and adaptive suboptimal control of connected vehicles: A global adaptive dynamic programming approach". *Journal of Intelligent & Robotic Systems*: 1–15.

Gao, W. and Z. P. Jiang (2018). "Learning-based adaptive optimal tracking control of strict-feedback nonlinear systems". *IEEE Transactions on Neural Networks and Learning Systems*. 29(6): 2614–2624.

Gao, W., Z. P. Jiang, F. L. Lewis, and Y. Wang (2018). "Leader-to-formation stability of multiagent systems: An adaptive optimal control approach". *IEEE Transactions on Automatic Control*. 63(10): 3581–3587.

Gao, W., J. Gao, K. Ozbay, and Z. P. Jiang (2019a). "Reinforcement-learning-based cooperative adaptive cruise control of buses in the Lincoln tunnel corridor with time-varying topology". *IEEE Transactions on Intelligent Transportation Systems*. 20(10): 3796–3805.

Gao, W., Y. Jiang, and M. Davari (2019b). "Data-driven cooperative output regulation of multi-agent systems via robust adaptive dynamic programming". *IEEE Transactions on Circuits and Systems II: Express Briefs*. 66(3): 447–451.

Gao, W., A. Odekunle, Y. Chen, and Z.-P. Jiang (2019c). "Predictive cruise control of connected and autonomous vehicles via reinforcement learning". *IET Control Theory & Applications*. 13(7): 2849–2855.

Glimcher, P. W. (2011). "Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis". *Proceedings of the National Academy of Sciences*. 108(Supplement 3): 15647–15654.

Goebel, R., R. G. Sanfelice, and A. R. Teel (2012). *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press.

Guo, G. and W. Yue (2014). "Sampled-data cooperative adaptive cruise control of vehicles with sensor failures". *IEEE Transactions on Intelligent Transportation Systems*. 15(6): 2404–2418.

Haddad, W. M., V. Chellaboina, and S. G. Nersesov (2014). *Impulsive and Hybrid Dynamical Systems: Stability, Dissipativity, and Control*. Princeton University Press.

Haith, A. M. and J. W. Krakauer (2013). "Model-based and model-free mechanisms of human motor learning". In: *Progress in Motor Control*. Ed. by M. J. Richardson, M. A. Riley, and K. Shockley. Vol. 782. *Advances in Experimental Medicine and Biology*. New York: Springer. Chap. 1. 1–21.

Hale, J. K. and S. M. V. Lunel (1993). *Introduction to Functional Differential Equations*. New York: Springer-Verlag.

Harris, C. M. and D. M. Wolpert (1998). "Signal-dependent noise determines motor planning". *Nature*. 394(6695): 780–784.

Haruno, M. and D. M. Wolpert (2005). "Optimal control of redundant muscles in step-tracking wrist movements". *Journal of Neurophysiology*. 94(6): 4244–4255.

Haussmann, U. (1971). "Optimal stationary control with state control dependent noise". *SIAM Journal on Control*. 9(2): 184–198.

Haussmann, U. (1973). "Stability of linear systems with control dependent noise". *SIAM Journal on Control*. 11(2): 382–394.

He, H. and X. Zhong (2018). "Learning without external reward". *IEEE Computational Intelligence Magazine*. 13(3): 48–54.

Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. Cambridge, MA: The MIT Press.

Huang, J. (2004). *Nonlinear Output Regulation: Theory and Applications*. Philadelphia, PA: SIAM.

Huang, J. and Z. Chen (2004). "A general framework for tackling the output regulation problem". *IEEE Transactions on Automatic Control*. 49(12): 2203–2218.

Huang, M., W. Gao, Y. Wang, and Z. P. Jiang (2019). "Data-driven shared steering control of semi-autonomous vehicles". *IEEE Transactions on Human-Machine Systems*. 49(4): 350–361.

Huang, V. S., A. Haith, P. Mazzoni, and J. W. Krakauer (2011). "Rethinking motor learning and savings in adaptation paradigms: Model-free memory for successful actions combines with internal models". *Neuron*. 70(4): 787–801.

Ioannou, P. A. and C. C. Chien (1993). "Autonomous intelligent cruise control". *IEEE Transactions on Vehicular Technology*. 42(4): 657–672.

Isidori, A. and C. I. Byrnes (1990). "Output regulation of nonlinear systems". *IEEE Transactions on Automatic Control*. 35(2): 131–140.

Jiang, Y. and Z. P. Jiang (2011). "Approximate dynamic programming for optimal stationary control with control-dependent noise". *IEEE Transactions on Neural Networks*. 22(12): 2392–2398.

Jiang, Y. and Z. P. Jiang (2012a). "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics". *Automatica*. 48(10): 2699–2704.

Jiang, Y. and Z. P. Jiang (2012b). "Robust adaptive dynamic programming for large-scale systems with an application to multimachine power systems". *IEEE Transactions on Circuits and Systems II: Express Briefs*. 59(10): 693–697.

Jiang, Y. and Z. P. Jiang (2013a). "Robust adaptive dynamic programming with an application to power systems". *IEEE Transactions on Neural Networks and Learning Systems*. 24(7): 1150–1156.

Jiang, Y. and Z. P. Jiang (2014a). "Adaptive dynamic programming as a theory of sensorimotor control". *Biological Cybernetics*. 108(4): 459–473.

Jiang, Y. and Z. P. Jiang (2014b). "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems". *IEEE Trans. Neural Networks and Learning Syst.* 25(5): 882–893.

Jiang, Y. and Z. P. Jiang (2015a). "A robust adaptive dynamic programming principle for sensorimotor control with signal-dependent noise". *Journal of Systems Science and Complexity*. 28(2): 261–288.

Jiang, Y. and Z. P. Jiang (2015b). "Global adaptive dynamic programming for continuous-time nonlinear systems". *IEEE Transactions on Automatic Control*. 60(11): 2917–2929.

Jiang, Y. and Z. P. Jiang (2017). *Robust Adaptive Dynamic Programming*. Hoboken, NJ: Wiley-IEEE Press.

Jiang, Z. P. and I. M. Y. Mareels (1997). "A small-gain control method for nonlinear cascaded systems with dynamic uncertainties". *IEEE Transactions on Automatic Control*. 42(3): 292–308.

Jiang, Z. P. and Y. Jiang (2013b). "Robust adaptive dynamic programming for linear and nonlinear systems: An overview". *European Journal of Control*. 19(5): 417–425.

Jiang, Z. P. and T. Liu (2018). "Small-gain theory for stability and control of dynamical networks: A survey". *Annual Reviews in Control*. 46: 58–79.

Jiang, Z. P., A. R. Teel, and L. Praly (1994). "Small-gain theorem for ISS systems and applications". *Mathematics of Control, Signals and Systems*. 7(2): 95–120.

Jiang, Z. P., I. M. Mareels, and Y. Wang (1996). "A Lyapunov formulation of the nonlinear small-gain theorem for interconnected ISS systems". *Automatica*. 32(8): 1211–1215.

Johnson, C. D. (1971). "Accommodation of external disturbances in linear regulator and servomechanism problems". *IEEE Transactions on Automatic Control*. 16: 635–644.

Kalman, R. E. (1960). "Contributions to the theory of optimal control". *Boletin de la Sociedad Matematica Mexicana*. 5: 102–119.

Kamalapurkar, R., H. Dinh, S. Bhasin, and W. E. Dixon (2015). "Approximate optimal trajectory tracking for continuous-time nonlinear systems". *Automatica*. 51: 40–48.

Kamalapurkar, R., P. Walters, J. Rosenfeld, and W. E. Dixon (2018). *Reinforcement Learning for Optimal Feedback Control: A Lyapunov-Based Approach*. Springer.

Karafyllis, I. and Z. P. Jiang (2011). *Stability and Stabilization of Nonlinear Systems*. London: Springer.

Karafyllis, I. and M. Krstić (2018). *Input-to-State Stability for PDEs*. London: Springer.

Khas'minskii, R. (1967). "Necessary and sufficient conditions for the asymptotic stability of linear stochastic systems". *Theory of Probability & Its Applications*. 12(1): 144–147.

Khas'minskii, R. (2012). *Stochastic Stability of Differential Equations*. Berlin, Heidelberg: Springer.

Kiumarsi, B., K. G. Vamvoudakis, H. Modares, and F. L. Lewis (2017). "Optimal and autonomous control using reinforcement learning: A survey". *IEEE Transactions on Neural Networks and Learning Systems*. 29(6): 32–50.

Kleinman, D. L. (1968). "On an iterative technique for Riccati equation computations". *IEEE Transactions on Automatic Control*. 13(1): 114–115.

Kleinman, D. L. (1969). "Optimal stationary control of linear systems with control-dependent noise". *IEEE Transactions on Automatic Control*. 14(6): 673–677.

Kober, J., J. A. Bagnell, and J. Peters (2013). "Reinforcement learning in robotics: A survey". *The Int. J. Robotics Research*. 32(11): 1228–1274.

Kolm, P. N., R. Tütüncü, and F. J. Fabozzi (2014). "60 years of portfolio optimization: Practical challenges and current trends". *European Journal of Operational Research*. 234(2): 356–371.

Krener, A. J. (1992). "The construction of optimal linear and nonlinear regulators". In: *Systems, Models and Feedback: Theory and Applications*. Ed. by A. Isidori and T. J. Tarn. Vol. 12. Boston, Birkhauser. 301–322.

Krstić, M. (2009). *Delay Compensation for Nonlinear, Adaptive, and PDE Systems*. Boston: Birkhäuser.

Krstić, M., I. Kanellakopoulos, and P. V. Kokotovic (1995). *Nonlinear and Adaptive Control Design*. New York, NY: John Wiley & Sons, Inc.

Kučera, V. (1973). "A review of the matrix Riccati equation". *Kybernetika*. 9(1): 42–61.

Kushner, H. J. (1967). "Optimal discounted stochastic control for diffusion processes". *SIAM Journal on Control*. 5(4): 520–531.

Kushner, H. J. and G. G. Yin (2003). *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer.

Lancaster, P. and L. Rodman (1995). *Algebraic Riccati Equations*. New York: Oxford University Press.

Lanzon, A., Y. Feng, B. D. O. Anderson, and M. Rotkowitz (2008). "Computing the positive stabilizing solution to algebraic Riccati equations with an indefinite quadratic term via a recursive method". *IEEE Transactions on Automatic Control*. 53(10): 2280–2291.

Leake, R. and R. Liu (1967). "Construction of suboptimal control sequences". *SIAM Journal on Control*. 5(1): 54–63.

Lee, T. C. and Z. P. Jiang (2008). "Uniform asymptotic stability of nonlinear switched systems with an application to mobile robots". *IEEE Trans. Automatic Control*. 53(5): 1235–1252.

Lee, J. and B. Park (2012). "Development and evaluation of a cooperative vehicle intersection control algorithm under the connected vehicles environment". *IEEE Transactions on Intelligent Transportation Systems*. 13(1): 81–90.

Lewis, F. L., D. M. Dawson, and C. T. Abdallah (2004). *Robot Manipulator Control: Theory and Practice*. 2nd Ed. New York, NY: Marcel Dekker, Inc.

Lewis, F. L. and D. Vrabie (2009). "Reinforcement learning and adaptive dynamic programming for feedback control". *IEEE Circuits and Systems Magazine*. 9(3): 32–50.

Lewis, F. L., D. Vrabie, and V. L. Syrmos (2012a). *Optimal Control*. 3rd Ed. John Wiley & Sons, Inc.

Lewis, F. L., D. Vrabie, and K. G. Vamvoudakis (2012b). "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers". *IEEE Control Systems*. 32(6): 76–105.

Liberzon, D. (2003). *Switching in Systems and Control*. Birkhauser.

Liberzon, D. (2012). *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton, NJ: Princeton University Press.

Lisberger, S. G. and J. F. Medina (2015). "How and why neural and motor variation are related". *Current Opinion in Neurobiology*. 33: 110–116.

Liu, D. and E. Todorov (2007). "Evidence for the flexible sensorimotor strategies predicted by optimal feedback control". *The Journal of Neuroscience*. 27(35): 9354–9368.

Liu, L., Z. Chen, and J. Huang (2009). "Parameter convergence and minimal internal model with an adaptive output regulation problem". *Automatica*. 45(5): 1306–1311.

Liu, T., Z. P. Jiang, and D. J. Hill (2014). *Nonlinear Control of Dynamic Networks*. New York, NY: CRC Press.

Liu, D., Q. Wei, D. Wang, X. Yang, and H. Li (2017). *Adaptive Dynamic Programming with Applications in Optimal Control*. Springer International Publishing.

Lo, C.-C. and X.-J. Wang (2006). "Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks". *Nature Neuroscience*. 9(7): 956–963.

Lorica, B. (2017). *Practical Applications of Reinforcement Learning in Industry: An Overview of Commercial and Industrial Applications of Reinforcement Learning*.

Luo, B., H.-N. Wu, T. Huang, and D. Liu (2014). "Data-based approximate policy iteration for affine nonlinear continuous-time optimal control design". *Automatica*. 50(12): 3281–3290.

Luo, B., D. Liu, T. Huang, and D. Wang (2016). "Model-free optimal tracking control via critic-only Q-learning". *IEEE Transactions on Neural Networks and Learning Systems*. 27(10): 2134–2144.

Mahadevan, S. (1996). "Average reward reinforcement learning: Foundations, algorithms, and empirical results". *Machine Learning*. 22(1–3): 159–195.

Marino, R. and P. Tomei (2003). "Output regulation for linear systems via adaptive internal model". *IEEE Transactions on Automatic Control*. 48(12): 2199–2202.

Meyn, S. P. and R. L. Tweedie (1993a). "Stability of Markovian processes II: Continuous-time processes and sampled chains". *Advances in Applied Probability*. 25(3): 487–517.

Meyn, S. P. and R. L. Tweedie (1993b). "Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes". *Advances in Applied Probability*. 25(3): 518–548.

Minsky, M. L. (1954). "Theory of neural-analog reinforcement systems and its application to the brain model problem". *PhD thesis*. Princeton University.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis (2015). "Human-level control through deep reinforcement learning". *Nature*. 518(7540): 529–533.

Mnih, V., A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu (2016). "Asynchronous methods for deep reinforcement learning". In: *Proceedings of the 33rd International Conference on Machine Learning*. New York, New York, USA. 1928–1937.

Modares, H. and F. L. Lewis (2014). "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning". *IEEE Transactions on Automatic Control*. 59(11): 3051–3056.

Munos, R. (2000). "A study of reinforcement learning in the continuous case by the means of viscosity solutions". *Machine Learning*. 40(3): 265–299.

Murray, J. J., C. J. Cox, G. G. Lendaris, and R. Saeks (2002). "Adaptive dynamic programming". *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*. 32(2): 140–153.

Nedić, A. and D. P. Bertsekas (2003). "Least squares policy evaluation algorithms with linear function approximation". *Discrete Event Dynamic Systems*. 13(1–2): 79–110.

Ni, Z., H. He, and J. Wen (2013). "Adaptive learning in tracking control based on the dual critic network design". *IEEE Transactions on Neural Networks and Learning Systems*. 24(6): 913–928.

Odekunle, A., W. Gao, M. Davari, and Z. P. Jiang (2020). "Reinforcement learning and non-zero-sum game output regulation for multi-player linear uncertain systems". *Automatica*. 112(2): 108672.

Oncu, S., J. Ploeg, N. van de Wouw, and H. Nijmeijer (2014). "Cooperative adaptive cruise control: Network-aware analysis of string stability". *IEEE Transactions on Intelligent Transportation Systems*. 15(4): 1527–1537.

Pang, B. and Z. P. Jiang (2020). "Adaptive optimal control of linear periodic systems: An off-policy value iteration approach". *IEEE Trans. Automatic Control*. DOI: 10.1109/TAC.2020.2987313.

Pang, B., T. Bian, and Z. P. Jiang (2019). "Adaptive dynamic programming for finite-horizon optimal control of linear time-varying discrete-time systems". *Control Theory and Technology*. 17(1): 73–84.

Pang, B., Z. P. Jiang, and I. Mareels (2020). "Reinforcement learning for adaptive optimal control of continuous-time linear periodic systems". *Automatica*. 118: 109035.

Park, J. and I. W. Sandberg (1991). "Universal approximation using radial-basis-function networks". *Neural Computation*. 3(2): 246–257.

Pekny, S. E., J. Izawa, and R. Shadmehr (2015). "Reward-dependent modulation of movement variability". *The Journal of Neuroscience*. 35(9): 4015–4024.

Pliska, S. R. (1986). "A stochastic calculus model of continuous trading: Optimal portfolios". *Mathematics of Operations Research*. 11(2): 371–382.

Praly, L. and Y. Wang (1996). "Stabilization in spite of matched unmodeled dynamics and an equivalent definition of input-to-state stability". *Mathematics of Control, Signals and Systems*. 9(1): 1–33.

Puterman, M. L. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ: John Wiley & Sons, Inc.

Rahnama, A. and P. J. Antsaklis (2019). "Learning-based event-triggered control for synchronization of passive multi-agent systems under attack". *IEEE Transactions on Automatic Control*. 65(10): 4170–4185.

Recht, B. (2019). "A tour of reinforcement learning: The view from continuous control". *Annual Review of Control, Robotics, and Autonomous Systems*. 2(1): 253–279.

Renart, A. and C. K. Machens (2014). "Variability in neural activity and behavior". *Current Opinion in Neurobiology*. 25: 211–220.

Rizvi, S. A. A. and Z. Lin (2019). "Reinforcement learning-based linear quadratic regulation of continuous-time systems using dynamic output feedback". *IEEE Transactions on Cybernetics*. Accepted.

Robbins, H. and S. Monro (1951). "A stochastic approximation method". *The Annals of Mathematical Statistics*. 22(3): 400–407.

Ross, S. A. (1976). “The arbitrage theory of capital asset pricing”. *Journal of Economic Theory*. 13(3): 341–360.

Russell, S. J. and P. Norvig (2010). *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Pearson Education.

Salvucci, D. D. and R. Gray (2004). “A two-point visual control model of steering”. *Perception*. 33(10): 1233–1248.

Sandell, N. R. (1974). “On Newton’s method for Riccati equation solution”. *IEEE Transactions on Automatic Control*. 19(3): 254–255.

Saridis, G. N. and C.-S. G. Lee (1979). “An approximation theory of optimal control for trainable manipulators”. *IEEE Transactions on Systems, Man and Cybernetics*. 9(3): 152–159.

Schmidhuber, J. (2015). “Deep learning in neural networks: An overview”. *Neural Networks*. 61(Jan.): 85–117.

Schwartz, A. (1993). “A reinforcement learning method for maximizing undiscounted rewards”. In: *Proceedings of the 10th International Conference on Machine Learning*. Amherst, MA. 298–305.

Seiler, P., A. Pant, and K. Hedrick (2004). “Disturbance propagation in vehicle strings”. *IEEE Transactions on Automatic Control*. 49(10): 1835–1842.

Serrani, A., A. Isidori, and L. Marconi (2001). “Semiglobal nonlinear output regulation with adaptive internal model”. *IEEE Transactions on Automatic Control*. 46(8): 1178–1194.

Shadmehr, R. and F. A. Mussa-Ivaldi (1994). “Adaptive representation of dynamics during learning of a motor task”. *The Journal of Neuroscience*. 14(5): 3208–3224.

Shadmehr, R. and S. Mussa-Ivaldi (2012). *Biological Learning and Control: How the Brain Builds Representations, Predicts Events, and Makes Decisions*. Cambridge, MA: The MIT Press.

Shayman, M. (1986). “Phase portrait of the matrix Riccati equation”. *SIAM Journal on Control and Optimization*. 24(1): 1–65.

Shladover, S., D. Su, and X.-Y. Lu (2012). “Impacts of cooperative adaptive cruise control on freeway traffic flow”. *Transportation Research Record*. 2324: 63–70.

Silver, D. (2015). “Reinforcement Learning (COMPM050/COMPGI13)”. Lecture Notes, University College London, Computer Science Department, London.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis (2016). "Mastering the game of Go with deep neural networks and tree search". *Nature*. 529(7587): 484–489.

Silver, D., J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis (2017). "Mastering the game of Go without human knowledge". *Nature*. 550(7676): 354–359.

Song, R., F. L. Lewis, Q. Wei, H. Zhang, Z. P. Jiang, and D. Levine (2015). "Multiple actor-critic structures for continuous-time optimal control using input–output data". *IEEE Transactions on Neural Networks and Learning Systems*. 26(4): 851–865.

Sontag, E. D. (2008). "Input to state stability: Basic concepts and results". In: *Nonlinear and Optimal Control Theory*. Ed. by P. Nistri and G. Stefani. Berlin, Heidelberg: Springer. 163–220.

Spall, J. C. (2003). *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Hoboken, NJ: John Wiley & Sons, Inc.

Su, Y. and J. Huang (2012). "Cooperative output regulation of linear multi-agent systems". *IEEE Transactions on Automatic Control*. 57(4): 1062–1066.

Sutton, R. S. and A. G. Barto (2018). *Reinforcement Learning: An Introduction*. 2nd Ed. Cambridge, MA: The MIT Press.

Sutton, R. S., A. G. Barto, and R. J. Williams (1992). "Reinforcement learning is direct adaptive optimal control". *IEEE Control Systems*. 12(2): 19–22.

Sutton, R. S., D. Precup, and S. P. Singh (1999). "Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning". *Artificial Intelligence*. 112(1–2): 181–211.

Sutton, R. S., D. A. McAllester, S. P. Singh, and Y. Mansour (2000). "Policy gradient methods for reinforcement learning with function approximation". In: *Advances in Neural Information Processing Systems 12*. Vol. 12. MIT Press. 1057–1063.

Szepesvári, C. (2010). *Algorithms for Reinforcement Learning.* Morgan & Claypool Publishers.

Tanner, H. G., G. J. Pappas, and V. Kumar (2004). "Leader-to-formation stability". *IEEE Transactions on Robotics and Automation.* 20(3): 443–455.

Tao, G. (2003). *Adaptive Control Design and Analysis.* Hoboken, NJ: John Wiley & Sons, Inc.

Theodorou, E. A., J. Buchli, and S. Schaal (2010). "A generalized path integral control approach to reinforcement learning". *Journal of Machine Learning Research.* 11(Dec.): 3137–3181.

Thrun, S. B. (1992). "Efficient exploration in reinforcement learning". *Tech. rep.* No. CMU-CS-92-102. Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.

Todorov, E. (2004). "Optimality principles in sensorimotor control". *Nature Neuroscience.* 7(9): 907–915.

Todorov, E. (2005). "Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system". *Neural Computation.* 17(5): 1084–1108.

Todorov, E. and M. I. Jordan (2002). "Optimal feedback control as a theory of motor coordination". *Nature Neuroscience.* 5(11): 1226–1235.

Tsitsiklis, J. N. (1994). "Asynchronous stochastic approximation and Q-learning". *Machine Learning.* 16(3): 185–202.

Tsitsiklis, J. N. and B. Van Roy (1997). "An analysis of temporal-difference learning with function approximation". *IEEE Transactions on Automatic Control.* 42(5): 674–690.

Tsitsiklis, J. N. and B. Van Roy (1999). "Average cost temporal-difference learning". *Automatica.* 35(11): 1799–1808.

Tsitsiklis, J. N. and B. Van Roy (2002). "On average versus discounted reward temporal-difference learning". *Machine Learning.* 49(2–3): 179–191.

Uno, Y., M. Kawato, and R. Suzuki (1989). "Formation and control of optimal trajectory in human multijoint arm movement". *Biological Cybernetics.* 61(2): 89–101.

Vamvoudakis, K. G. (2017). "Q-learning for continuous-time linear systems: A model-free infinite horizon optimal control approach". *Systems & Control Letters*. 100: 14–20.

Vamvoudakis, K. G. and H. Ferraz (2018). "Model-free event-triggered control algorithms for continuous-time linear systems with optimal performance". *Systems & Control Letters*. 87: 412–420.

van Hasselt, H. (2012). "Reinforcement learning in continuous state and action spaces". In: *Reinforcement Learning: State-of-the-Art*. Ed. by M. Wiering and M. Otterlo. Vol. 12. *Adaptation, Learning, and Optimization*. Berlin, Heidelberg: Springer. Chap. 7. 207–251.

van Hasselt, H. and M. A. Wiering (2007). "Reinforcement learning in continuous action spaces". In: *IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning*. 272–279.

van der Schaft, A. J. (2017). *L_2-Gain and Passivity Techniques in Nonlinear Control*. 3rd Ed. Springer International Publishing.

van der Schaft, A. J. and H. Schumacher (2000). *An Introduction to Hybrid Dynamical Systems*. Springer.

Vaswani, P. A., L. Shmuelof, A. M. Haith, R. J. Delnicki, V. S. Huang, P. Mazzoni, R. Shadmehr, and J. W. Krakauer (2015). "Persistent residual errors in motor adaptation tasks: Reversion to baseline and exploratory escape". *The Journal of Neuroscience*. 35(17): 6969–6977.

Vrabie, D., O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis (2009). "Adaptive optimal control for continuous-time linear systems based on policy iteration". *Automatica*. 45(2): 477–484.

Vrabie, D., K. G. Vamvoudakis, and F. L. Lewis (2013). *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*. London, UK: Institution of Engineering and Technology.

Wang, D. and C. Mu (2019). *Adaptive Critic Control with Robust Stabilization for Uncertain Nonlinear Systems*. Singapore: Springer.

Wang, D., H. He, and D. Liu (2017). "Adaptive critic nonlinear robust control: A survey". *IEEE Transactions on Cybernetics*. 47(10): 3429–3451.

Wang, F.-Y., H. Zhang, and D. Liu (2009). "Adaptive dynamic programming: An introduction". *IEEE Computational Intelligence Magazine*. 4(2): 39–47.

Wang, J. X., Z. Kurth-Nelson, D. Kumaran, D. Tirumala, H. Soyer, J. Z. Leibo, D. Hassabis, and M. Botvinick (2018). "Prefrontal cortex as a meta-reinforcement learning system". *Nature Neuroscience*. 21(6): 860–868.

Wang, X., Y. Hong, J. Huang, and Z. P. Jiang (2010). "A distributed control approach to a robust output regulation problem for multi-agent linear systems". *IEEE Transactions on Automatic Control*. 55(12): 2891–2895.

Werbos, P. (1968). "The elements of intelligence". *Cybernetica (Namur)*. 11(3): 131.

Werbos, P. (2013). "Reinforcement learning and approximate dynamic programming (RLADP) – Foundations, common misconceptions and the challenges ahead". In: *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. Ed. by F. L. Lewis and D. Liu. Hoboken, NJ: Wiley. 3–30.

Werbos, P. J. (2014). "From ADP to the brain: Foundations, roadmap, challenges and research priorities". In: *Proceedings of the 2014 International Joint Conference on Neural Networks*. Beijing, China. 107–111.

Werbos, P. J. (2018). "AI intelligence for the grid 16 years later: Progress, challenges and lessons for other sectors". In: *Proceedings of the 2018 International Joint Conference on Neural Networks*. 1–8.

Willems, J. L. (1971). "Least squares stationary optimal control and the algebraic Riccati equation". *IEEE Transactions on Automatic Control*. 16(6): 621–634.

Wise, R. A. (2004). "Dopamine, learning and motivation". *Nature Reviews Neuroscience*. 5(June): 483–494.

Wolpert, D. M. and Z. Ghahramani (2000). "Computational principles of movement neuroscience". *Nature Neuroscience*. 3: 1212–1217.

Wolpert, D. M., J. Diedrichsen, and J. R. Flanagan (2011). "Principles of sensorimotor learning". *Nature Reviews Neuroscience*. 12(12): 739–751.

Wonham, W. (1967). "Optimal stationary control of a linear system with state-dependent noise". *SIAM Journal on Control*. 5(3): 486–500.

Wu, H. G., Y. R. Miyamoto, L. N. G. Castro, B. P. Olveczky, and M. A. Smith (2014). "Temporal structure of motor variability is dynamically regulated and predicts motor learning ability". *Nature Neuroscience*. 17(2): 312–321.

Yang, Y., L. Wang, H. Modares, D. Ding, Y. Yin, and D. Wunsch (2019). "Data-driven integral reinforcement learning for continuous-time non-zero-sum games". *IEEE Access*. 7(1): 82901–82912.

Yu, H. and D. P. Bertsekas (2009). "Convergence results for some temporal difference methods based on least squares". *IEEE Transactions on Automatic Control*. 54(7): 1515–1531.

Zakai, M. (1969). "A Lyapunov criterion for the existence of stationary probability distributions for systems perturbed by noise". *SIAM Journal on Control*. 7(3): 390–397.

Zhou, S.-H., J. Fong, V. Crocher, Y. Tan, D. Oetomo, and I. M. Y. Mareels (2016). "Learning control in robot-assisted rehabilitation of motor skills—A review". *Journal of Control and Decision*. 3(1): 19–43.