Synchronous Reinforcement Learning-Based Control for Cognitive Autonomy

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

Synchronous Reinforcement Learning-Based Control for Cognitive Autonomy

Kyriakos G. Vamvoudakis
Georgia Institute of Technology
USA
kyriakos@gatech.edu

Nick-Marios T. Kokolakis
Georgia Institute of Technology
USA
nmkokolakis@gatech.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

K. G. Vamvoudakis and N.-M. T. Kokolakis. *Synchronous Reinforcement Learning-Based Control for Cognitive Autonomy*. Foundations and Trends® in Systems and Control, vol. 8, no. 1–2, pp. 1–175, 2020.

ISBN: 978-1-68083-745-2

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

Synchronous Reinforcement Learning-Based Control for Cognitive Autonomy

Kyriakos G. Vamvoudakis[1] and Nick-Marios T. Kokolakis[2]

[1] *Georgia Institute of Technology, USA; kyriakos@gatech.edu*
[2] *Georgia Institute of Technology, USA; nmkokolakis@gatech.edu*

ABSTRACT

This monograph provides an exposition of recently developed reinforcement learning-based techniques for decision and control in human-engineered cognitive systems. The developed methods learn the solution to optimal control, zero-sum, non zero-sum, and graphical game problems completely online by using measured data along the system trajectories and have proved stability, optimality, and robustness. It is true that games have been shown to be important in robust control for disturbance rejection, and in coordinating activities among multiple agents in networked teams. We also consider cases with intermittent (an analogous to triggered control) instead of continuous learning and apply those techniques for optimal regulation and optimal tracking. We also introduce a bounded rational model to quantify the cognitive skills of a reinforcement learning agent. In order to do that, we leverage ideas from behavioral psychology to formulate differential games where the interacting learning agents have different intelligence skills, and we introduce an iterative method of optimal responses that determine the policy of an agent in adversarial environments. Finally, we present applications of reinforcement learning to motion planning and collaborative target tracking of bounded rational unmanned aerial vehicles.

Kyriakos G. Vamvoudakis and Nick-Marios T. Kokolakis (2020), “Synchronous Reinforcement Learning-Based Control for Cognitive Autonomy”, Foundations and Trends® in Systems and Control: Vol. 8, No. 1–2, pp 1–175. DOI: 10.1561/2600000022.

1

Introduction

1.1 A Unified Approach

This monograph describes the use of principles of reinforcement learning (RL) to design feedback policies for continuous-time dynamical systems that combine features of adaptive control and optimal control. *Adaptive control* (Ioannou and Fidan, 2006) and *optimal control* (Lewis *et al.*, 2012a) represent different philosophies for designing feedback controllers. These methods have been developed by the control systems community.

Optimal controllers minimize user-prescribed performance functions and are normally designed offline, i.e., performing all the calculations before being implemented into a system, by solving Hamilton–Jacobi–Bellman (HJB) equations, for example, the Riccati equation, using complete knowledge of the system dynamics. Determining optimal control policies for nonlinear systems requires the offline solution of nonlinear HJB equations.

Adaptive controllers learn online, i.e., process data and decide in real-time, to control unknown systems using data measured along the system trajectories. In fact, adaptive control is a powerful tool that uses online tuning of parameters to provide effective controllers for non-linear or linear systems with modeling uncertainties and disturbances.

Closed-loop stability while learning the parameters is guaranteed, often by using Lyapunov design techniques. Parameter convergence, however, often requires that the measured signals carry sufficient information about the unknown parameters known as a persistence of excitation (PE) condition, that is similar to exploration and exploitation in the learning terminology. Nevertheless, adaptive controllers are not usually designed to be optimal in the sense of minimizing user-prescribed performance functions. Indirect adaptive controllers use system identification techniques to first identify the system parameters and then use the obtained model to solve optimal design equations (Ioannou and Fidan, 2006). Adaptive controllers may satisfy certain inverse optimality conditions (Li and Krstic, 1997).

Several machine learning techniques have been employed for enabling adaptive autonomy (Vamvoudakis *et al.*, 2015). Machine learning is grouped, in supervised, unsupervised or *RL*, depending on the amount and quality of feedback about the system or task. In supervised learning, the feedback information provided to learning algorithms is a labeled training data set, and the objective is to build the system model representing the learned relation between the input, output and system parameters. In unsupervised learning, no feedback information is provided to the algorithm and the objective is to classify the sample sets to different groups based on the similarity between the input samples. Finally, RL, that is the subject of this monograph, is a goal-oriented learning tool wherein the agent, decision maker or controller learns a policy to optimize a long-term reward by interacting with the environment. At each step, an RL agent gets evaluative feedback about the performance of its action, allowing it to improve the performance of subsequent actions (Bertsekas and Tsitsiklis, 1996; Cao, 2007; Liu *et al.*, 2017; Sutton and Barto, 2018; Wiering and Van Otterlo, 2012).

In a control engineering context, *RL* bridges the gap between traditional optimal control and adaptive control algorithms (Bertsekas, 2019; Hovakimyan and Cao, 2010; Ioannou and Fidan, 2006; Jiang and Jiang, 2013; Kamalapurkar *et al.*, 2018; Krstić and Kanellakopoulos, 1995; Lewis *et al.*, 2012a,b; Tao, 2003; Zhang *et al.*, 2020). In our framework the goal is to learn the optimal policy and value function for a potentially uncertain physical system. Nevertheless, it is worth pointing out that the application of RL to the control discipline is not restricted solely in learning the optimal strategy and value function, but rather it is applicable in diverse applications such as system identification, adaptive control and even to the coordination of multi-agent systems (Hunt *et al.*, 1992; Mannor and Shamma, 2007; Poveda *et al.*, 2019; Sontag, 1993; Sontag and Sussmann, 1997; Wang and Hill, 2009). Unlike traditional optimal control, RL finds the solution to the HJB equation online. On the other hand, unlike traditional adaptive controllers, that are not usually designed to be optimal in the sense of minimizing cost functionals, RL algorithms are optimal. This has motivated control system researchers to enable adaptive and cognitive autonomy in an optimal manner by developing RL-based controllers. In continuous-time (CT) linear systems with multiple decision makers and quadratic costs, one has to rely on solving complicated matrix Riccati equations that require complete knowledge of the system matrices and need to be solved offline and then implemented online in the controller. In the era of complex and big data systems, modeling the processes exactly is most of the time infeasible and offline solutions make the systems vulnerable to parameter changes (drift).

Q-learning is a model-free action-dependent RL technique, i.e., does not require information about the environment, developed primarily for discrete-time systems (Watkins, 1989). It learns an action-dependent value function that ultimately gives the expected utility of taking a given action in a given state and following the optimal policy thereafter. When such an action-dependent value function is learned, the optimal policy can be computed easily. The *biggest strength* of Q-learning is that it is model-free. It has been proven in Watkins (1989) that for any finite Markov Decision Process, Q-learning eventually finds an optimal policy. In complex-systems Q-learning needs to store massive amounts of data, which makes the algorithm infeasible. This problem can be solved effectively by using adaptation techniques. Specifically, Q-learning can be improved by using the universal function approximation property that allow us to solve difficult optimization problems online and forward in time. This makes it possible to apply the algorithm to larger problems, even when the state space is continuous, and infinitely large.

Synchronous RL arises from a combination of techniques based on model-free and model-based RL. Specifically, RL techniques are used to design adaptive systems with novel structures that learn the solutions to optimization-based problems by observing data along the system trajectories. We term these as *optimal adaptive controllers*. These adaptive controllers are learned online and the policies converge to the optimal ones by tuning all parameters in all loops *simultaneously*, giving rise to *synchronous RL*. This is accomplished by developing two learning networks that interact with each other as they learn, and so mutually tune their parameters together *simultaneously* without any iterations. This learning mechanism is composed of an actor/critic structure, wherein there are two networks in two control loops – critic-network that evaluates the performance of current control policies and an actor-network that computes those current policies.

Game theory develops mathematical models allowing us to capture the strategic interaction among rational decision-makers/players (Başar and Olsder, 1999; Myerson, 2013). A rational agent can be thought of as an agent that has clear preferences, models uncertainty via expected values, and always chooses to perform the policy with the optimal expected outcome for itself from among all feasible actions. The solutions of several types of non-cooperative games (the cooperation among the agents is not allowed), namely the equilibrium strategies of the game, rely on the assumption of perfect rationality (Myerson, 2013). However, in real-world problems, the assumption of perfect rationality turns out to be quite strong and incapable of interpreting the actual behavior of the players (Crawford and Iriberri, 2007), thereby giving rise in bounded rationality (Simon, 1984) wherein the agents are bounded rational in the sense that the intelligence of the agents is limited by the information they have, the cognitive limitations of their minds, and the finite amount of time they have to make a decision. In the framework of RL, game theory is regraded as a bounded-rational interpretation of how equilibrium may result. Finally, based on the above, it follows that the *synchronous* RL can constitute a means for enabling online gaming by allowing the agents to learn their optimal policies online by measuring data along the players' trajectories, even when the environment is unknown or subject to changes.

1.2 RL and Cognitive Autonomy

Autonomy means having the freedom to act or function independently, i.e., self-government. Concerning the terminology of this term, it originally came from the Greek word "autonomia," which is a combination of the Greek words "auto" (self) and "nomy" (a system of rules). In the discipline of control engineering, this means that the agents can make a decision, namely to select a control policy, without involving a supervisor. Systems featuring these properties are the so-termed "Intelligent Autonomous Systems" (IAS), examples include Unmanned Aerial Vehicles (UAVs), Autonomous Underwater Vehicles (AUVs), office and residential buildings that regulate their energy consumption while adapting to the needs of their inhabitants (smart buildings), safety systems and environmentally friendly energy systems in automobiles (smart cars, smart highways) (Antsaklis *et al.*, 1991; Asama *et al.*, 2013; Vamvoudakis *et al.*, 2015). However, the IAS should be designed so that they are capable of dealing with the endogenous uncertainty imposing by the environment involving the presence of modeling uncertainties, the unavailability of the model, the possibility of cooperative along with non-cooperative goals, and malicious attacks compromising the security of teams of complex systems (Lamnabhi-Lagarrigue *et al.*, 2017). Nevertheless, it is evident that the *Synchronous RL* with the flexibility that it offers in tackling uncertainty, it has facilitated the evolution of cognitive autonomy aiming towards building fully autonomous IAS that are highly cognitive, reflective, multitask-able, and effective in knowledge discovery without external intervention. Ideally, moving towards full autonomy, the control engineering community desires to construct IAS, which should perhaps have the ability to perform even hardware repair if any of their components fails.

In general, there is a need for approaches that respond to situations not programmed or anticipated in the design. Therefore, by leveraging ideas from the recent advances of *Synchronous RL* and game theory, we bring together and combine interdisciplinary ideas from different fields as pictorially illustrated in Figure 1.1, i.e., computational intelligence, game theory, control theory, and information theory to endow IAS with novel cognitive learning algorithms intending to ensuring full autonomy

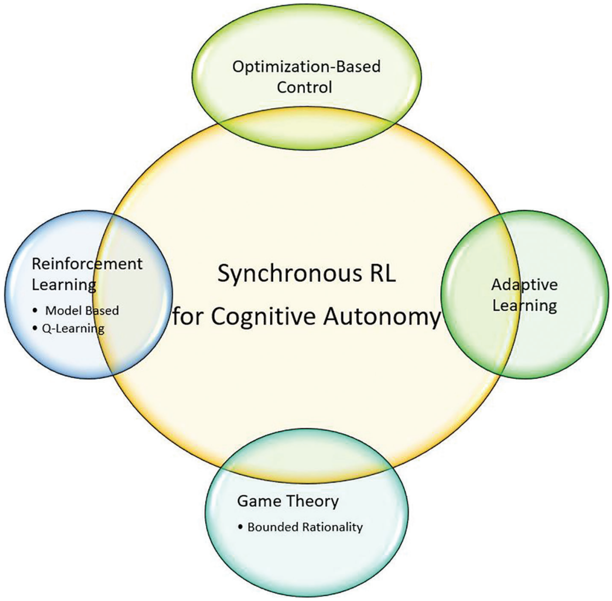

Figure 1.1: The *Synchronous RL*-based framework for enabling cognitive autonomy arises from the intersection of several diverse fields including, optimization-based control, adaptive learning, game theory, and RL.

and secure operation. Exploiting the adaptive nature of *Synchronous RL*, we apply the ideas of synchronous RL to kinodynamic motion planning algorithms that enable IAS to navigate securely and explore an unknown, challenging, environment with obstacles while guaranteeing the avoidance of collision with them. Furthermore, in the aerospace community is of profound importance to develop algorithms that will enable the coordination of autonomous swarms of UAVs to apprehend malicious vehicles that enter a protected zone, a phenomenon that has already been observed. To address that problem, we enforce "geofencing" protocols by constructing *cognitive* hierarchy-based algorithms inspired by the human brain, to coordinate a team of bounded rational UAVs for tracking an intelligent invading moving target. Finally, from the aforementioned, it is obvious that the *Synchronous RL*-based algorithms are featured by strong abilities of learning, and thus, the complex systems will be fully autonomous and tolerant to failures.

In this monograph we present a family of model-free, and model-based online adaptive learning algorithms for single and multi-agent systems using measurements along the system trajectories with continuous and intermittent feedback. The algorithms developed here are based on *Synchronous RL* principles, and rely on actor/critic-network schemes involving simultaneous tuning of the actor/critic neural networks (NNs) while providing online solutions to complex Hamilton–Jacobi (HJ) equations. However, it is worth mentioning that several of these techniques can be implemented without knowing the complete system dynamics, enabling cognitive autonomy.

1.3 Organization

The remainder of this monograph is structured as follows. Section 2 presents an adaptive method based on actor/critic RL for solving online the optimal control problem for deterministic CT input-affine nonlinear systems with known or partially unknown dynamics as well as with saturating and non-saturating actuators. In Section 3, under the assumption of perfect rationality, we develop adaptive controllers that learn optimal solutions for several differential game theory problems, including zero-sum, multi-player non-zero-sum, as well as graphical games. In the sequel, Section 4 proposes online Q-learning algorithms for solving the optimal control problem of a system with completely uncertain/unknown dynamics and shows its applications to differential game theory. Model-free and model-based intermittent control algorithms are displayed in Section 5 using ideas from RL. Next, by relaxing the assumption of perfect rationality, Section 6 introduces the non-equilibrium differential game theory and demonstrates its applications to cyber-physical systems security (CPS). Section 7 applies synchronous RL-based decision-making algorithms to motion planning in robotics as well as to coordinated target tracking using a team of bounded rational UAVs. Finally, Section 8 provides concluding remarks and potential future research perspectives on the area of synchronous RL-based control for cognitive autonomy.

Moreover, it is worth mentioning that throughout the monograph, we omit to include the proofs of the theorems as well as simulation results to avoid breaking the flow of the document. Nevertheless, we refer the reader to particular references wherein there are complete proofs, and simulation results verifying the efficiency of the presented control algorithms. Last but not least, note that instead of having a "centralized" literature review in this introductory section, and in following with the spirit of this monograph, we adopt a "distributed" literature review approach, where each section itself contains a review of the references that are relevant to the particular section content.

1.4 Notation

The notation used here is standard. $\mathbb{R}_+$ is the set of positive real numbers. $\|\cdot\|$ denotes the Euclidean norm of a vector. The superscript $\star$ is used to denote the optimal solution of an optimization problem, $\underline{\lambda}(A)$ is the minimum eigenvalue of a matrix A, $\bar{\lambda}(A)$ is the maximum eigenvalue of a matrix A, $\mathrm{tr}(A)$ is the trace of a matrix A, and $\mathbf{1}_m$ is the column vector with m ones. The gradient of a scalar-valued function with respect to a vector-valued variable x is defined as a column vector, and is denoted by $\nabla \coloneqq \partial/\partial x$. The $\mathrm{vec}(A)$ and the $\mathrm{vech}(A)$ denote the vectorization and the half-vectorization of a symmetric $n \times n$ matrix A, respectively. The notations $\overline{K}$, $|K|$, and ∂K denote the closure, the cardinality, and the limit points of the set K, respectively. The $U \otimes V$ denotes the Kronecker product of two vectors. The $\oplus$ is the Minkowski sum of two sets.

References

Abouheaf, M. I., F. L. Lewis, K. G. Vamvoudakis, S. Haesaert, and R. Babuska (2014). "Multi-agent discrete-time graphical games and reinforcement learning solutions". *Automatica*. 50(12): 3038–3053.

Abou-Kandil, H. and P. Bertrand (1985). "Analytical solution for an open-loop Stackelberg game". *IEEE Transactions on Automatic Control*. 30(12): 1222–1224.

Abu-Khalaf, M. and F. L. Lewis (2005). "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach". *Automatica*. 41(5): 779–791.

Abu Khalaf, M., F. L. Lewis, A. Al Tamimi, and D. Vrabie (2006). "Model-free adaptive dynamic programming for unknown systems". In: *Proceedings of the First International Conference on Computer Science & Education ICCSE'2006*. 105–114.

Abuzainab, N., W. Saad, and H. V. Poor (2016). "Cognitive hierarchy theory for heterogeneous uplink multiple access in the Internet of things". In: *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 1252–1256.

Alpcan, T. and T. Başar (2010). *Network Security: A Decision and Game-Theoretic Approach*. Cambridge University Press.

Al-Tamimi, A., F. L. Lewis, and M. Abu-Khalaf (2008). "Discrete-time nonlinear HJB solution using approximate dynamic programming: Convergence proof". *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 38(4): 943–949.

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.

Arthur, W. B. (2018). *The Economy as an Evolving Complex System II*. CRC Press.

Asama, H., T. Fukuda, T. Arai, and I. Endo (2013). *Distributed Autonomous Robotic Systems 2*. Springer Science & Business Media.

Bagchi, A. and T. Başar (1981). "Stackelberg strategies in linear-quadratic stochastic differential games". *Journal of Optimization Theory and Applications*. 35(3): 443–464.

Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Başar, T. and G. Olsder (1999). *Dynamic Noncooperative Game Theory*. 2nd edn. Philadelphia, PA: SIAM.

Başar, T. and P. Bernhard (2008). *H-Infinity Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Springer.

Beck, C. L. and R. Srikant (2012). "Error bounds for constant step-size Q-learning". *Systems & Control Letters*. 61(12): 1203–1208.

Benosman, M. (2018). "Model-based vs data-driven adaptive control: An overview". *International Journal of Adaptive Control and Signal Processing*. 32(5): 753–776.

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

Bertsekas, D. P. and J. N. Tsitsiklis (1995). "Neuro-dynamic programming: An overview". In: *Proceedings of the 34th IEEE Conference on Decision and Control*. Vol. 1. IEEE. 560–564.

Bertsekas, D. P. and J. N. Tsitsiklis (1996). *Neuro-Dynamic Programming*. Athena Scientific, MA.

Bhasin, S., R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon (2013). "A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems". *Automatica*. 49(1): 89–92.

Brams, S. and D. M. Kilgour (1988). *Game Theory and National Security*. New York: Basil Blackwell.

Bryson, A. E. (2018). *Applied Optimal Control: Optimization, Estimation and Control*. Routledge.

Busoniu, L., R. Babuska, and B. De Schutter (2008). "A comprehensive survey of multiagent reinforcement learning". *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 38(2): 156–172.

Busoniu, L., R. Babuska, B. De Schutter, and D. Ernst (2010). *Reinforcement Learning and Dynamic Programming Using Function Approximators*. Vol. 39. CRC Press.

Camerer, C. F. (2011). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004). "A cognitive hierarchy model of games". *The Quarterly Journal of Economics*. 119(3): 861–898.

Cao, M. (2020). "Merging game theory and control theory in the era of AI and autonomy". *National Science Review*. 7(7): 1122–1124.

Cao, X.-R. (2007). *Stochastic Learning and Optimization*. US: Springer.

Chen, C. and J. Cruz (1972). "Stackelburg solution for two-person games with biased information patterns". *IEEE Transactions on Automatic Control*. 17(6): 791–798.

Chen, T., X. Liu, and W. Lu (2007). "Pinning complex networks by a single controller". *IEEE Transactions on Circuits and Systems I: Regular Papers*. 54(6): 1317–1326.

Chong, J.-K., T.-H. Ho, and C. Camerer (2016). "A generalized cognitive hierarchy model of games". *Games and Economic Behavior*. 99: 257–274.

Chowdhary, G. and E. Johnson (2010). "Concurrent learning for convergence in adaptive control without persistency of excitation". In: *2010 49th IEEE Conference on Decision and Control (CDC)*. IEEE. 3674–3679.

Crawford, V. P. and N. Iriberri (2007). "Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions?" *Econometrica*. 75(6): 1721–1770.

Daskalakis, C., P. W. Goldberg, and C. H. Papadimitriou (2009). "The complexity of computing a Nash equilibrium". *SIAM Journal on Computing*. 39(1): 195–259.

Devraj, A. M., A. Bušić, and S. Meyn (2019). "Zap Q-learning – A user's guide". In: *2019 Fifth Indian Control Conference (ICC)*. 10–15.

Dierks, T. and S. Jagannathan (2010). "Optimal control of affine nonlinear continuous-time systems using an online Hamilton–Jacobi–Isaacs formulation". In: *49th IEEE Conference on Decision and Control (CDC)*. IEEE. 3048–3053.

Dong, L., Y. Tang, H. He, and C. Sun (2016a). "An event-triggered approach for load frequency control with supplementary ADP". *IEEE Transactions on Power Systems*. 32(1): 581–589.

Dong, L., X. Zhong, C. Sun, and H. He (2016b). "Event-triggered adaptive dynamic programming for continuous-time systems with control constraints". *IEEE Transactions on Neural Networks and Learning Systems*. 28(8): 1941–1952.

Donkers, M. and W. Heemels (2012). "Output-based event-triggered control with guaranteed-gain and improved and decentralized event-triggering". *IEEE Transactions on Automatic Control*. 57(6): 1362–1376.

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

Engwerda, J. (2005). *LQ Dynamic Optimization and Differential Games*. John Wiley & Sons.

Erev, I. and A. E. Roth (1998). "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria". *American Economic Review*. 88(4): 848–881.

Euwe, M., M. Blaine, and J. Rumble (1982). *The Logical Approach to Chess*. Courier Corporation.

Fax, J. A. and R. M. Murray (2004). "Information flow and cooperative control of vehicle formations". *IEEE Transactions on Automatic Control*. 49(9): 1465–1476.

Fisk, C. (1984). “Game theory and transportation systems modelling”. *Transportation Research Part B: Methodological*. 18(4–5): 301–313.

Fleming, W. H. and W. M. McEneaney (2000). “A max-plus-based algorithm for a Hamilton–Jacobi–Bellman equation of nonlinear filtering”. *SIAM Journal on Control and Optimization*. 38(3): 683–710.

Freiling, G., G. Jank, and H. Abou-Kandil (1996). “On global existence of solutions to coupled matrix Riccati equations in closed-loop Nash games”. *IEEE Transactions on Automatic Control*. 41(2): 264–269.

Freiling, G., G. Jank, and S. Lee (2001). “Existence and uniqueness of open-loop Stackelberg equilibria in linear-quadratic differential games”. *Journal of Optimization Theory and Applications*. 110(3): 515–544.

Freiling, G., G. Jank, and D. Kremer (2003). “Solvability condition for a nonsymmetric Riccati equation appearing in stackelberg games”. In: *2003 European Control Conference (ECC)*. IEEE. 963–967.

Fudenberg, D., F. Drew, D. K. Levine, and D. K. Levine (1998). *The Theory of Learning in Games*. Vol. 2. MIT Press.

Gajic, Z. and T. Li (1988). “Simulation results for two new algorithms for solving coupled algebraic Riccati equations”. In: *Third Int. Symp. on Differential Games. Sophia, Antipolis, France*.

Gao, W., Y. Jiang, Z.-P. Jiang, and T. Chai (2016). “Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming”. *Automatica*. 72: 37–45.

Garcia, E. and P. J. Antsaklis (2013). “Model-based event-triggered control for systems with quantization and time-varying network delays”. *IEEE Transactions on Automatic Control*. 58(2): 422–434.

Goretkin, G., A. Perez, R. Platt, and G. Konidaris (2013). “Optimal sampling-based planning for linear-quadratic kinodynamic systems”. In: *IEEE International Conference on Robotics and Automation*. 2429–2436.

Grondman, I., L. Busoniu, G. A. Lopes, and R. Babuska (2012). “A survey of actor-critic reinforcement learning: Standard and natural policy gradients”. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 42(6): 1291–1307.

Haddad, W. M., V. Chellaboina, and S. G. Nersesov (2006). *Impulsive and Hybrid Dynamical Systems: Stability, Dissipativity, and Control.* Princeton University Press.

Haykin, S. S. (2009). *Neural Networks and Learning Machines.* Vol. 3. Pearson Upper Saddle River.

He, X., A. Prasad, S. P. Sethi, and G. J. Gutierrez (2007). "A survey of Stackelberg differential game models in supply and marketing channels". *Journal of Systems Science and Systems Engineering.* 16(4): 385–413.

He, X., H. Dai, and P. Ning (2016). "Faster learning and adaptation in security games by exploiting information asymmetry". *IEEE Transactions on Signal Processing.* 64(13): 3429–3443.

Heemels, W., J. Sandee, and P. Van Den Bosch (2008). "Analysis of event-driven controllers for linear systems". *International Journal of Control.* 81(4): 571–590.

Hespanha, J. P. (2017). *Noncooperative Game Theory: An Introduction for Engineers and Computer Scientists.* Princeton University Press.

Hespanha, J. P., D. Liberzon, and A. R. Teel (2008). "Lyapunov conditions for input-to-state stability of impulsive systems". *Automatica.* 44(11): 2735–2744.

Ho, T.-H. and X. Su (2010). "A dynamic level-k model in games". *Tech. rep.* Haas School of Business, University of California, Berkeley.

Hornik, K., M. Stinchcombe, and H. White (1989). "Multilayer feedforward networks are universal approximators". *Neural Networks.* 2(5): 359–366.

Hornik, K., M. Stinchcombe, and H. White (1990). "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks". *Neural Networks.* 3(5): 551–560.

Hovakimyan, N. and C. Cao (2010). *L1 Adaptive Control Theory: Guaranteed Robustness with Fast Adaptation.* Philadelphia, PA: Advances in Design, Control, Society for Industrial, and Applied Mathematics (SIAM).

Hunt, K. J., D. Sbarbaro, R. Żbikowski, and P. J. Gawthrop (1992). "Neural networks for control systems—A survey". *Automatica.* 28(6): 1083–1112.

Ioannou, P. and B. Fidan (2006). “Adaptive control tutorial, society for industrial and applied mathematics”. In: *Advances in Design and Control*. PA: SIAM.

Ioannou, P. A. and J. Sun (2012). *Robust Adaptive Control*. Courier Corporation.

Jadbabaie, A., J. Lin, and A. S. Morse (2003). “Coordination of groups of mobile autonomous agents using nearest neighbor rules”. *IEEE Transactions on Automatic Control*. 48(6): 988–1001.

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

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

Johnson, M., T. Hiramatsu, N. Fitz-Coy, and W. E. Dixon (2010). “Asymptotic Stackelberg optimal control design for an uncertain Euler Lagrange system”. In: *49th IEEE Conference on Decision and Control (CDC)*. IEEE. 6686–6691.

Johnson, M., R. Kamalapurkar, S. Bhasin, and W. E. Dixon (2015). “Approximate N-player nonzero-sum game solution for an uncertain continuous nonlinear system”. *IEEE Transactions on Neural Networks and Learning Systems*. 26(8): 1645–1658.

Kaelbling, L. P., M. L. Littman, and A. W. Moore (1996). “Reinforcement learning: A survey”. *Journal of Artificial Intelligence Research*. 4(1): 237–285.

Kamalapurkar, R., P. Walters, J. Rosenfeld, and W. Dixon (2018). *Reinforcement Learning for Optimal Feedback Control*. Springer.

Kanellopoulos, A. and K. G. Vamvoudakis (2019). “Non-equilibrium dynamic games and cyber–physical security: A cognitive hierarchy approach”. *Systems & Control Letters*. 125: 59–66.

Kearns, M. (2007). “Graphical games”. In: *Algorithmic Game Theory*. Ed. by N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani. Vol. 3. Cambridge, UK: Cambridge University Press. 159–180.

Kiumarsi, B., F. L. Lewis, H. Modares, A. Karimpour, and M.-B. Naghibi-Sistani (2014). "Reinforcement Q-learning for optimal tracking control of linear discrete-time systems with unknown dynamics". *Automatica*. 50(4): 1167–1175.

Kiumarsi, B., F. L. Lewis, and Z. P. Jiang (2017). "H_∞ control of linear discrete-time systems: Off-policy reinforcement learning". *Automatica*. 78: 144–152.

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

Kokolakis, N.-M. T. and N. T. Koussoulas (2018). "Coordinated stand-off tracking of a ground moving target and the phase separation problem". In: *2018 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE. 473–482.

Kokolakis, N.-M. T., A. Kanellopoulos, and K. G. Vamvoudakis (2020). "Bounded rational unmanned aerial vehicle coordination for adversarial target tracking". In: *2020 American Control Conference (ACC)*. IEEE. 2508–2513.

Kontoudis, G. P. and K. G. Vamvoudakis (2019). "Kinodynamic motion planning with continuous-time Q-learning: An online, model-free, and safe navigation framework". *IEEE Transactions on Neural Networks and Learning Systems*. 30(12): 3803–3817.

Krstić, M. and I. Kanellakopoulos (1995). *Nonlinear and Adaptive Control Design. Adaptive and Learning Systems for Signal Processing, Communication and Control*. NY: Wiley.

Lamnabhi-Lagarrigue, F., A. Annaswamy, S. Engell, A. Isaksson, P. Khargonekar, R. M. Murray, H. Nijmeijer, T. Samad, D. Tilbury, and P. Van den Hof (2017). "Systems and control for the future of humanity, research agenda: Current and future roles, impact and grand challenges". *Annual Reviews in Control*. 43: 1–64.

LaValle, S. M. and S. A. Hutchinson (1998). "Optimal motion planning for multiple robots having independent goals". *IEEE Transactions on Robotics and Automation*. 14(6): 912–925.

Lee, J. Y., J. B. Park, and Y. H. Choi (2012). "Integral Q-learning and explorized policy iteration for adaptive optimal control of continuous-time linear systems". *Automatica*. 48(11): 2850–2859.

Lemmon, M. (2010). "Event-triggered feedback in control, estimation, and optimization". In: *Networked Control Systems*. Springer. 293–358.

Lewis, F. L. and K. G. Vamvoudakis (2010). "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data". *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 41(1): 14–25.

Lewis, F. L. and D. Liu (2013). *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. Vol. 17. John Wiley & Sons.

Lewis, F., S. Jagannathan, and A. Yesildirak (1998). *Neural Network Control of Robot Manipulators and Non-Linear Systems*. CRC Press.

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

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

Lewis, F. L., H. Zhang, K. Hengster-Movric, and A. Das (2014). *Cooperative Control of Multi-Agent Systems-Optimal and Adaptive Design Approaches*. New York: Springer-Verlag.

Li, J., T. Chai, F. L. Lewis, Z. Ding, and Y. Jiang (2018). "Off-policy interleaved Q-learning: Optimal control for affine nonlinear discrete-time systems". *IEEE Transactions on Neural Networks and Learning Systems*. 30(5): 1308–1320.

Li, N., D. Oyler, M. Zhang, Y. Yildiz, A. Girard, and I. Kolmanovsky (2016). "Hierarchical reasoning game theory based approach for evaluation and testing of autonomous vehicle control systems". In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 727–733.

Li, Z.-H. and M. Krstic (1997). "Optimal design of adaptive tracking controllers for nonlinear systems". In: *American Control Conference, 1997. Proceedings of the 1997*. Vol. 2. IEEE. 1191–1197.

Liang, X. and Y. Xiao (2009). “Studying bio-inspired coalition formation of robots for detecting intrusions using game theory”. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 40(3): 683–693.

Littman, M. L. (2001). “Value-function reinforcement learning in Markov games”. *Cognitive Systems Research*. 2(1): 55–66.

Liu, D., H. Li, and D. Wang (2013). “Neural-network-based zero-sum game for discrete-time nonlinear systems via iterative adaptive dynamic programming algorithm”. *Neurocomputing*. 110: 92–100.

Liu, D., H. Li, and D. Wang (2014). “Online synchronous approximate optimal learning algorithm for multi-player non-zero-sum games with unknown dynamics”. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 44(8): 1015–1027.

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

Luo, B., H.-N. Wu, and T. Huang (2014). “Off-policy reinforcement learning for H_∞ control design”. *IEEE Transactions on Cybernetics*. 45(1): 65–76.

Lyshevski, S. (1996). “Constrained optimization and control of nonlinear systems: New results in optimal control”. In: *Proceedings of the 35th IEEE Conference on Decision and Control*. Vol. 1. IEEE. 541–546.

Lyshevski, S. E. (1998). “Optimal control of nonlinear continuous-time systems: Design of bounded controllers via generalized nonquadratic functionals”. In: *American Control Conference, 1998. Proceedings of the 1998*. Vol. 1. IEEE. 205–209.

Ma, Z., D. S. Callaway, and I. A. Hiskens (2011). “Decentralized charging control of large populations of plug-in electric vehicles”. *IEEE Transactions on Control Systems Technology*. 21(1): 67–78.

MacKenzie, A. B. and S. B. Wicker (2001). “Game theory in communications: Motivation, explanation, and application to power control”. In: *GLOBECOM’01. IEEE Global Telecommunications Conference (Cat. No. 01CH37270)*. Vol. 2. IEEE. 821–826.

Mannor, S. and J. S. Shamma (2007). “Multi-agent learning for engineers”. *Artificial Intelligence*. 171(7): 417–422.

Marden, J. R. and J. S. Shamma (2015). "Game theory and distributed control". In: *Handbook of Game Theory with Economic Applications*. Vol. 4. Elsevier. 861–899.

Marden, J. R. and J. S. Shamma (2018a). "Game theory and control". *Annual Review of Control, Robotics, and Autonomous Systems*. 1: 105–134.

Marden, J. R. and J. S. Shamma (2018b). "Game-theoretic learning in distributed control". *Handbook of Dynamic Game Theory*: 511–546.

McEneaney, W. M. (2006). *Max-Plus Methods for Nonlinear Control and Estimation*. Springer Science & Business Media.

McKelvey, R. D. and T. R. Palfrey (1995). "Quantal response equilibria for normal form games". *Games and Economic Behavior*. 10(1): 6–38.

Mehta, P. and S. Meyn (2009). "Q-learning and Pontryagin's minimum principle". In: *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference*. IEEE. 3598–3605.

Melo, F. S., S. P. Meyn, and M. I. Ribeiro (2008). "An analysis of reinforcement learning with function approximation". In: *Proceedings of the 25th International Conference on Machine Learning*. 664–671.

Modares, H. and F. L. Lewis (2014). "Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning". *Automatica*. 50(7): 1780–1792.

Modares, H., F. L. Lewis, and M. B. Naghibi-Sistani (2013). "Adaptive optimal control of unknown constrained-input systems using policy iteration and neural Networks". *IEEE Transactions on Neural Networks and Learning Systems*. 24(10): 1513–1525.

Modares, H., F. L. Lewis, and M.-B. Naghibi-Sistani (2014). "Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems". *Automatica*. 50(1): 193–202.

Modares, H., F. L. Lewis, and Z. P. Jiang (2015). "H_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning". *IEEE Transactions on Neural Networks and Learning Systems*. 26(10): 2550–2562.

Molin, A. and S. Hirche (2013). “On the optimality of certainty equivalence for event-triggered control systems”. *IEEE Transactions on Automatic Control*. 58(2): 470–474.

Morgenstern, O. and J. Von Neumann (1953). *Theory of Games and Economic Behavior*. Princeton University Press.

Mu, C., K. Wang, Q. Zhang, and D. Zhao (2020). “Hierarchical optimal control for input-affine nonlinear systems through the formulation of Stackelberg game”. *Information Sciences*. 517: 1–17.

Myerson, R. B. (2013). *Game Theory*. Harvard University Press.

Nash, J. (1951). “Non-cooperative games”. *Annals of Mathematics*. 54(2): 286–295.

Olfati-Saber, R. and R. M. Murray (2004). “Consensus problems in networks of agents with switching topology and time-delays”. *IEEE Transactions on Automatic Control*. 49(9): 1520–1533.

Olfati-Saber, R., J. A. Fax, and R. M. Murray (2007). “Consensus and cooperation in networked multi-agent systems”. *Proceedings of the IEEE*. 95(1): 215–233.

Paccagnan, D., B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros (2016a). “Distributed computation of generalized Nash equilibria in quadratic aggregative games with affine coupling constraints”. In: *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE. 6123–6128.

Paccagnan, D., M. Kamgarpour, and J. Lygeros (2016b). “On aggregative and mean field games with applications to electricity markets”. In: *2016 European Control Conference (ECC)*. IEEE. 196–201.

Paccagnan, D., B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros (2019). “Nash and Wardrop equilibria in aggregative games with coupling constraints”. *IEEE Transactions on Automatic Control*. 64(4): 1373–1388.

Palanisamy, M., H. Modares, F. L. Lewis, and M. Aurangzeb (2014). “Continuous-time Q-learning for infinite-horizon discounted cost linear quadratic regulator problems”. *IEEE Transactions on Cybernetics*. 45(2): 165–176.

Pavel, L. (2006). “A noncooperative game approach to OSNR optimization in optical networks”. *IEEE Transactions on Automatic Control*. 51(5): 848–852.

Pavlov, I. P. and W. Gantt (1928). *Lectures on Conditioned Reflexes: Twenty-Five Years of Objective Study of the Higher Nervous Activity (Behaviour) of Animals*. Liverwright Publishing Corporation.

Perez, A., R. Platt, G. Konidaris, L. Kaelbling, and T. Lozano-Perez (2012). "LQR-RRT*: Optimal sampling-based motion planning with automatically derived extension heuristics". In: *IEEE International Conference on Robotics and Automation*. 2537–2542.

Pita, J., M. Jain, J. Marecki, F. Ordóñez, C. Portway, M. Tambe, C. Western, P. Paruchuri, and S. Kraus (2008). "Deployed ARMOR protection: The application of a game theoretic model for security at the Los Angeles International Airport". In: *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems: Industrial Track*. 125–132.

Polycarpou, M., J. Farrell, and M. Sharma (2003). "On-line approximation control of uncertain nonlinear systems: Issues with control input saturation". In: *Proceedings of the 2003 American Control Conference, 2003*. Vol. 1. IEEE. 543–548.

Poveda, J. I., M. Benosman, and A. R. Teel (2019). "Hybrid online learning control in networked multiagent systems: A survey". *International Journal of Adaptive Control and Signal Processing*. 33(2): 228–261.

Qu, Z. (2009). *Cooperative Control of Dynamical Systems: Applications to Autonomous Vehicles*. Springer Science & Business Media.

Rabin, M. O. (1957). "Effective computability of winning strategies". *Contributions to the Theory of Games*. 3(39): 147–157.

Rantzer, A. (2008). *Using Game Theory for Distributed Control Engineering*. Department of Automatic Control, Lunds University.

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

Ren, W. and R. W. Beard (2005). "Consensus seeking in multiagent systems under dynamically changing interaction topologies". *IEEE Transactions on Automatic Control*. 50(5): 655–661.

Ren, W. and R. W. Beard (2008). *Distributed Consensus in Multi-Vehicle Cooperative Control*. Springer.

Ren, W., R. W. Beard, and E. M. Atkins (2005). “A survey of consensus problems in multi-agent coordination”. In: *Proceedings of the 2005, American Control Conference, 2005.* IEEE. 1859–1864.

Roth, A. E. and I. Erev (1995). “Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term”. *Games and Economic Behavior*. 8(1): 164–212.

Roy, S., C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya, and Q. Wu (2010). “A survey of game theory as applied to network security”. In: *2010 43rd Hawaii International Conference on System Sciences*. IEEE. 1–10.

Rudin, W. (1964). *Principles of Mathematical Analysis*. Vol. 3. New York: McGraw-Hill.

Saad, W., Z. Han, H. V. Poor, and T. Başar (2012). “Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications”. *IEEE Signal Processing Magazine*. 29(5): 86–105.

Sahoo, P. P. and K. G. Vamvoudakis (2020). “On–off adversarially robust Q-learning”. *IEEE Control Systems Letters*. 4(3): 749–754.

Saleheen, F. and C.-H. Won (2019). “Statistical Stackelberg game control: Open-loop minimal cost variance case”. *Automatica*. 101: 338–344.

Schaft, A. J. van der (1992). “L_2-gain analysis of nonlinear systems and nonlinear state-feedback H_∞ control”. *IEEE Transactions on Automatic Control*. 37(6): 770–784.

Schwalbe, U. and P. Walker (2001). “Zermelo and the early history of game theory”. *Games and Economic Behavior*. 34(1): 123–137.

Semsar-Kazerooni, E. and K. Khorasani (2009). “Multi-agent team cooperation: A game theory approach”. *Automatica*. 45(10): 2205–2213.

Simaan, M. and J. B. Cruz (1973a). “On the Stackelberg strategy in nonzero-sum games”. *Journal of Optimization Theory and Applications*. 11(5): 533–555.

Simaan, M. and J. B. Cruz (1973b). “Additional aspects of the Stackelberg strategy in nonzero-sum games”. *Journal of Optimization Theory and Applications*. 11(6): 613–626.

Simon, H. A. (1984). *Models of Bounded Rationality, Volume 1: Economic Analysis and Public Policy*. Vol. 1. MIT Press Books.

Solowjow, F. and S. Trimpe (2020). "Event-triggered learning". *Automatica*. 117: 109009.

Sontag, E. D. (1993). "Neural networks for control". In: *Essays on Control: Perspectives in the Theory and Its Applications*. Ed. by H. L. Trentelman and J. C. Willems. Boston, MA: Birkhäuser Boston. 339–380.

Sontag, E. and H. Sussmann (1997). "Mathematical theory of neural networks". *Tech. rep.* Rutgers, The State University. New Brunswick, NJ.

Strzalecki, T. (2014). "Depth of reasoning and higher order beliefs". *Journal of Economic Behavior & Organization*. 108: 108–122.

Sutton, R. S. and A. G. Barto (2018). *Reinforcement Learning: An Introduction*. 2nd edn. Vol. 1. Cambridge: MIT Press.

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

Tabuada, P. (2007). "Event-triggered real-time scheduling of stabilizing control tasks". *IEEE Transactions on Automatic Control*. 52(9): 1680–1685.

Tambe, M. (2011). *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*. Cambridge University Press.

Tao, G. (2003). "Adaptive control design and analysis". *Adaptive, Cognitive Dynamic Systems: Signal Processing, Learning, Communication and Control*. NY: Wiley.

Tijs, S. (2003). *Introduction to Game Theory*. Springer.

Tsai, J., S. Rathi, C. Kiekintveld, F. Ordonez, and M. Tambe (2009). "IRIS-a tool for strategic security allocation in transportation networks". *AAMAS (Industry Track)*: 37–44.

Tsitsiklis, J. (1984). *Problems in Decentralized Decision Making and Computation*. Ph.D. Thesis, MIT.

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

Ungureanu, V. (2018). *Pareto–Nash–Stackelberg Game and Control Theory*. Springer.

Vamvoudakis, K. G. (2014a). "Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems". *IEEE/CAA Journal of Automatica Sinica*. 1(3): 282–293.

Vamvoudakis, K. (2014b). "An online actor/critic algorithm for event-triggered optimal control of continuous-time nonlinear systems". In: *American Control Conference (ACC), 2014*. 1–6.

Vamvoudakis, K. G. (2015). "Non-zero sum Nash Q-learning for unknown deterministic continuous-time linear systems". *Automatica*. 61: 274–281.

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 F. L. Lewis (2010). "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem". *Automatica*. 46(5): 878–888.

Vamvoudakis, K. G. and F. L. Lewis (2011). "Multi-player non-zero-sum games: Online adaptive learning solution of coupled Hamilton–Jacobi equations". *Automatica*. 47(8): 1556–1569.

Vamvoudakis, K. G. and F. Lewis (2012). "Online solution of nonlinear two-player zero-sum games using synchronous policy iteration". *International Journal of Robust and Nonlinear Control*. 22(13): 1460–1483.

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

Vamvoudakis, K. G. and J. P. Hespanha (2018). "Cooperative Q-learning for rejection of persistent adversarial inputs in networked linear quadratic systems". *IEEE Transactions on Automatic Control*. 63(4): 1018–1031.

Vamvoudakis, K. G. and A. Kanellopoulos (2019). "Non-equilibrium learning and cyber-physical security". In: *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 1–6.

Vamvoudakis, K. G., D. Vrabie, and F. Lewis (2011). “Online learning algorithm for zero-sum games with integral reinforcement learning”. *Journal of Artificial Intelligence and Soft Computing Research*. 1(4): 315–332.

Vamvoudakis, K. G., F. L. Lewis, and G. R. Hudas (2012). “Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality”. *Automatica*. 48(8): 1598–1611.

Vamvoudakis, K. G., D. Vrabie, and F. L. Lewis (2014). “Online adaptive algorithm for optimal control with integral reinforcement learning”. *International Journal of Robust and Nonlinear Control*. 24(17): 2686–2710.

Vamvoudakis, K. G., P. J. Antsaklis, W. E. Dixon, J. P. Hespanha, F. L. Lewis, H. Modares, and B. Kiumarsi (2015). “Autonomy and machine intelligence in complex systems: A tutorial”. In: *2015 American Control Conference (ACC)*. IEEE. 5062–5079.

Vamvoudakis, K. G., M. F. Miranda, and J. P. Hespanha (2016). “Asymptotically stable adaptive–optimal control algorithm with saturating actuators and relaxed persistence of excitation”. *IEEE Transactions on Neural Networks and Learning Systems*. 27(11): 2386–2398.

Vamvoudakis, K. G., H. Modares, B. Kiumarsi, and F. L. Lewis (2017a). “Game theory-based control system algorithms with real-time reinforcement learning: How to solve multiplayer games online”. *IEEE Control Systems Magazine*. 37(1): 33–52.

Vamvoudakis, K. G., A. Mojoodi, and H. Ferraz (2017b). “Event-triggered optimal tracking control of nonlinear systems”. *International Journal of Robust and Nonlinear Control*. 27(4): 598–619.

Vamvoudakis, K. G., F. L. Lewis, and W. E. Dixon (2019). “Open-loop Stackelberg learning solution for hierarchical control problems”. *International Journal of Adaptive Control and Signal Processing*. 33(2): 285–299.

Velupillai, K. V. (2011). “Non-linear dynamics, complexity and randomness: Algorithmic foundations”. *Journal of Economic Surveys*. 25(3): 547–568.

Von Stackelberg, H. (2010). *Market Structure and Equilibrium*. Springer Science & Business Media.

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*. Vol. 81. IET.

Wang, C. and D. J. Hill (2009). *Deterministic Learning Theory for Identification, Recognition, and Control*. Vol. 32. CRC Press.

Wang, D., C. Mu, H. He, and D. Liu (2016). "Event-driven adaptive robust control of nonlinear systems with uncertainties through NDP strategy". *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 47(7): 1358–1370.

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

Wang, D., H. He, and D. Liu (2017b). "Improving the critic learning for event-based nonlinear H_∞ control design". *IEEE Transactions on Cybernetics*. 47(10): 3417–3428.

Wang, D., H. He, X. Zhong, and D. Liu (2017c). "Event-driven nonlinear discounted optimal regulation involving a power system application". *IEEE Transactions on Industrial Electronics*. 64(10): 8177–8186.

Wang, D., C. Mu, D. Liu, and H. Ma (2017d). "On mixed data and event driven design for adaptive-critic-based nonlinear H_∞ control". *IEEE Transactions on Neural Networks and Learning Systems*. 29(4): 993–1005.

Wang, X. and M. D. Lemmon (2011). "Event-triggering in distributed networked control systems". *IEEE Transactions on Automatic Control*. 56(3): 586–601.

Wang, X. F. and G. Chen (2002). "Pinning control of scale-free dynamical networks". *Physica A: Statistical Mechanics and Its Applications*. 310(3–4): 521–531.

Watkins, C. (1989). *Learning from Delayed Rewards*. Ph.D. Thesis, Cambridge University, Cambridge, UK.

Watkins, C. J. and P. Dayan (1992). "Q-learning". *Machine Learning*. 8(3–4): 279–292.

Webb, D. J. and J. van den Berg (2013). "Kinodynamic RRT*: Asymptotically optimal motion planning for robots with linear dynamics". In: *IEEE International Conference on Robotics and Automation*. 5054–5061.

Wei, Q., R. Song, and P. Yan (2015). "Data-driven zero-sum neuro-optimal control for a class of continuous-time unknown nonlinear systems with disturbance using ADP". *IEEE Transactions on Neural Networks and Learning Systems*. 27(2): 444–458.

Werbos, P. J. (1992). "Approximate dynamic programming for real-time control and neural modeling". *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*. 15: 493–525.

Werbos, P. J. (2007). "Using ADP to understand and replicate brain intelligence: The next level design?" In: *Neurodynamics of Cognition and Consciousness*. Springer. 109–123.

Wiering, M. and M. Van Otterlo (2012). "Reinforcement learning". *Adaptation, Learning, and Optimization*. 12: 3.

Wu, H.-N. and B. Luo (2012). "Neural network based online simultaneous policy update algorithm for solving the HJI equation in nonlinear H_∞ control". *IEEE Transactions on Neural Networks and Learning Systems*. 23(12): 1884–1895.

Xu, H., S. Jagannathan, and F. L. Lewis (2012). "Stochastic optimal control of unknown linear networked control system in the presence of random delays and packet losses". *Automatica*. 48(6): 1017–1030.

Xu, H., Q. Zhao, and S. Jagannathan (2014). "Optimal regulation of uncertain dynamic systems using adaptive dynamic programming". *Journal of Control and Decision*. 1(3): 226–256.

Yang, X. and H. He (2018). "Adaptive critic designs for event-triggered robust control of nonlinear systems with unknown dynamics". *IEEE Transactions on Cybernetics*. 49(6): 2255–2267.

Yang, Y., K. G. Vamvoudakis, and H. Modares (2020a). "Safe reinforcement learning for dynamical games". *International Journal of Robust and Nonlinear Control*. 30(9): 3706–3726.

Yang, Y., K. G. Vamvoudakis, H. Modares, Y. Yin, and D. C. Wunsch (2020b). "Safe intermittent reinforcement learning with static and dynamic event generators". *IEEE Transactions on Neural Networks and Learning Systems*.

Ye, M. and G. Hu (2017). "Distributed Nash equilibrium seeking by a consensus based approach". *IEEE Transactions on Automatic Control*. 62(9): 4811–4818.

Zames, G. and B. Francis (1983). "Feedback, minimax sensitivity, and optimal robustness". *IEEE Transactions on Automatic Control*. 28(5): 585–601.

Zhang, H., L. Cui, and Y. Luo (2012). "Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP". *IEEE Transactions on Cybernetics*. 43(1): 206–216.

Zhang, H., T. Feng, G.-H. Yang, and H. Liang (2014a). "Distributed cooperative optimal control for multiagent systems on directed graphs: An inverse optimal approach". *IEEE Transactions on Cybernetics*. 45(7): 1315–1326.

Zhang, H., C. Qin, B. Jiang, and Y. Luo (2014b). "Online adaptive policy learning algorithm for H_∞ state feedback control of unknown affine nonlinear discrete-time systems". *IEEE Transactions on Cybernetics*. 44(12): 2706–2718.

Zhang, H., J. Zhang, G.-H. Yang, and Y. Luo (2014c). "Leader-based optimal coordination control for the consensus problem of multiagent differential games via fuzzy adaptive dynamic programming". *IEEE Transactions on Fuzzy Systems*. 23(1): 152–163.

Zhang, Q., D. Zhao, and D. Wang (2016a). "Event-based robust control for uncertain nonlinear systems using adaptive dynamic programming". *IEEE Transactions on Neural Networks and Learning Systems*. 29(1): 37–50.

Zhang, Q., D. Zhao, and Y. Zhu (2016b). "Event-triggered H_∞ control for continuous-time nonlinear system via concurrent learning". *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 47(7): 1071–1081.

Zhang, Y., S. Li, and X. Zhou (2020). *Deep Reinforcement Learning with Guaranteed Performance*. Springer.

Zhong, X. and H. He (2016). "An event-triggered ADP control approach for continuous-time system with unknown internal states". *IEEE Transactions on Cybernetics*. 47(3): 683–694.

Zhu, Y., D. Zhao, H. He, and J. Ji (2016). "Event-triggered optimal control for partially unknown constrained-input systems via adaptive dynamic programming". *IEEE Transactions on Industrial Electronics*. 64(5): 4101–4109.