Editorial
Special Issue on Adaptive Dynamic Programming and Reinforcement Learning

THE PAST decade has witnessed a surge in research activities related to adaptive dynamic programming (ADP) and reinforcement learning (RL), particularly for control applications. Several books [item 1)–5) in the Appendix] and survey papers [item 6)–10) in the Appendix] have been published on the subject. Both ADP and RL provide approximate solutions to dynamic programming problems. In a 1995 article by Barto *et al.* [item 11) in the Appendix], they introduced the so-called "adaptive real-time dynamic programming," which was specifically to apply ADP for real-time control. Later, in 2002, Murray *et al.* [item 12) in the Appendix] developed an ADP algorithm for optimal control of continuous-time affine nonlinear systems. On the other hand, the most famous algorithms in RL are the temporal difference algorithm [item 13) in the Appendix] and the Q-learning algorithm [item 14) and 15) in the Appendix].

It all started more than 40 years' ago, when Werbos outlined an approach [item 16) in the Appendix] which was later named "adaptive critic designs" (ACD) [item 17)–20) in the Appendix]. By the time when he published the 1992 book chapter [item 18) in the Appendix], Werbos has used the three terms "ACD," "approximate dynamic programming (ADP)," and "RL" interchangeably, and later, also interchangeably with "adaptive dynamic programming (ADP)" [item 21)–23) in the Appendix].

The main focus of ADP/RL research is to solve the Hamilton–Jacobi–Bellman (HJB) equation with manageable amount of computation to avoid the "curse of dimensionality." Many methods of direct solution and iterative solution have been proposed to solve the HJB equation, but new ideas are still wanting. A very important research direction would be to find novel and efficient approaches for solving the HJB equation.

Dynamic programming is an important subject in optimal control [item 24) in the Appendix]. It is therefore expected that ADP/RL plays an important role in solving optimal control problems of complex nonlinear systems. Furthermore, the most famous application of ADP/RL is the AlphaGo [item 25) in the Appendix] computer game, when combined with deep learning. Applications of ADP/RL to real-world problems which can bring about more significant social and economic impact are expected.

This special issue contains 22 articles from those accepted by the Transactions during 2018–2019. It can only give a glimpse of the rapid development of the field of ADP/RL. On the one hand, we were not able to put all articles accepted by the Transactions on ADP/RL into this special issue. On the other hand, quite a few other journals including several IEEE TRANSACTIONS have also published a good number of articles on ADP/RL.

The 22 articles are divided into three groups: 1) theoretical foundations of ADP/RL; 2) ADP/RL for controls; and 3) applications of ADP/RL.

The first group consists of eight articles that deal with recent theoretical development of ADP/RL. It starts with the article by Luo *et al.* on the adaptive RL method for control of discrete-time nonlinear systems, where a balancing parameter is added to integrate value iteration (VI) and policy iteration (PI), in order to accelerate VI and avoid the need of an initial admissible control of PI. Invariant ADP is studied in the article by Zhu *et al.* to develop a new PI algorithm for optimal control of discrete-time systems, where a given policy is evaluated in its invariantly admissible region at each iteration and a new policy and a new region are updated for the next iteration. The next article by Liang *et al.* studies the optimal control problem for discrete stochastic processes, by developing a novel local PI ADP algorithm, where the iterative decision rules are updated in a subspace of the state to reduce the computational burden for the CPU in comparison with the conventional PI algorithm. In the article by Zhang *et al.* a novel integral RL method based on VI is proposed for the H_∞ control of continuous-time constrained nonlinear systems, where a nonquadratic function is introduced to reconstruct the L_2-gain condition and the implementation using critic-only structure is employed. In the next article, Su *et al.* develop a novel ADP algorithm for infinite-horizon H_∞ optimal control problems of unknown continuous-time nonlinear systems subject to external disturbances, where the generalized fuzzy hyperbolic models are utilized to establish an identifier–critic architecture. In the article by Song *et al.* an optimal critic learning algorithm based on single neural network (NN) PI is developed for continuous-time 2-player zero-sum games, where a single NN-based online PI algorithm is employed with the NN weights obtained by the least-squares method. Finally, the article by Song *et al.* considers the tasks of feature selection and policy evaluation based on linear value function approximation in RL problems, where an efficient l_1-regularized recursive least-squares-based online algorithm is

Digital Object Identifier 10.1109/TSMC.2020.3025549

2168-2216 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

developed through nested optimization decomposition to solve a series of standard optimization problems.

The second group has 11 articles that studies ADP/RL for control applications to nonlinear systems and linear systems. These applications include tracking control, robust control, guaranteed cost control, decentralized control, networked control, sliding-mode control, and consensus control. This group starts with the article by Fan *et al.* presenting a new model-free approach for the output regulation problem of networked control systems with the possibility of loosing the system state in the feedback process, where an off-policy algorithm based on RL is developed to calculate the feedback gain using the measured data when dropout occurs. Next, the article by Yang and He develops a decentralized event-triggered control strategy for nonlinear systems with uncertain interconnections, employing gradient descent approach and experience replay technique in critic network training to obtain asymptotically stable overall system performance. Then, Mu *et al.* study an ADP-based optimal control strategy for the tracking control of a class of continuous-time nonlinear systems with unmatched uncertainties, by transforming the unmatched uncertain term into an auxiliary system that contains a steady part and a transient part with the related controllers solved separately. The article by Guo *et al.* presents an integral RL-based adaptive NN tracking control for continuous-time nonlinear systems with unknown control directions, where the reinforcement signal is explicitly embedded in the updating law of the action NN to minimize the long-term performance index. Liu *et al.* study the H_∞ tracking control problem for a class of discrete time-delay systems using the data-based ADP algorithm, where novel data-based Bellman equation is derived and solved. In the article by Yang *et al.* they develop a novel event-triggered integral sliding-mode control strategy for constrained nonlinear systems with disturbances via robust ADP, where a mixed event-triggering scheme with two different triggering parts is developed and the optimal performance of sliding-mode dynamics with input constrains is ensured by a critic-only NN-based ADP scheme employing concurrent learning. The next article by Hou *et al.* develops an H_∞ optimal tracking controller for completely unknown discrete-time nonlinear systems with control constraints by using an iterative adaptive learning algorithm, where an action–disturbance–critic NN structure is employed for the iterative dual heuristic programming algorithm with convergence guarantee of the costate function and the control policy. The article by Peng *et al.* studies an online reinforcement Q-learning algorithm for H_∞ tracking control of unknown discrete-time linear systems, where both state-data-driven and output-data-driven reinforcement Q-learning algorithms toward finding the control policies are elaborated. Wang *et al.* develop in their article a local Q-function-based ADP method for the optimal consensus control problem for unknown discrete-time nonlinear multiagent systems, where the local Q-function is defined using the local consensus error and the actions of the agent and its neighbors and a distributed PI technique is employed with guaranteed convergence. Zhang *et al.* present a novel event-driven guaranteed cost control method in their article for nonlinear systems subject to actuator faults, where an RL algorithm is utilized to develop a sliding-mode control strategy. Finally, Wu and Liu develop a data-driven guaranteed cost control design for linear systems with time-varying parameter uncertainties, where a modified simultaneous policy update algorithm is proposed for solving the algebraic Riccati equation using off-policy RL and experience replay technique.

The third group contains four articles on practical applications of ADP/RL. In the article by Liu *et al.* a switching-based control formulation for multi-intersection and multiphase traffic light systems is developed based on ADP, which can handle different cycle times across intersections without the need for synchronization constraints. In the next article, Li *et al.* proposed an RL-based adaptive control algorithm to solve the tracking problem of a wheeled mobile robot by transforming the robot model into an affine nonlinear system, and introducing a delay matrix function and appropriate Lyapunov–Krasovskii functionals to overcome the problems caused by the state and input time delays. The application considered in the article by Mu *et al.* is about automating the configuration of MapReduce, where an RL-based scheme for automatically configuring the MapReduce parameters is proposed and two variations are evaluated which are traditional RL-based and deep RL-based, respectively. Finally, Chen *et al.* study deep RL for improving the efficiency of searching for the most informative sampling locations in wireless sensor networks with mobile robotic sensor nodes.

We hope you enjoy reading this special issue and hope it may inspire some future endeavors.

DERONG LIU
School of Automation
Guangdong University of Technology
Guangzhou 510006, China

FRANK L. LEWIS
UTA Research Institute
University of Texas at Arlington
Fort Worth, TX 76118 USA

QINGLAI WEI
State Key Laboratory of Management and Control for Complex Systems
Institute of Automation
Chinese Academy of Sciences
Beijing 100190, China
University of Chinese Academy of Sciences
Beijing 100049, China

APPENDIX
RELATED WORK

1) W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Hoboken, NJ, USA: Wiley, 2007.

2) D. Vrabie, K. Vamvoudakis, and F. L. Lewis, *Optimal Adaptive Control and Differential Games by Reinforcement Learning Principles*. London, U.K.: IET Press, 2012.

3) H. Zhang, D. Liu, Y. Luo, and D. Wang, *Adaptive Dynamic Programming for Control: Algorithms and Stability*. London, U.K.: Springer, 2013.
4) F. L. Lewis and D. Liu, *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. Hoboken, NJ, USA: Wiley, 2013.
5) D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming With Applications in Optimal Control*. Cham, Switzerland: Springer, 2017.
6) F.-Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: An introduction," *IEEE Comput. Intell. Mag.*, vol. 4, no. 2, pp. 39–47, May 2009.
7) F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits Syst. Mag.*, vol. 9, no. 3, pp. 32–50, Aug. 2009.
8) F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Syst. Mag.*, vol. 32, no. 6, pp. 76–105, Dec. 2012.
9) D. Wang, H. He, and D. Liu, "Adaptive critic nonlinear robust control: A survey," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3429–3451, Oct. 2017.
10) Y. Zhu and D. Zhao, "Comprehensive comparison of online ADP algorithms for continuous-time optimal control," *Artif. Intell. Rev.*, vol. 49, no. 4, pp. 531–547, Apr. 2018.
11) A. G. Barto, S. J. Bradtke, and S. P. Singh, "Learning to act using real-time dynamic programming," *Artif. Intell.*, vol. 72, no. 1–2, pp. 81–138, Jan. 1995.
12) J. J. Murray, C. J. Cox, G. G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 32, no. 2, pp. 140–153, May 2002.
13) R. S. Sutton, "Learning to predict by the method of temporal differences," *Mach. Learn.*, vol. 3, no. 1, pp. 9–44, Aug. 1988.
14) C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, Cambridge Univ., Cambridge, U.K., 1989.
15) C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, May 1992.
16) P. J. Werbos, "Advanced forecasting methods for global crisis warning and models of intelligence," *Gen. Syst.*, vol. 22, pp. 25–38, Jan. 1977.
17) P. J. Werbos, "A menu of designs for reinforcement learning over time," in *Neural Networks for Control*, W. T. Miller, R. S. Sutton, and P. J. Werbos, Eds. Cambridge, MA, USA: MIT Press, 1990, ch. 3.
18) P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," in *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, D. A. White and D. A. Sofge, Eds. New York, NY, USA: Van Nostrand Reinhold, 1992, ch. 13.
19) D. V. Prokhorov, R. A. Santiago, and D. C. Wunsch, "Adaptive critic designs: A case study for neurocontrol," *Neural Netw.*, vol. 8, no. 9, pp. 1367–1372, 1995.
20) D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 997–1007, Sep. 1997.
21) P. J. Werbos, "ADP: Goals, opportunities and principles," in *Handbook of Learning and Approximate Dynamic Programming*, J. Si, A. G. Barto, W. B. Powell, and D. C. Wunsch, Eds. New York, NY, USA: Wiley, 2004, ch. 1.
22) P. J. Werbos, "Using ADP to understand and replicate brain intelligence: The next level design," in *Proc. IEEE Int. Symp. Approximate Dyn. Program. Reinforcement Learn.*, Honolulu, HI, USA, Apr. 2007, pp. 209–216.
23) P. J. Werbos, "ADP: The key direction for future research in intelligent control and understanding brain intelligence," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 4, pp. 898–900, Aug. 2008.
24) F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*, 3rd ed. Hoboken, NJ, USA: Wiley, 2012.
25) D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.

Derong Liu (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1994.

He was a Staff Fellow with General Motors Research and Development Center, Warren, MI, USA, from 1993 to 1995. He was an Assistant Professor with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, from 1995 to 1999. He joined the University of Illinois at Chicago, Chicago, IL, USA, in 1999, and became a Full Professor of Electrical and Computer Engineering and of Computer Science in 2006. He served as the Associate Director of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2010 to 2015. He has published 19 books.

Prof. Liu received the Faculty Early Career Development Award from the National Science Foundation in 1999, the University Scholar Award from the University of Illinois from 2006 to 2009, the Overseas Outstanding Young Scholar Award from the National Natural Science Foundation of China in 2008, the Outstanding Achievement Award from Asia–Pacific Neural Network Assembly in 2014, the INNS Gabor Award in 2018, the IEEE TNNLS Outstanding Paper Award in 2018, the IEEE SMC Society Andrew P. Sage Best Transactions Paper Award in 2018, and the IEEE/CCA JAS Hsue-Shen Tsien Paper Award in 2019. He was selected for the 100 Talents Program by the Chinese Academy of Sciences in 2008. He is the Editor-in-Chief of *Artificial Intelligence Review* (Springer). He was the Editor-in-Chief of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2010 to 2015. He is a Fellow of the International Neural Network Society and the International Association of Pattern Recognition.

Frank L. Lewis (Life Fellow, IEEE) received the B.S. degree in electrical engineering and physics from Rice University, Houston, TX, USA, in 1971, the M.S.E.E. degree from Rice University, the M.S. degree in aeronautical engineering from the University of West Florida, Pensacola, FL, USA, in 1977, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta, GA, USA, in 1981.

He is a UTA Charter Distinguished Scholar Professor, a UTA Distinguished Teaching Professor, and a Moncrief-O'Donnell Chair with the University of Texas at Arlington Research Institute, Fort Worth, TX, USA. He works in feedback control, intelligent systems, reinforcement learning, cooperative control systems, and nonlinear systems. He has authored seven U.S. patents, numerous journal special issues, 445 journal papers, 20 books, including the textbooks Optimal Control, Aircraft Control, Optimal Estimation, and Robot Manipulator Control.

Dr. Lewis received the Fulbright Research Award, the NSF Research Initiation Grant, the ASEE Terman Award, the International Neural Network Society's Gabor Award, the U.K. Inst Measurement & Control Honeywell Field Engineering Medal, the IEEE Computational Intelligence Society Neural Networks Pioneer Award, the AIAA Intelligent Systems Award, and the AACC Ragazzini Award. He is ranked at position 84 worldwide, 66 in the USA, and 3 in Texas of all scientists in Computer Science and Electronics, by Guide2Research.com. He has a total of 65 000 Google Scholar Citations. He has received over $12M in 100 research grants from NSF, ARO, ONR, AFOSR, DARPA, and USA industry contracts. He helped win the U.S. SBA Tibbets Award in 1996 as the Director of the UTA Research Institute SBIR Program. He is a Fellow of the National Academy of Inventors, IFAC, AAAS, and the U.K. Institute of Measurement & Control, a PE of Texas, and a U.K. Chartered Engineer.

Qinglai Wei (Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2002 and 2009, respectively.

From 2009 to 2011, he was a Postdoctoral Fellow with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, where he is currently a Professor. He has authored one book and published over 70 international journal papers. His research interests include adaptive dynamic programming, neural-network-based control, optimal control, nonlinear systems, and their industrial applications.

Prof. Wei was a recipient of the Shuang-Chuang Talents in Jiangsu Province, China, in 2014, the Outstanding Paper Award of Acta Automatica Sinica in 2011, the Zhang Siying Outstanding Paper Award of Chinese Control and Decision Conference in 2015, and the Young Researcher Award of Asia Pacific Neural Network Society in 2016. He has been an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS since 2019, IEEE TRANSACTION ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS since 2016, *Information Sciences* since 2016, *Neurocomputing* since 2016, *Optimal Control Applications and Methods* since 2016, and *Acta Automatica Sinica* since 2015. He has been the Secretary of the IEEE Computational Intelligence Society Beijing Chapter since 2015. He was the Registration Chair of the 12th World Congress on Intelligent Control and Automation in 2016, the 2014 IEEE World Congress on Computational Intelligence, the 2013 International Conference on Brain Inspired Cognitive Systems, and the 8th International Symposium on Neural Networks in 2011. He was the Publication Chair of the 5th International Conference on Information Science and Technology in 2015 and the 9th International Symposium on Neural Networks in 2012. He was the Finance Chair of the 4th International Conference on Intelligent Control and Information Processing in 2013 and the Publicity Chair of the 2012 International Conference on Brain Inspired Cognitive Systems. He was a guest editor for several international journals.