8-1-2008

Issues on Stability of ADP Feedback Controllers for Dynamical Systems

S. N. Balakrishnan
Missouri University of Science and Technology, bala@mst.edu

Jie Ding

F. L. Lewis

Issues on Stability of ADP Feedback Controllers for Dynamical Systems

S. N. Balakrishnan, *Member, IEEE*, Jie Ding, and Frank L. Lewis, *Fellow, IEEE*

(Invited Paper)

***Abstract*—This paper traces the development of neural-network (NN)-based feedback controllers that are derived from the principle of adaptive/approximate dynamic programming (ADP) and discusses their closed-loop stability. Different versions of NN structures in the literature, which embed mathematical mappings related to solutions of the ADP-formulated problems called "adaptive critics" or "action-critic" networks, are discussed. Distinction between the two classes of ADP applications is pointed out. Furthermore, papers in "model-free" development and model-based neurocontrollers are reviewed in terms of their contributions to stability issues. Recent literature suggests that work in ADP-based feedback controllers with assured stability is growing in diverse forms.**

***Index Terms*—Adaptive/approximate dynamic programming (ADP), feedback controllers, neural networks (NNs), nonlinear control, stability.**

I. Introduction

DYNAMICAL systems are ubiquitous in nature and include naturally occurring systems such as the cell and more complex biological organisms and man-made systems such as automobiles, aircraft, missiles, and retail inventories. von Bertalanffy [10] and Whitehead [54] were among the first to provide a modern theory of systems at the beginning of the century. Three components that define a system are its outputs that can be measured, inputs to it that can be manipulated, and its internal dynamics. Feedback control develops suitable control inputs, based on the difference between observed and desired behaviors, for a dynamical system such that, with time, the observed behavior reaches a desired behavior prescribed by the user. All biological systems employ feedback for survival, with even the simplest of cells using chemical diffusion based on feedback to create a potential difference across the membrane to maintain its homeostasis, or required equilibrium condition for survival. Volterra was the first to show that feedback is responsible for the balance of two populations of fish in a pond, and Darwin showed that feedback over extended time periods provides the subtle pressures that cause the evolution of species.

There is a large and well-established body of design and analysis techniques for feedback control systems, which have been responsible for successes in the industrial revolution, the ship and aircraft design, and the space age. Most of these techniques, such as classical design methods, multivariable control, linear quadratic regulators, robust control, and H-infinity control, are based on linear systems theory. Nonlinear design approaches include nonlinear optimal control, adaptive control, sliding mode control, backstepping control, and others [24]. Feedback control synthesis is further complicated due to unknown dynamics, modeling errors, and various sorts of disturbances, uncertainties, and noise. This, coupled with the increasing complexity of today's dynamical systems, creates a need for advanced control design techniques that overcome limitations on traditional feedback control techniques.

In recent years, the field of biologically inspired control is growing where the feedback control systems mimic the functions of living biological systems. Ideas from swarm behavior, immune systems, and brain functions have found their way to optimization and control. Theory and applications of nonlinear neural networks (NNs), which embed brainlike structures in the field of feedback control, have been well documented [27], [28]. It is generally understood that NNs provide an elegant extension of adaptive control techniques to nonlinearly parameterized learning systems.

II. NN for Feedback Control

NN literature on feedback control is already large and increasing at a rapid rate. The use of NNs in feedback control systems was first proposed by Werbos [50]. Since then, NN control has been studied by many researchers. Recently, NN has entered the mainstream of control theory as a natural extension of adaptive control to systems that are nonlinear in the tunable parameters. The state of NN control is well illustrated by papers in the Automatica Special Issue on NN control [32]. Overviews of the initial work in NN control are provided by Miller *et al.* [30] and the *Handbook of Intelligent Control* [53], which highlighted a host of difficulties to be addressed for closed-loop control applications. Note that NN applications in closed-loop control are fundamentally different from open-loop applications such as classification and image processing. The basic multilayer NN tuning strategy is backpropagation [49]. Basic problems that had to be addressed for closed-loop NN control [51], [52] included weight initialization for feedback

The work of S. N. Balakrishnan and J. Ding was supported by the National Science Foundation (NSF) under Grant ECS-0601706. The work of F. L. Lewis was supported in part by the NSF under Grant ECS-0501451 and in part by the Army Research Office under Grant W91NF-05-1-0314.

S. N. Balakrishnan and J. Ding are with the Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, MO 65401 USA (e-mail: bala@mst.edu; jdr5c@mst.edu).

F. L. Lewis is with the Automation and Robotics Research Institute, University of Texas at Arlington, Fort Worth, TX 76118-7115 USA (e-mail: lewis@uta.edu).

Digital Object Identifier 10.1109/TSMCB.2008.926599

stability, determining the gradients needed for backpropagation tuning, determining what to backpropagate, obviating the need for preliminary offline tuning, modifying backprop so that it tunes the weights forward through time, and providing efficient computer code for implementation.

A. ADP Paradigms

Many papers in the literature have coopted other nonlinear control concepts such as feedback linearization, backstepping, and sliding mode control to make the NN-based feedback controllers more powerful. Artificial-NN literature on solving optimal control problems with NNs really got an impetus with Werbos' [51], [52] work presentation of the notions of "adaptive critics" to solve optimal control problems formulated on an adaptive/approximate dynamic programming (ADP) framework. It is well known that the dynamic programming formulation offers the most comprehensive solution to nonlinear optimal control; however, a huge amount of computational and storage requirements are needed to solve the associated Hamilton–Jacobi–Bellman (HJB) equation (also known as the Bellman equation) [29]. Werbos [52] proposed a means to get around this numerical complexity by using "approximate dynamic programming" formulations. His methods approximate the original problem with a discrete formulation. Solution to the ADP formulation is obtained through the two-NN adaptive critic approach. He proposed two basic versions. In one version called the heuristic dynamic programming (HDP), one network, which is called the "action network" or the "actor," represents the mapping between the state variables of a dynamic system and control, and the second network, which is called the "critic," outputs the cost function or the value function with the state variables as its inputs. In the second version of the adaptive critic approach called the dual heuristic programming (DHP), the action network remains the same, but the second network, which is called the critic, outputs the costates with the state variables as its inputs. These ADP processes, through the nonlinear function approximation capabilities of NNs, overcome the computational complexity that plagued the dynamic programming formulation of optimal control problems. Werbos [52] further advanced two other versions called "action-dependent critics," namely, ADHDP and ADDHP, which add control also as inputs to the networks. An account of the various adaptive critic designs is found in [36]. A thorough treatment of neurodynamic programming is given in the seminal book by Bertsekas and Tsitsiklis [11].

Adaptive critic formulations attracted two schools of researchers. The first group consists of researchers who used ADP formulations to mainly solve problems with finite state spaces with applications to behavioral and computer sciences, operations research, and robotics (for example, [8], [9], [11], and [35]). Note that adaptive critics are also reinforcement learning designs [8], [9]. Many of these formulations essentially use Markov decision processes, and their problems primarily employ cost-function-based adaptive critics. A major reason is the fact that the value function derivatives are not well defined in a stochastic setting. The second group of researchers uses system science principles to formulate their problems with applications to real-time feedback control of dynamic systems for example (see [41] and [53] for multiple papers).

III. Stability Issues of ADP-Based Controllers

Howard [22] showed the convergence of an algorithm relying on the successive policy iteration solution of a nonlinear Lyapunov equation for the cost (value) and an optimizing equation for the control (action). This algorithm relies on perfect knowledge of the system dynamics and is an offline technique. Later, various online dynamic-programming-based reinforcement learning algorithms emerged and were mainly based on Werbos' HDP [52], Sutton's temporal difference (TD) learning methods [42], and Q-Learning, which was introduced by Watkins [48] and Werbos [50]. Critic and action network tuning was provided by recursive least squares, gradient techniques, or backpropagation algorithm [49]. Early work on dynamic-programming-based reinforcement learning focused on discrete finite state and action spaces. These depended on lookup tables or linear function approximators. Convergence results were shown in this case, such as [15].

For applications in finite state spaces, convergence of the training/learning algorithms for the ADP-based network structure and boundedness of the cost are the major critical points with respect to stability. There are several papers in current literature on this subject, such as [11], [18], and [44].

For continuous state and action spaces, convergence results are more challenging as adaptive critics require the use of nonlinear function approximators. More recently, Ferrari and Stengel [19] provide a proof of convergence of adaptive critic training on arguments similar to Howard's. The linear quadratic regulation (LQR) problem [29] served as a testbed for many of these studies. Solid convergence results were obtained for various adaptive critic designs for the LQR problem. We mention the work of Bradtke *et al.* [12] where Q-Learning was shown to converge when using the nonlinear function approximators. An important persistence of excitation notion was included. Further work was done by Landelius [26] who studied the four adaptive critic architectures. He demonstrated convergence results for all four cases in the LQR case and discussed when the design is model free. Prokhorov and Feldkamp [37] look at the Lyapunov stability analysis.

A. Model-Based Systems

It should be noted that for the use of ADP-based neurocontrollers for dynamic systems, treatment of stability issues depends on the needs of the system/plant. Consider a model-based ADP controller. Many system models can be derived from first principles or constitutive relations underlying the physics of the problem. If the need is to design a feedback controller for a deterministic model-based system, a controller resulting from a typical adaptive critic technique assures stability. This is because an ADP-based controller is basically an optimal controller. Optimal control guarantees stability with the nonexistence of conjugate points [13]. Model-based synthesis of adaptive-critic-based controllers presented by Balakrishnan and Biega [7], Prokhorov and Wunsch [36], and Venayagamoorthy *et al.* [45], for systems driven by ordinary differential equations, and by Padhi and Balakrishnan [33], for distributed parameter systems, demonstrates instances where the ADP-based controllers were shown to stabilize the plants quite successfully. Convergence of the training algorithm is the major issue in these cases as regards stability.

A convergent action network output guarantees closed-loop stability.

B. Model-Free and Uncertain Systems

Closed-loop stability for the ADP-based controllers for model-free and uncertain systems, however, is different. Model noise and exploration noise are discussed in [20] where adaptive critic is viewed as a stochastic approximation technique. Anderson *et al.* [4] showed convergence and stability for a reinforcement learning scheme. These results were done for the discrete-time case.

Q-Learning is not well posed when sampling times become small, and thus, it is not useful for extension to continuous-time systems. Continuous-time dynamic-programming-based reinforcement learning was reformulated by using the so-called advantage learning by Baird [6], who defines a differential increment from the optimal solution and explicitly takes into account the sampling interval Δt. Doya [16] derived results for online updating of the critic using techniques from continuous-time nonlinear optimal control. The advantage function follows naturally from this approach and, in fact, coincides with the continuous-time Hamiltonian function. Doya gave relations with the TD (0) and TD (λ) techniques of Sutton [42]. Murray *et al.* [31] proved the convergence of an algorithm that uses system state measurements to find the cost to go. An array of initial conditions is needed. Unknown plant dynamics in the linear case is confronted by estimating a matrix of state derivatives. The cost functional is shown to be a Lyapunov function and approximated by using either quadratic functions or a radial basis function NN. Saridis and Lee [39] showed the convergence of an offline algorithm relying on the successive iteration solution of a nonlinear Lyapunov equation for the cost (value) and an optimizing equation for the control (action). This is the continuous-time equivalent of Howard's [22] work. Applications of adaptive critics in the continuous-time domain were mainly done through the discretization and through the application of the well-established discrete-time results (e.g., [43]). Various continuous-time nondynamic reinforcement learnings were discussed by Campos and Lewis [14] and Rovithakis [38], who approximated a Lyapunov function derivative. In [58], the HJB equation of dynamic programming is approximated by a Riccati equation, and a suboptimal controller-based on NN feedback linearization is implemented with full stability and convergence proofs.

Anderson *et al.* [4], [5] investigated the stability of the ADP-based control combining ideas from reinforcement learning and robust control. Typical "robust control" techniques consider a range of uncertainties, and the resulting design is guaranteed to be stable to the predefined disturbance boundaries. Anderson *et al.* [5] use reinforcement learning with the integral quadratic constraints in developing neurocontrollers. By adding an adaptive reinforcement signal, conservatism with a typical robust control is mitigated. They provide some simple examples to demonstrate their concepts [5] and implement them in an HVAC system [5].

C. ADP-Based Techniques for Model-Free Systems

There has been great interest recently in "universal model-free controllers" that do not need a mathematical model of the controlled plant but mimic the functions of biological processes to learn about the systems that they are controlling online, so that performance improves automatically. Advantage of such controllers is the fact that the controller is portable between derivatives of similar plants. This is a huge factor in areas such as high-performance aircraft or missile where the controller redesign time and, therefore, the cost for different versions of the plants can be reduced dramatically. Enns and Si [17] present a lucid article on model-free approach to helicopter control. Recent works by Lewis *et al.* and Jagannathan *et al.* have been quite rigorous in theory and useful in practical applications. Jagannathan [40] has extended stability proofs for systems with observers in the feedback loop. Al-Tamimi *et al.* [1] use the HDP and the DHP structures to solve problems formulated with game theoretic notions. Their formulation leads to a forward-in-time reinforcement learning algorithm that converges to the Nash equilibrium of the corresponding zero-sum game. They have provided performance comparisons with an F-16 autopilot problem. Convergence proof of the algorithms has been given. While this paper requires a model, Al-Tamirni *et al.* [2], [3] extend these results to a model-free environment for linear systems [2] for the control of a power system generator. In this paper, they present online model-free adaptive critic schemes based on ADP to solve optimal control problems in both discrete- and continuous-time domains for linear systems with unknown dynamics. In the discrete-time case, the solution process leads to solving the underlying game algebraic Riccati equation (GARE) of the corresponding optimal control problem or zero-sum game. In the continuous-time domain, their ADP scheme solves the underlying ARE of the optimal control problem. They show that their continuous-time ADP scheme is nothing but a quasi-Newton method to solve the ARE. In both time domains, the adaptive critic algorithms are easy to initialize considering that initial policies are not required to be stabilizing. As with the model-based paper, the authors have proved the convergence of the presented algorithm.

Vrabie *et al.* [46] proposed a new policy iteration technique to solve online the continuous-time LQR problem for a partially model-free system (internal dynamics unknown). They present an online adaptive critic algorithm in which the actor performs continuous-time control, whereas the critic's correction of the actor's behavior is discrete in time until best performance is obtained. The critic evaluates the actor performance over a period of time and formulates it in a parameterized form. Policy update is a function of the critic's evaluation of the actor. Convergence of the proposed algorithm is established by proving equivalence with an established algorithm [25]. Numerical results using the short period dynamics of an F-16 aircraft are presented. In [47], the authors illustrate the same ideas in a power system application.

Shih *et al.* [40] formulated a novel reinforcement-learning-based output-adaptive NN controller to track a desired trajectory for a class of complex nonlinear discrete-time systems in the presence of bounded and unknown disturbances. The controller structure includes an observer for estimating states and the outputs, critic, and two action NNs for generating virtual and actual control inputs. The NN weights are adapted online to minimize a performance index. A Lyapunov function proves the uniformly ultimate boundedness (UUB) of the closed-loop tracking error, network weights, and observer estimation error.

Performance of this controller is evaluated on a spark ignition engine operating with high exhaust gas recirculation levels, and good experimental results are reported. Javaherian *et al.* [23] also reported good model approximation and performance results with DHP for an engine application.

Zheng and Jagannathan [57] investigated the use of NNs toward obtaining nearly optimal solutions to the control of nonlinear discrete-time systems. The method is based on least squares successive approximation solution of the generalized HJB (GHJB) equation. An NN is used to approximate the GHJB solution, and the authors show through examples, which include a planar two-link robot, that the control laws are optimal for linear systems and suboptimal for nonlinear systems.

Yang and Jagannathan [56] applied an adaptive-critic-based controller to an atomic force microscope-based force controller to push nanoparticles on the substrates. A block phase correlation-based algorithm is embedded into the controller for the compensation of the thermal drift which is considered as the main external uncertainty during nanomanipulation. They prove the convergence of the states and NN weight estimates. He and Jagannathan [21] presented a reinforcement learning scheme in discrete time for the NN controller, where the action generating NN learning is performed based on a critic-supplied performance measure. By using the Lyapunov approach and with novel weight updates, the UUB of the closed-loop tracking error and the weight estimates are shown. The adaptive critic NN does not require an offline learning phase, and the weights can be initialized at zero or randomly. It is shown via simulation that taking magnitude constraints on the input helps reduce transients. Simulation results justify the theoretical analysis.

Recently, an improvement and modification to the two-network ac architecture, which is called the "single network adaptive critic (SNAC)," has been presented [33]–[55]. This approach eliminates the action network. As a consequence, the SNAC architecture offers three potential advantages: a simpler architecture, lesser computational load (about half of the dual network ac algorithms), and no approximation error due to the eliminated network. The SNAC approach is applicable to a wide class of nonlinear systems where the optimal control (stationary) equation can be explicitly expressed in terms of the state and the costate variables. Most of the problems in aerospace, automobile, robotics, and other engineering disciplines can be characterized by nonlinear control-affine equations that yield such a relation. SNAC-based controllers yield excellent tracking results in applications to microelectromechanical systems, chemical reactor, and high-speed reentry problems. Padhi *et al.* [34] have proved that for linear systems (where the mapping between the costate at stage $k+1$ and the state at stage k is linear), the SNAC structure on convergence converges to the discrete Riccati equation.

IV. Conclusion

New and exciting results in the ADP-based controllers for systems with increased complexity are continually emerging. The breadth of applications is truly amazing. As systems get large, diverse, and complex, uncertainties in modeling go hand in hand. Hence, it is crucial that new ideas and formulations of closed-loop stability are generated. Guarantees on closed-loop stability will be the enabling piece that transforms the ADP-based controllers from simulations into implementations in various applications.

References

[1] A. Al-Tamimi, M. Abu-Khalaf, and F. L. Lewis, "Adaptive critic designs for discrete-time zero-sum games with application to H_∞ control," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 1, pp. 240–247, Feb. 2007.

[2] A. Al-Tamirni, F. L. Lewis, and Y. Wang, "Model-free H-infinity load-frequency controller design for power systems," in *Proc. IEEE Int. Symp. Intell. Control*, 2007, pp. 118–125.

[3] A. Al-Tamimi, D. Vrabie, M. Abu-Khalaf, and F. L. Lewis, "Model-free approximate dynamic programming schemes for linear systems," in *Proc. IJCNN*, 2007, pp. 371–378.

[4] C. W. Anderson, R. M. Kretchner, P. M. Young, and D. C. Hittle, "Robust reinforcement learning control with static and dynamic stability," *Int. J. Robust Nonlinear Control*, vol. 11, pp. 1469–1500, 2001.

[5] C. W. Anderson, D. Hittle, M. Kretchmar, and P. Young, "Robust reinforcement learning for heating, ventilation, and air conditioning control of buildings," in *Handbook of Learning and Approximate Dynamic Programming*. New York: Wiley-IEEE Press, Aug. 2004, ch. 20.

[6] L. Baird, "Reinforcement learning in continuous time: Advantage updating," in *Proc. Int. Conf. Neural Netw.*, Orlando, FL, Jun. 1994, pp. 2448–2453.

[7] S. N. Balakrishnan and V. Biega, "Adaptive critic based neural networks for aircraft optimal control," *AIAA J. Guid. Control Dyn.*, vol. 19, no. 4, pp. 731–739, 1996.

[8] A. G. Barto, "Connectionist learning for control," in *Neural Networks for Control*. Cambridge, MA: MIT Press, 1991.

[9] A. G. Barto and T. G. Dieterich, "Reinforcement learning and its relation to supervised learning," in *Handbook of Learning and Approximate Dynamic Programming*. New York: Wiley-IEEE Press, Aug. 2004, ch. 1.

[10] L. von Bertalanffy, *General System Theory*. New York: Braziller, 1968.

[11] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific, 1996.

[12] S. Bradtke, B. Ydstie, and A. Barto, "Adaptive linear quadratic control using policy iteration," Univ. Massachusetts, Amherst, MA, pp. 3475–3479, CMPSCI-94-49, Jun. 1994.

[13] A. E. Bryson and Y. C. Ho, *Applied Optimal Control: Optimization, Estimation and Control*. Waltham, MA: Ginn & Company, 1969.

[14] J. Campos and F. L. Lewis, "Adaptive critic neural network for feedforward compensation," in *Proc. Amer. Control Conf.*, San Diego, CA, Jun. 1999, pp. 2813–2818.

[15] P. Dayan, "The convergence of TD (λ) for general λ," *Mach. Learn.*, vol. 8, no. 3/4, pp. 341–362, May 1992.

[16] K. Doya, "Reinforcement learning in continuous time and space," in *Neural Computation*, vol. 12. Cambridge, MA: MIT Press, 2000, pp. 219–245.

[17] R. Enns and J. Si, "Apache helicopter stabilization using neural dynamic programming," *AIAA J. Guid. Control Dyn.*, vol. 25, no. 1, pp. 19–25, 2002.

[18] D. P. Farias, "The linear programming approach to approximate dynamic programming," in *Handbook of Learning and Approximate Dynamic Programming*. New York: Wiley-IEEE Press, Aug. 2004, ch. 6.

[19] S. Ferrari and R. Stengel, "An adaptive critic global controller," in *Proc. Amer. Control Conf.*, 2002, pp. 2665–2670.

[20] S. Hagen and B. Kröse, "Linear quadratic regulation using reinforcement learning," in *Proc. 8th Belgian–Dutch Conf. Mach. Learn.*, F. Verdenius and W. van den Broek, Eds., Oct. 1998, pp. 39–46.

[21] P. He and S. Jagannathan, "Reinforcement learning neural-network-based controller for nonlinear discrete-time systems with input constraints," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 2, pp. 425–436, Apr. 2007.

[22] R. Howard, *Dynamic Programming and Markov Processes*. Cambridge, MA: MIT Press, 1960.

[23] H. Javaherian, D. Liu, and O. Kovalenko, "Automotive engine torque and air–fuel ratio control using dual heuristic dynamic programming," in *Proc. IJCNN*, 2006, pp. 518–525.

[24] H. Khalil, *Nonlinear Systems*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[25] D. L. Kleinman, "On an iterative technique for Riccati equation computations," *IEEE Trans. Autom. Control*, vol. AC-13, no. 1, pp. 114–115, Feb. 1968.

[26] T. Landelius, "Reinforcement learning and distributed local model synthesis," Ph.D. dissertation, Linköping Univ., Linköping, Sweden, 1997.

[27] F. L. Lewis and S. S. Ge, "Neural networks in feedback control," in *Mechanical Engineers' Handbook*. New York: Wiley, 2006, ch. 19.
[28] F. L. Lewis, J. P. Huang, D. V. T. Prokhorov, and D. C. Wunsch, "Guest editorial special issue on neural nets for feedback control systems," *IEEE Trans. Neural Netw.*, vol. 18, no. 4, pp. 969–972, Jul. 2007.
[29] F. L. Lewis and V. Syrmos, *Optimal Control*, 2nd ed. New York: Wiley, 1995.
[30] *Neural Networks for Control*, MIT Press, Cambridge, MA, 1991.
[31] 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.
[32] K. S. Narendra and F. L. Lewis, "Introduction to the special issue on neural network feedback control," *Automatica*, vol. 37, no. 8, pp. 1147–1148, Aug. 2001.
[33] R. Padhi and S. N. Balakrishnan, "Optimal management of beaver population using a reduced-order distributed parameter model and single network adaptive critics," *IEEE Trans. Control Syst. Technol.*, vol. 14, no. 4, pp. 628–640, Jul. 2006.
[34] R. Padhi, N. Unnikrishnan, X. Wang, and S. N. Balakrishnan, "A single network adaptive critic (SNAC) architecture for optimal control synthesis for a class of nonlinear systems," *Neural Netw.*, vol. 19, no. 10, pp. 1648–1660, Dec. 2006.
[35] W. Powell and B. Van Roy, "ADP for high-dimensional resource allocation problems," in *Handbook of Learning and Approximate Dynamic Programming*. New York: Wiley-IEEE Press, Aug. 2004, ch. 10.
[36] D. V. Prokhorov and D. C. Wunsch, II, "Adaptive critic designs," *IEEE Trans. Neural Netw.*, vol. 8, no. 5, pp. 997–1007, Sep. 1997.
[37] D. V. Prokhorov and L. A. Feldkamp, "Analyzing for Lyapunov stability with adaptive critics," in *Proc. Int. Conf. Syst., Man, Cybern.*, 1998, pp. 1658–1661.
[38] G. A. Rovithakis, "Stable adaptive neuro-control design via Lyapunov function derivative estimation," *Automatica*, vol. 37, no. 8, pp. 1213–1221, Aug. 2001.
[39] G. Saridis and C. S. Lee, "An approximation theory of optimal control for trainable manipulators," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 3, pp. 152–159, Mar. 1979.
[40] P. Shih, B. Kaul, S. Jagannathan, and J. Drallmeier, "Near optimal output-feedback control of nonlinear discrete-time systems in nonstrict feedback form with application to engines," in *Proc. IJCNN Conf.*, 2007, pp. 396–401.
[41] J. Si, P. A. G. Barto, and W. B. D. Wunsch, Eds., *Handbook of Learning and Approximate Dynamic Programming*, IEEE Press Series on Computational Intelligence. New York: Wiley-IEEE Press, Aug. 2004.
[42] R. Sutton, "Learning to predict by the methods of temporal differences," *Mach. Learn.*, vol. 3, no. 1, pp. 9–44, Aug. 1988.
[43] J. N. Tsitsiklis, "Efficient algorithms for globally optimal trajectories," *IEEE Trans. Autom. Control*, vol. 40, no. 9, pp. 1528–1538, Sep. 1995.
[44] B. Van Roy, D. P. Bertsekas, Y. Lee, and J. N. Tsitsiklis, "A neuro-dynamic programming approach to retailer inventory management," in *Proc. IEEE Conf. Decision Control*, San Diego, CA, 1997, pp. 4052–4057.
[45] G. K. Venayagamoorthy, R. G. Harley, and D. C. Wunsch, "Dual heuristic programming excitation neurocontrol for generators in a multimachine power system," *IEEE Trans. Ind. Appl.*, vol. 39, no. 2, pp. 382–384, Mar./Apr. 2003.
[46] D. Vrabie, O. Pastravanu, and F. L. Lewis, "Policy iteration for continuous-time systems with unknown internal dynamics," in *Proc. Control Autom.*, 2007, pp. 1–6.
[47] D. Vrabie, M. Abu-Khalaf, F. L. Lewis, and Y. Wang, "Continuous-time ADP for linear systems with partially unknown dynamics," in *Proc. IEEE Int. Symp. ADPRL*, 2007, pp. 247–253.
[48] C. Watkins, "Learning from delayed rewards," Ph.D. dissertation, Cambridge Univ., Cambridge, U.K., 1989.
[49] P. J. Werbos, "Beyond regression: New tools for prediction and analysis in the behavior sciences," Ph.D. dissertation, Committee Appl. Math., Harvard Univ., Cambridge, MA, 1974.
[50] P. J. Werbos, "Neural networks for control and system identification," in *Proc. IEEE Conf. Decision Control*, Tampa, FL, 1989, vol. 1, pp. 260–265.
[51] 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: MIT Press, 1991, pp. 67–95.
[52] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," in *Handbook of Intelligent Control*, D. A. White and D. A. Sofge, Eds. New York: Van Nostrand Reinhold, 1992, ch. 13.
[53] D. A. White and D. A. Sofge, Eds., "Applied learning-optimal control for manufacturing," in *Handbook of Intelligent Control*. New York: Van Nostrand Reinhold, 1992, ch. 9.
[54] A. N. Whitehead, *The Concept of Nature*. Cambridge, U.K.: Cambridge Univ. Press, 1920.
[55] V. Yadav, R. Padhi, and S. N. Balakrishnan, "Robust/optimal temperature profile control of a high-speed aerospace vehicle using neural networks," *IEEE Trans. Neural Netw.*, vol. 18, no. 4, pp. 1115–1128, Jul. 2007.
[56] Q. Yang and S. Jagannathan, "Adaptive critic neural network force controller for atomic force microscope-based nanomanipulation," in *Proc. IEEE Int. Symp. Intell. Control*, 2006, pp. 464–469.
[57] C. Zheng and S. Jagannathan, "Generalized Hamilton–Jacobi–Bellman formulation-based neural network control of affine nonlinear discrete-time systems," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 90–106, Jan. 2008.
[58] Y. H. Kim and F. L. Lewis, "High-level feedback control with neural networks," *World Scientific*, Singapore, 1998.

S. N. Balakrishnan (M'05) received the Ph.D. degree from the University of Texas, Austin.

He is currently a Professor with the Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla. His research interests include neural networks, optimal control, and large-scale and impulse systems. His papers from the development of techniques in these areas include applications to missiles, spacecraft, aircraft, robotics, temperature, and animal population control.

Jie Ding is currently working toward the Ph.D. degree in the Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla.

Her research interests include neural networks and optimal control.

Frank L. Lewis (S'78–M'81–SM'86–F'94) received the B.S. degree in physics/electrical engineering and the M.S. degree in electrical engineering from Rice University, Houston, TX, the M.S. degree in aeronautical engineering from the University of West Florida, Pensacola, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta.

He is currently a Distinguished Scholar Professor and the Moncrief-O'Donnell Chair with the Automation and Robotics Research Institute, University of Texas at Arlington, Fort Worth. He works in feedback control, intelligent systems, and sensor networks. He is the holder of five U.S. patents and the author of 186 journal papers, 305 conference papers, and 12 books.

Dr. Lewis is a Fellow of the International Federation of Automatic Control, a Fellow of the U.K. Institute of Measurement and Control, a Professional Engineer in Texas, and a U.K. Chartered Engineer. He was the Founding Member of the Board of Governors of the Mediterranean Control Association. He was appointed to the NAE Committee on Space Station in 1995. He is an elected Guest Consulting Professor at both the South China University of Technology, Guangzhou, and Shanghai Jiao Tong University, Shanghai, China. He received the Fulbright Research Award, the NSF Research Initiation Grant, the ASEE Terman Award, and the Outstanding Service Award from the Dallas IEEE Section. He was selected as the Engineer of the Year by the Forth Worth IEEE Section and listed in the Forth Worth Business Press Top 200 Leaders in Manufacturing. He helped win the IEEE Control Systems Society Best Chapter Award (as the Founding Chairman of DFW Chapter), the National Sigma Xi Award for Outstanding Chapter (as the President of UTA Chapter), and the U.S. SBA Tibbets Award in 1996 (as the Director of ARRI's SBIR Program).