Robust Adaptive Dynamic Programming for Optimal Nonlinear Control Design

Yu Jiang and Zhong-Ping Jiang
Department of Electrical and Computer Engineering
Polytechnic Institute of New York University, Brooklyn, NY 11201
Email: yjiang06@students.poly.edu, zjiang@poly.edu

Abstract—**This paper studies the robust optimal control design for uncertain nonlinear systems from a perspective of robust adaptive dynamic programming (robust-ADP). The objective is to fill up a gap in the past literature of ADP where dynamic uncertainties or unmodeled dynamics are not addressed. A key strategy is to integrate tools from modern nonlinear control theory, such as the robust redesign and the backstepping techniques as well as the nonlinear small-gain theorem, with the theory of ADP. The proposed robust-ADP methodology can be viewed as a natural extension of ADP to uncertain nonlinear systems. A practical learning algorithm is developed in this paper, and has been applied to a sensorimotor control problem.**

I. INTRODUCTION

Reinforcement learning (RL) [30] is an important branch in machine learning theory. It is concerned with how an agent should modify its actions based on the reward from its reactive unknown environment so as to achieve a long term goal. In 1968, Werbos pointed out that the policy iteration technique devised in [6] for dynamic programming can be employed to perform RL [34]. Starting from then, many real-time RL methods for finding online optimal control policies have emerged and they are broadly called approximate/adaptive dynamic programming (ADP) [35], [36] or neurodynamic programming [5]. See [1], [2], [7], [22], [24], [31], [32], [33], for some recently developed results.

In the past literature of ADP, it is commonly assumed that the system order is known and the state variables are either fully available or reconstructible from the output; see [21], [22] and reference therein. However, in practice, the system order may be unknown due to the presence of dynamic uncertainties (or unmodeled dynamics), which are motivated by engineering applications in situations where the exact mathematical model of a physical system is not easy to be obtained. Of course, dynamic uncertainties also make sense for the mathematical modeling in other branches of science such as biology and economics. This problem, often formulated as robust control, cannot be viewed as a special case of output feedback control, and the ADP methods developed in the past literature may not only fail to guarantee optimality, but also the stability of the closed-loop system when dynamic uncertainty occurs. To fill up the above-mentioned gap in the past literature of

ADP, we recently developed a new theory of robust adaptive dynamic programming (robust-ADP) [8], [10], [11], which can be viewed as a natural extension of ADP to linear and partially linear systems with dynamic uncertainties.

The primary objective of this paper is to study robust-ADP designs for genuinely nonlinear systems in the presence of dynamic uncertainties. We first decompose the open-loop system into two parts: The *system model* (ideal environment) with known system order and fully accessible state, and the *dynamic uncertainty*, with unknown system order and dynamics, interacting with the ideal environment. In order to handle the dynamic interaction between two systems, we then resort to the gain assignment idea [14], [15], [26]. More specifically, we need to assign a suitable gain for the system model with disturbance in the sense of Sontag's input-to-state stability (ISS) [29]. The backstepping, robust redesign, and small-gain techniques in modern nonlinear control theory are incorporated into the robust-ADP theory, such that the system model is made ISS with an arbitrarily small gain. At last, the nonlinear small-gain theorem [15] is applied to analyze the stability for the interconnected systems.

Throughout this paper, vertical bars $|\cdot|$ represent the Euclidean norm for vectors, or the induced matrix norm for matrices. For any piecewise continuous function u, $\|u\|$ denotes $\sup\{|u(t)|, t \geq 0\}$. A function $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be of class $\mathcal{K}$ if it is continuous, strictly increasing with $\gamma(0) = 0$. It is of class $\mathcal{K}_\infty$ if additionally $\gamma(s) \to \infty$ as $s \to \infty$. A function $\beta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is of class $\mathcal{KL}$ if $\beta(\cdot, t)$ is of class $\mathcal{K}$ for every fixed $t \geq 0$, and $\beta(s, t) \to 0$ as $t \to \infty$ for each fixed $s \geq 0$. The notation $\gamma_1 > \gamma_2$ means $\gamma_1(s) > \gamma_2(s)$, $\forall s > 0$.

II. PRELIMINARIES

In this section, let us review a policy iteration technique to solve optimal control problems [27].

To begin with, consider the system

$$\dot{x} = f(x) + g(x)u \tag{1}$$

where $x \in \mathbb{R}^n$ is the system state, $u \in \mathbb{R}$ is the control input, $f, g : \mathbb{R}^n \to \mathbb{R}^n$ are locally Lipschitz functions. For any initial condition $x_0 \in \mathbb{R}^n$, the cost function associated with (1) is

defined as

$$J(x_0) = \int_0^\infty \left[Q(x) + ru^2\right] dt, \quad x(0) = x_0 \qquad (2)$$

where $Q(\cdot)$ is a positive definite function, and $r > 0$ is a constant. In addition, assume there exists an *admissible* control policy $u = u_0(x)$ in the sense that, under this policy, the system (1) is globally asymptotically stable and the cost (2) is finite. By [20], the control policy that minimizes the cost (2) can be solved from the following Hamilton-Jacobi-Bellman (HJB) equation:

$$0 = \nabla V(x)f(x) + Q(x) - \frac{1}{4r}\left[\nabla V(x)g(x)\right]^2 \qquad (3)$$

with the boundary condition $V(0) = 0$. Indeed, if the solution $V^*(x)$ of (3) exists, the optimal control policy is given by

$$u^*(x) = -\frac{1}{2r}g(x)^T \nabla V^*(x)^T. \qquad (4)$$

In general, the analytical solution of (3) is difficult to be solved. However, if $V^*(x)$ exists, it can be approximated using the policy iteration technique [27]:

1) Find an admissible control policy $u_0(x)$.
2) For any integer $i \geq 0$, solve for $V_i(x)$, with $V_i(0) = 0$, using

$$0 = \nabla V_i(x)\left[f(x) + g(x)u_i(x)\right] + Q(x) + ru_i(x)^2. \qquad (5)$$

3) Update the control policy using

$$u_{i+1}(x) = -\frac{1}{2r}g(x)^T \nabla V_i(x)^T. \qquad (6)$$

Convergence of the policy iteration (5) and (6) is concluded in the following theorem, which can be seen as a trivial extension of Theorem 4 in [27].

Theorem 2.1: Consider $V_i(x)$ and $u_i(x)$ defined in (5) and (6). Then, for all $i = 0, 1, \cdots$,

$$V_{i+1}(x) \leq V_i(x), \quad \forall x \in R^n \qquad (7)$$

and $u_i(x)$ is admissible. In addition, if the solution $V^*(x)$ of (3) exists, then for each fixed x, $V_i(x)$ and $u_i(x)$ converge pointwise to $V^*(x)$ and $u^*(x)$, respectively.

III. ONLINE LEARNING VIA ROBUST-ADP

In this section, we develop the robust-ADP methodology for nonlinear systems as follows:

$$\dot{w} = \Delta_w(w, x) \qquad (8)$$
$$\dot{x} = f(x) + g(x)\left[u + \Delta(w, x)\right] \qquad (9)$$

where $x \in \mathbb{R}^n$ is the measured component of the state available for feedback control, $w \in \mathbb{R}^p$ is the unmeasurable part of the state with unknown order p, $u \in \mathbb{R}$ is the control input, $\Delta_w : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}^p$, $\Delta : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}$ are unknown locally Lipschitz functions, f and g are defined the same as in (1) but are assumed to be unknown.

Our design objective is to find online the control policy which stabilizes the system at the origin. Also, in the absence of the dynamic uncertainty (i.e., $\Delta = 0$ and the w-subsystem is absent), the control policy becomes the optimal control policy that minimizes (2).

A. Online policy iteration

The iterative technique introduced in Section 2 relies on the knowledge of $f(x)$ and $g(x)$. To remove this requirement, we develop a novel online policy iteration technique, which can be viewed as the nonlinear extension of [7].

To begin with, notice that (9) can be rewritten as

$$\dot{x} = f(x) + g(x)u_i(x) + g(x)v_i \qquad (10)$$

where $v_i = u + \Delta - u_i$. For each $i \geq 0$, the time derivative of $V_i(x)$ along the solutions of (10) satisfies

$$\dot{V}_i(x) = -Q(x) - ru_i^2(x) - 2ru_{i+1}(x)v_i. \qquad (11)$$

Integrating both sides of (11) on any time interval $[t, t+T]$, it follows that

$$V_i(x(t+T)) - V_i(x(t))$$
$$= \int_t^{t+T}\left[-Q(x) - ru_i^2(x) - 2ru_{i+1}(x)v_i\right]dt. \qquad (12)$$

Notice that if $u_i(x)$ is given, the unknown functions $V_i(x)$ and $u_{i+1}(x)$ can be approximated using (12). To be more specific, for any given compact set $\Omega \subset \mathbb{R}^n$ containing the origin as an interior point, let $\{\phi_j(x)\}_{j=1}^\infty$ be an infinite sequence of linearly independent smooth basis functions on Ω, where $\phi_j(0) = 0$ for all $j = 1, 2, \cdots$. Then, for each $i = 0, 1, \cdots$, the cost function and the control policy are approximated by $\hat{V}_i(x) = \sum_{j=1}^{N_1} \hat{c}_{i,j}\phi_j(x)$, and $\hat{u}_{i+1}(x) = \sum_{j=1}^{N_2} \hat{w}_{i,j}\phi_j(x)$, respectively, where $N_1 > 0$, $N_2 > 0$ are two sufficiently large integers, and $\hat{c}_{i,j}$, $\hat{w}_{i,j}$ are constant weights to be determined.

Replacing $V_i(x)$, $u_i(x)$, and $u_{i+1}(x)$ in (12) with their approximations, we obtain

$$\sum_{j=1}^{N_1} \hat{c}_{i,j}\left[\phi_j(x(t_{k+1})) - \phi_j(x(t_k))\right]$$
$$= -\int_{t_k}^{t_{k+1}} 2r\sum_{j=1}^{N_2}\hat{w}_{i,j}\phi_j(x)\hat{v}_i dt \qquad (13)$$
$$-\int_{t_k}^{t_{k+1}}\left[Q(x) + r\hat{u}_i^2(x)\right]dt + e_{i,k}$$

where $\hat{u}_0 = u_0$, $\hat{v}_i = u + \Delta - \hat{u}_i$, and $\{t_k\}_{k=0}^l$ is a strictly increasing sequence with $l > 0$ a sufficiently large integer. Then, the weights $\hat{c}_{i,j}$ and $\hat{w}_{i,j}$ can be solved in the sense of least-squares (i.e., by minimizing $\sum_{k=0}^l e_{i,k}^2$).

Now, starting from $u_0(x)$, two sequences $\{\hat{V}_i(x)\}_{i=0}^\infty$, and $\{\hat{u}_{i+1}(x)\}_{i=0}^\infty$ can be generated via the online policy iteration technique (13). Next, we show the convergence of the sequences to $V_i(x)$ and $u_{i+1}(x)$, respectively.

Assumption 3.1: There exist $l_0 > 0$ and $\delta > 0$, such that for all $l \geq l_0$, we have

$$\frac{1}{l}\sum_{k=0}^l \theta_{i,k}^T\theta_{i,k} \geq \delta I_{N_1+N_2} \qquad (14)$$

where

$$
\theta_{i,k}^T = \begin{bmatrix} \phi_1(x(t_{k+1})) - \phi_1(x(t_k)) \\ \phi_2(x(t_{k+1})) - \phi_2(x(t_k)) \\ \vdots \\ \phi_{N_1}(x(t_{k+1})) - \phi_{N_1}(x(t_k)) \\ 2r \int_{t_k}^{t_{k+1}} \phi_1(x)\hat{v}_i(x)dt \\ 2r \int_{t_k}^{t_{k+1}} \phi_2(x)\hat{v}_i(x)dt \\ \vdots \\ 2r \int_{t_k}^{t_{k+1}} \phi_{N_2}(x)\hat{v}_i(x)dt \end{bmatrix} \in \mathbb{R}^{N_1+N_2}.
$$

Assumption 3.2: For all $t \geq 0$, we have $x(t) \in \Omega$.

Notice that, Assumption 3.2 is not very restrictive and can be satisfied if Ω is an invariant set for the x-subsystem. This issue will be further elaborated in Section V.

Theorem 3.1: Under Assumptions 3.1 and 3.2, for each $i \geq 0$, we have

$$
\lim_{N_1,N_2 \to \infty} \hat{V}_i(x) = V_i(x), \tag{15}
$$

$$
\lim_{N_1,N_2 \to \infty} \hat{u}_{i+1}(x) = u_{i+1}(x), \tag{16}
$$

for all $x \in \Omega$.

Proof: See the Appendix. ∎

Corollary 3.1: Under Assumptions 3.1 and 3.2, for any arbitrary $\epsilon > 0$, there exist integers $i^* > 0$, $N_1^* > 0$ and $N_2^* > 0$, such that

$$
|\hat{V}_i(x) - V^*(x)| \leq \epsilon, \quad \text{and} \quad |\hat{u}_{i+1}(x) - u^*(x)| \leq \epsilon,
$$

for all $x \in \Omega$, if $i > i^*$, $N_1 > N_1^*$, and $N_2 > N_2^*$.

B. Robust redesign

In the presence of the dynamic uncertainty, we redesign the approximated optimal control policy so as to achieve asymptotic stability. This method is an integration of optimal control theory [20] with the gain assignment technique [15], [26]. To begin with, let us assume the following:

Assumption 3.3: There exists a function $\underline{\alpha}$ of class $\mathcal{K}_\infty$, such that for $i = 0, 1, \cdots$,

$$
\underline{\alpha}(|x|) \leq V_i(x), \quad \forall x \in \mathbb{R}^n. \tag{17}
$$

In addition, assume there exists a constant $\epsilon > 0$ such that $Q(x) - \epsilon^2 |x|^2$ is a positive definite function.

Notice that, we can also find a class $\mathcal{K}_\infty$ function $\bar{\alpha}$, such that for $i = 0, 1, \cdots$,

$$
V_i(x) \leq \bar{\alpha}(|x|), \quad \forall x \in \mathbb{R}^n. \tag{18}
$$

Assumption 3.4: Consider (8). There exist functions $\underline{\lambda}, \bar{\lambda} \in \mathcal{K}_\infty$, $\kappa_1, \kappa_2, \kappa_3 \in \mathcal{K}$, and positive definite functions W and κ_4, such that for all $w \in \mathbb{R}^p$ and $x \in \mathbb{R}^n$, we have

$$
\underline{\lambda}(|w|) \leq W(w) \leq \bar{\lambda}(|w|), \tag{19}
$$

$$
|\Delta(w,x)| \leq \max\{\kappa_1(|w|), \kappa_2(|x|)\}, \tag{20}
$$

together with the following implication:

$$
W(w) \geq \kappa_3(|x|) \Rightarrow \nabla W(w)\Delta_w(w,x) \leq -\kappa_4(w). \tag{21}
$$

Assumption 3.4 implies that the w-system (8) is input-to-state stable (ISS) [29] when x is considered as the input, i.e.,

$$
|w(t)| \leq \beta_w(|w(0)|, t) + \gamma_w(\|x\|) \tag{22}
$$

where β_w is of class $\mathcal{K}\mathcal{L}$ and γ_w is of class $\mathcal{K}$.

Now, consider the following type of control policy

$$
u_{ro}(x) = \left[1 + \frac{r}{2}\rho^2(|x|^2)\right] \hat{u}_{i^*+1}(x) \tag{23}
$$

where $i^* > 0$ is a sufficiently large integer as defined in Corollary 3.1, ρ is a smooth, non decreasing function, with $\rho(s) > 0$ for all $s \geq 0$. Notice that u_{ro} can be viewed as a robust redesign of the approximated optimal control law $\hat{u}_{i^*+1}$.

As in [14], let us define a class $\mathcal{K}_\infty$ function γ by

$$
\gamma(s) = \frac{1}{2}\epsilon\rho(s^2)s, \quad \forall s \geq 0. \tag{24}
$$

In addition, define

$$
\begin{aligned}
e_{ro}(x) &= \frac{r}{2}\rho^2(|x|^2)\left[\hat{u}_{i^*+1}(x) - u_{i^*+1}(x)\right] \\
&\quad + \hat{u}_{i^*+1}(x) - u_{i^*}(x).
\end{aligned} \tag{25}
$$

Theorem 3.2: Under Assumptions 3.3 and 3.4, suppose

$$
\gamma > \max\{\kappa_2, \kappa_1 \circ \underline{\lambda}^{-1} \circ \kappa_3 \circ \underline{\alpha}^{-1} \circ \bar{\alpha}\}, \tag{26}
$$

and the following implication holds for some constant $d > 0$:

$$
0 < V_{i^*}(x) \leq d \Rightarrow |e_{ro}(x)| < \gamma(|x|). \tag{27}
$$

Then, the closed-loop system composed of (8), (9), and (23) is asymptotically stable at the origin. In addition, there exists $\sigma \in \mathcal{K}_\infty$, such that $\Omega_{i^*} = \{(w,x) : \max[\sigma(V_{i^*}(x)), W(w)] \leq \sigma(d)\}$ is an estimate of the region of attraction of the closed-loop system.

Proof: See the Appendix. ∎

Remark 3.1: In the absence of the dynamic uncertainty (i.e., $\Delta = 0$ and the w-system is absent), the control policy (23) can be replaced by $\hat{u}_{i^*+1}(x)$, which is an approximation of the optimal control policy $u^*(x)$ that minimizes the following cost function

$$
J(x_0) = \int_0^\infty \left[Q(x) + ru^2\right] dt, \quad x(0) = x_0. \tag{28}
$$

IV. ROBUST-ADP WITH UNMATCHED DYNAMIC UNCERTAINTY

In this section, we extend the robust-ADP methodology to nonlinear systems with unmatched dynamic uncertainties. To begin with, consider the system:

$$
\dot{w} = \Delta_w(w,x) \tag{29}
$$

$$
\dot{x} = f(x) + g(x)\left[z + \Delta(w,x)\right] \tag{30}
$$

$$
\dot{z} = f_1(x,z) + u + \Delta_1(w,x,z) \tag{31}
$$

where $[x^T, z]^T \in \mathbb{R}^n \times \mathbb{R}$ is the measured component of the state available for feedback control; w, u, Δ_w, f, g, and Δ are defined in the same way as in (8)-(9); $f_1 : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ and $\Delta_1 : \mathbb{R}^p \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ are locally Lipschitz functions and are assumed to be unknown.

Assumption 4.1: There exist class $\mathcal{K}$ functions $\kappa_5, \kappa_6, \kappa_7$, such that the following inequality holds:

$$|\Delta_1(w, x, z)| \leq \max\{\kappa_5(|w|), \kappa_6(|x|), \kappa_7(|z|)\}. \quad (32)$$

A. Online learning

Let us define a virtual control policy $\xi = u_{ro}$, as defined in (23). Then, a state transformation can be performed as $\zeta = z - \xi$. Along the trajectories of (30)-(31), it follows that

$$\dot{\zeta} = \bar{f}_1(x, z) + u + \Delta_1 - \bar{g}_1(x)\Delta \quad (33)$$

where $\bar{f}_1(x, z) = f_1(x, z) - \frac{\partial \xi}{\partial x} f(x) - \frac{\partial \xi}{\partial x} g(x)z$, and $\bar{g}_1(x) = \frac{\partial \xi}{\partial x} g(x)$ are two unknown functions that can be approximated by $\hat{f}_1(x, z) = \sum_{j=1}^{N_3} \hat{w}_{f,j} \psi_j(x, z)$ and $\hat{g}_1(x) = \sum_{j=0}^{N_4-1} \hat{w}_{g,j} \phi_j(x)$, respectively, where $\{\psi_j(x, z)\}_{j=1}^{\infty}$ is a sequence of linearly independent basis functions on some compact set $\Omega_1 \in \mathbb{R}^{n+1}$ containing the origin as its interior, $\phi_0(x) \equiv 1$, $\hat{w}_{f,j}$ and $\hat{w}_{g,j}$ are constant weights to be trained. As in the matched case, Ω_1 is selected to be an invariant set for the system (30) and (31).

1) Phase-one learning: To approximate the virtual control input ξ for the x-subsystem, the same procedure as in (13) can be applied, with $\hat{v}_i = z + \Delta - \hat{u}_i$.

2) Phase two learning: To approximate the unknown functions $\bar{f}_1$ and $\bar{g}_1$, The constant weights can be solved, in the sense of least-squares, from

$$\frac{1}{2}\zeta^2(t'_{k+1}) - \frac{1}{2}\zeta^2(t'_k)$$
$$= \int_{t'_k}^{t'_{k+1}} \left[\sum_{j=1}^{N_3} \hat{w}_{f,j}\psi_j(x,z) - \sum_{j=0}^{N_4-1} \hat{w}_{g,j}\phi_j(x)\Delta \right] \zeta dt$$
$$+ \int_{t'_k}^{t'_{k+1}} (u + \Delta_1)\zeta dt + \bar{e}_k \quad (34)$$

where $\{t'_k\}_{k=1}^{l}$ is a strictly increasing positive constant sequence with $l > 0$ a sufficiently large integer, and $\bar{e}_k$ denotes the approximation error. Similarly as in the previous section, let us introduce the following assumption:

Assumption 4.2: There exist $l_1 > 0$ and $\delta_1 > 0$, such that for all $l \geq l_1$, we have

$$\frac{1}{l} \sum_{k=0}^{l} \bar{\theta}_k^T \bar{\theta}_k \geq \delta_1 I_{N_3+N_4} \quad (35)$$

where

$$\bar{\theta}_k^T = \begin{bmatrix} \int_{t'_k}^{t'_{k+1}} \psi_1(x, z)\zeta dt \\ \int_{t'_k}^{t'_{k+1}} \psi_2(x, z)\zeta dt \\ \vdots \\ \int_{t'_k}^{t'_{k+1}} \psi_{N_3}(x, z)\zeta dt \\ \int_{t'_k}^{t'_{k+1}} \phi_0(x)\Delta\zeta dt \\ \int_{t'_k}^{t'_{k+1}} \phi_1(x)\Delta\zeta dt \\ \vdots \\ \int_{t'_k}^{t'_{k+1}} \phi_{N_4-1}(x)\Delta\zeta dt \end{bmatrix} \in \mathbb{R}^{N_3+N_4}.$$

Theorem 4.1: Consider $(x(0), z(0)) \in \Omega_1$. Then, under Assumption 4.2 we have

$$\lim_{N_3, N_4 \to \infty} \hat{f}(x, z) = \bar{f}_1(x, z), \quad (36)$$
$$\lim_{N_3, N_4 \to \infty} \hat{g}(x) = \bar{g}_1(x), \quad \forall (x, z) \in \Omega_1. \quad (37)$$

Theorem 4.1 can be proved following the same idea as in the proof of Theorem 3.1, and is omitted here for want of space.

B. Robust redesign

Next, we study robust stabilization of the system (29)-(31). To this end, let κ_8 be a function of $\mathcal{K}$ such that

$$\kappa_8(|x|) \geq |\xi(x)|, \quad \forall x \in \mathbb{R}^n. \quad (38)$$

Then, Assumption 4.1 implies

$$|\Delta_1| \leq \max\{\kappa_5(|w|), \kappa_6(|x|), \kappa_7(|z|)\}$$
$$\leq \max\{\kappa_5(|w|), \kappa_6(|x|), \kappa_7(|\xi| + \kappa_8(|x|))\}$$
$$\leq \max\{\kappa_5(|w|), \kappa_9(|X_1|)\}$$

where $\kappa_9(s) = \max\{\kappa_6, \kappa_7 \circ \kappa_8 \circ (2s), \kappa_7 \circ (2s)\}, \forall s \geq 0$. In addition, we denote $\tilde{\kappa}_1 = \max\{\kappa_1, \kappa_5\}$, $\tilde{\kappa}_2 = \max\{\kappa_2, \kappa_9\}$, $\gamma_1(s) = \frac{1}{2}\epsilon\rho(\frac{1}{2}s^2)s$, and

$$U_{i^*}(X_1) = V_{i^*}(x) + \frac{1}{2}\zeta^2. \quad (39)$$

Notice that, under Assumptions 3.3 and 3.4, there exist $\bar{\alpha}_1, \underline{\alpha}_1 \in \mathcal{K}_\infty$, such that $\underline{\alpha}_1(|X_1|) \leq U_{i^*}(X_1) \leq \bar{\alpha}_1(|X_1|)$.

The control policy can be approximated by

$$u_{ro1} = -\hat{f}_1(x, z) + 2r\hat{u}_{i^*+1}(x)$$
$$- \frac{\hat{g}^2(x)\rho_1^2(|X_1|^2)\zeta}{4} - \epsilon^2\zeta \quad (40)$$
$$- \frac{\rho_1^2(|X_1|^2)\zeta}{4} - \frac{\epsilon^2\rho^2(\zeta^2)\zeta}{2\rho^2(|x|^2)}$$

where $X_1 = [x^T, \zeta]^T$, and $\rho_1(s) = 2\rho(\frac{1}{2}s)$.

Next, define the approximation error as

$$e_{ro1}(X_1) = -\bar{f}_1(x, z) + \hat{f}_1(x, z)$$
$$+ 2r[u_{i^*+1}(x) - \hat{u}_{i^*+1}(x)]$$
$$- \frac{[\bar{g}_1^2(x) - \hat{g}_1^2(x)]\rho_1^2(|X_1|^2)\zeta}{4} \quad (41)$$

Then, conditions for asymptotic stability are summarized in the following Theorem:

Theorem 4.2: Under Assumptions 3.3, 3.4, and 4.1, if

$$\gamma_1 > \max\{\tilde{\kappa}_2, \tilde{\kappa}_1 \circ \underline{\lambda}^{-1} \circ \kappa_3 \circ \underline{\alpha}_1^{-1} \circ \bar{\alpha}_1\}, \quad (42)$$

and if the following implication holds for some constant $d_1 > 0$:

$$0 < U_{i^*}(X_1) \leq d_1 \Rightarrow \max\{|e_{ro1}(X_1)|, |e_{ro}(x)|\} < \gamma_1(|X_1|),$$

then the closed-loop system comprised of (29)-(31), and (40) is asymptotically stable at the origin. In addition, there exists $\sigma_1 \in \mathcal{K}_\infty$, such that

$$\Omega_{1,i^*} = \{(w, X_1) : \max[\sigma_1(U_{i^*}(X_1)), W(w)] \leq \sigma_1(d_1)\}$$

is an estimate of the region of attraction.

Proof: See the Appendix. ∎

Remark 4.1: In the absence of the dynamic uncertainty (i.e., $\Delta = 0$, $\Delta_1 = 0$ and the w-system is absent), the smooth functions ρ and ρ_1 can all be replaced by 0, and the system dynamics becomes

$$\dot{X}_1 = F_1(X_1) + G_1 u_{o1} \tag{43}$$

where $F_1(X_1) = \begin{bmatrix} f(x) + g(x)\zeta + g(x)\xi \\ -\nabla V_{i^*}(x)g(x) \end{bmatrix}$, $G_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, and $u_{o1} = -\epsilon^2 \zeta^2$. As a result, it can be concluded that the control policy $u = u_{o1}$ is an approximate optimal control policy with respect to the cost function

$$J_1(X_1(0)) = \int_0^\infty \left[Q_1(x, \zeta) + \frac{1}{2\epsilon^2}u^2 \right] dt \tag{44}$$

with $X_1(0) = [x_0^T, z_0 - u_{i^*}(x_0)]^T$ and $Q_1(x,\zeta) = Q(x) + \frac{1}{4r}[\nabla V_{i^*}(x)g(x)]^2 + \frac{\epsilon^2}{2}\zeta^2$.

V. IMPLEMENTATION ISSUES

In this section, we study a few implementation issues on the robust-ADP based online learning methodology, and give a practical algorithm. Due to the space limitation, we will mainly focus on the systems with matched dynamic uncertainties. These results can be easily extended to the unmatched case.

A. The compact set for approximation

Assumption 5.1: The closed-loop system composed of (8), (9), and

$$u = u_0(x) + e \tag{45}$$

is ISS when e, the exploration noise, is considered to be the input.

The reason for imposing Assumption 5.1 is two fold. First, like in many other policy iteration based ADP algorithms, an initial admissible control policy is desired. In this paper we further assume the initial control policy is stabilizing in the presence of dynamic uncertainties. Such an assumption is feasible and realistic by means of the designs in [14], [26]. Second, by adding the exploration noise, we are able to satisfy Assumptions 3.1 and 4.2, and at the same time keep the system solutions bounded.

Under Assumption 5.1, we can find a compact set Ω_0 which is an invariant set of the closed-loop system compose of (8), (9), and $u = u_0(x)$. In addition, we can also let Ω_0 contain Ω_{i^*} as its subset. Then, the compact set for approximation can be selected as $\Omega = \{x | \exists w, \text{ s.t. } (x,w) \in \Omega_0\}$.

B. Two-loop optimization scheme

In general cases, it may be difficult to determine the number of basis functions to be used for approximation. In this paper we propose a two-loop online optimization scheme as shown in Fig. 1. In the inner loop, least-squares method is used to train the weights. If the residual sum of errors is greater than a given threshold $\bar{\epsilon} > 0$, in the outer loop the number of basis functions are increased and online data are recollected to solve the minimization problem until sufficient small residual error can be obtained.
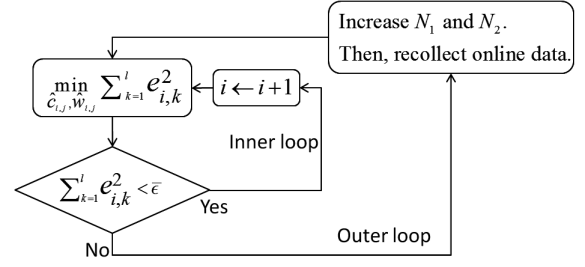

Fig. 1. Two-loop online optimization scheme

C. Robust-ADP algorithm

The robust-ADP algorithm is given in Algorithm 1.

Algorithm 1 Robust-ADP Algorithm

---

1. Let $(w(0), x(0)) \in \Omega_{i^*}$, employ the initial control policy (45) and collect the system state and input information.
2. Apply the online policy iteration using (13), and re-design the control policy using (23).
3. Terminate the exploration noise e.
4. If $(w(t), x(t)) \in \Omega_{i^*}$, apply the approximate robust optimal control policy (23).

VI. APPLICATION TO A SINGLE-JOINT HUMAN ARM MOVEMENT CONTROL PROBLEM

In this section, we apply the proposed online learning strategy to study a sensorimotor control problem. A linear version of this problem has been studied in [12].

Consider a single-joint arm movement as shown in Fig. 2, where the position of the elbow is fixed. The dynamic model is shown below [28].

$$I\ddot{\theta} = -mgl\cos(\theta) + n + T_m \tag{46}$$

where m is the mass of segment, I is the inertia, g is the gravitational constant, l is the distance of the center of mass from the joint, θ is the joint angular position, T_m is the input to the muscle from the motorneurons, and n denotes the inputs from the neural integrator, which can be modeled by a low pass filter as follows with a time constant τ_N.

$$\dot{n} = -\frac{n}{\tau_N} + T_m. \tag{47}$$

Let us define $x_1 = \theta - \theta_0$, $x_2 = \dot{\theta}$, $w = n - \frac{\tau_N mgl\cos(\theta_0)}{\tau_N+1} - Ix_2$, $u = T_m - \frac{mgl\cos(\theta_0)}{\tau_N+1}$, where θ_0 is the desired end point angular position. Then, the system can be converted to

$$\dot{w} = -\frac{1+\tau_N}{\tau_N}(w + Ix_2) \tag{48}$$
$$-2mgl\sin(\frac{x_1}{2})\sin(\frac{x_1}{2} + \theta_0) \tag{49}$$
$$\dot{x}_1 = x_2 \tag{50}$$
$$\dot{x}_2 = \frac{2mgl}{I}\sin(\frac{x_1}{2})\sin(\frac{x_1}{2} + \theta_0) \tag{51}$$
$$+\frac{1}{I}(u + Ix_2 + w)$$

Fig. 2.   Single-joint arm movement control problem.

Fig. 3.   Comparison of the approximated cost functions.

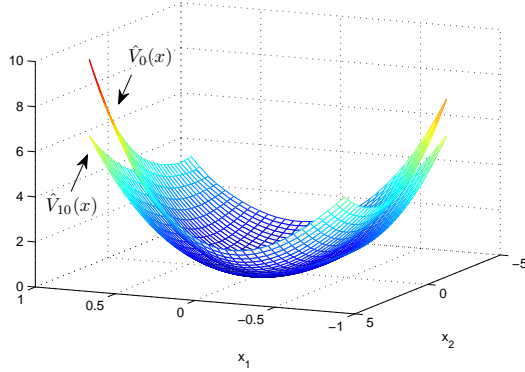Fig. 4.   Comparison of the speed profiles.

To apply the proposed robust-ADP method, the basis functions we used are polynomials with degrees less than or equal to five. The invariant set is chosen to contain the region $\{(w, x_1, x_2) : |w| \leq 1, |x_1| \leq 0.8, |x_2| \leq 3.5\}$. Only for simulation purpose, we set $\theta_0 = \frac{\pi}{4}$, $m = 1.65$, $l = 0.179$, $g = 9.81$, $I = 0.0779$. An initial control policy is set to be $u_0 = -0.5x_1 - 0.5x_2$. The initial condition is set to be $w(0) = 1$, $x_1(0) = -\frac{\pi}{4}$, and $x_2(0) = 0$. The optimal cost is specified as $J = \int_0^\infty \left(100x_1^2 + x_2^2 + u^2 \right) dt$.

In this simulation, the convergence is attained after 10 iterations. It can be seen from Fig. 3 that the approximated cost function $\hat{V}_{10}(x)$ is remarkably reduced compared with the initial approximated cost $\hat{V}_0(x)$. Also, in Fig. 4, we compare the speed curves under the initial control policy, the policy after two iterations, and the policy after 10 iterations. Clearly, after enough iteration steps, the speed profile becomes a bell-shaped curve which is consistent with experimental observations (see, for example, [3]).

VII. Conclusions

In this paper, computational robust optimal controller design has been studied for nonlinear systems with dynamic uncertainties. Both the matched and the unmatched cases are studied. We have presented for the first time a recursive, online, adaptive optimal controller design when dynamic uncertainties, characterized by input-to-state stable systems with unknown order and states/dynamics, are taken into consideration. We have achieved this goal by integration of approximate/adaptive dynamic programming (ADP) theory and several tools recen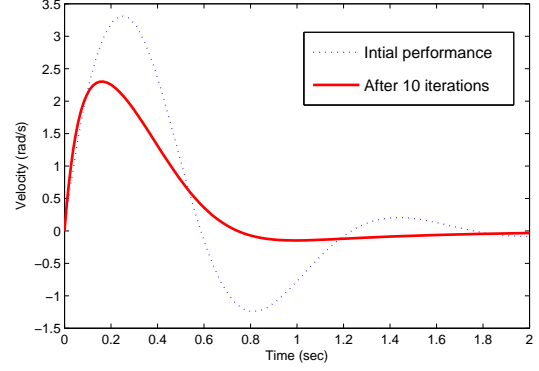tly developed within the nonlinear control community. Systematic robust-ADP based online learning algorithm has been developed. Rigorous stability analysis based on Lyapunov and small-gain techniques is carried out. The effectiveness of the proposed methodology has been validated by its application to a single-joint arm movement control problem.

Appendix

Proof of Theorem 3.1

To begin with, given $\hat{u}_i$, let $\tilde{V}_i(x)$ be the solution of the following equation with $\tilde{V}_i(0) = 0$.

$$\nabla \tilde{V}_i(x) \left(f(x) + g(x)\hat{u}_i(x) \right) + Q(x) + r\hat{u}_i^2(x) = 0 \qquad (52)$$

and denote $\tilde{u}_{i+1}(x) = -\dfrac{1}{2r}g(x)^T \nabla \tilde{V}_i(x)^T$.

Lemma A.1: For each $i \geq 0$, we have $\displaystyle \lim_{N_1, N_2 \to \infty} \hat{V}_i(x) = \tilde{V}_i(x)$, $\displaystyle \lim_{N_1, N_2 \to \infty} \hat{u}_{i+1}(x) = \tilde{u}_{i+1}(x)$, $\forall x \in \Omega$.

Proof: By definition

$$\tilde{V}_i(x(t_{k+1})) - \tilde{V}_i(x(t_k))$$
$$= -\int_{t_k}^{t_{k+1}} [Q(x) + r\hat{u}_i^2(x) + 2r\tilde{u}_{i+1}(x)\hat{v}_i(x)]dt \qquad (53)$$

Let $\tilde{c}_{i,j}$ and $\tilde{w}_{i,j}$ be the constant weights such that $\tilde{V}_i(x) = \sum_{j=1}^\infty \tilde{c}_{i,j}\phi_j(x)$ and $\tilde{u}_{i+1}(x) = \sum_{j=1}^\infty \tilde{w}_{i,j}\phi_j(x)$. Then, by (13) and (53), we have $e_{i,k} = \theta_{i,k}^T \bar{W}_i + \xi_{i,k}$, where

$$\bar{W}_i = \begin{bmatrix} \tilde{c}_{i,1} & \tilde{c}_{i,2} & \cdots & \tilde{c}_{i,N_1} & \tilde{w}_{i,1} & \tilde{w}_{i,2} & \cdots & \tilde{w}_{i,N_2} \end{bmatrix}^T$$
$$\quad - \begin{bmatrix} \hat{c}_{i,1} & \hat{c}_{i,2} & \cdots & \hat{c}_{i,N_1} & \hat{w}_{i,1} & \hat{w}_{i,2} & \cdots & \hat{w}_{i,N_2} \end{bmatrix}^T,$$
$$\xi_{i,k} = \sum_{j=N_1+1}^\infty \tilde{c}_{i,j} \left[\phi_j(x(t_{k+1})) - \phi_j(x(t_k)) \right]$$
$$\quad + \sum_{j=N_2+1}^\infty \tilde{w}_{i,j} \int_{t_k}^{t_{k+1}} 2r\phi_j(x)\hat{v}_i dt.$$

Since the weights are found using the least-squares method, we have

$$\sum_{k=1}^l e_{i,k}^2 \leq \sum_{k=1}^l \xi_{i,k}^2$$

Also notice that,

$$\sum_{k=1}^{l} \bar{W}_i^T \theta_{i,k}^T \theta_{i,k} \bar{W}_i = \sum_{k=1}^{l} (e_{i,k} - \xi_{i,k})^2$$

Then, under Assumption 3.1, it follows that

$$|\bar{W}_i|^2 \leq \frac{4|\Xi_{i,l}|^2}{l\delta} = \frac{4}{\delta} \max_{1 \leq k \leq l} \xi_{i,k}^2.$$

Therefore, given any arbitrary $\epsilon > 0$, we can find $N_{10} > 0$ and $N_{20} > 0$, such that when $N_1 > N_{10}$ and $N_2 > N_{20}$, we have

$$|\hat{V}_i(x) - \tilde{V}_i(x)| \tag{54}$$

$$\leq \sum_{j=1}^{N_1} |c_{i,j} - \hat{c}_{i,j}||\phi_j(x)| + \sum_{j=N_1+1}^{\infty} |c_{i,j}\phi_j(x)| \tag{55}$$

$$\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \quad \forall x \in \Omega. \tag{56}$$

Similarly, $|\hat{u}_{i+1}(x) - \tilde{u}_{i+1}(x)| \leq \epsilon$. The proof is complete. ∎

We now prove Theorem 3.1 by induction:
1) If $i = 0$ we have $\tilde{V}_0(x) = V_0(x)$, and $\tilde{u}_1(x) = u_1(x)$. Hence, the convergence can directly be proved by Lemma A.1.
2) Suppose for some $i > 0$, we have $\lim_{N_1, N_2 \to \infty} \hat{V}_{i-1}(x) = V_{i-1}(x)$, $\lim_{N_1, N_2 \to \infty} \hat{u}_i(x) = u_i(x)$, $\forall x \in \Omega$. By definition, we have

$$|V_i(x(t)) - \tilde{V}_i(x(t))|$$

$$= r|\int_t^{\infty} \left[\hat{u}_i(x)^2 - u_i(x)^2\right] dt|$$

$$+ 2r|\int_t^{\infty} u_{i+1}(x)g(x)\left[\hat{u}_i(x) - u_i(x)\right] dt|$$

$$+ 2r|\int_t^{\infty} \left[\hat{u}_{i+1}(x) - u_{i+1}(x)\right] g(x)\hat{v}_i dt|, \quad \forall x \in \Omega.$$

By the induction assumptions, we known

$$\lim_{N_1, N_2 \to \infty} \int_t^{\infty} \left[\hat{u}_i(x)^2 - u_i(x)^2\right] dt = 0 \tag{57}$$

$$\lim_{N_1, N_2 \to \infty} \int_t^{\infty} u_{i+1}(x)g(x)\left[\hat{u}_i(x) - u_i(x)\right] dt = 0 \tag{58}$$

Also, by Assumption 3.1, we conclude

$$\lim_{N_1, N_2 \to \infty} |u_{i+1}(x) - \hat{u}_{i+1}(x)| = 0 \tag{59}$$

and

$$\lim_{N_1, N_2 \to \infty} |V_i(x) - \tilde{V}_i(x)| = 0. \tag{60}$$

Finally, since

$$|\hat{V}_i(x) - V_i(x)| \leq |V_i(x) - \tilde{V}_i(x)| + |\tilde{V}_i(x) - \hat{V}_i(x)|$$

and by the induction assumption, we have

$$\lim_{N_1, N_2 \to \infty} |V_i(x) - \hat{V}_i(x)| = 0. \tag{61}$$

The proof is thus complete.

PROOF OF THEOREM 3.2

Define

$$\bar{e}_{ro}(x) = \begin{cases} e_{ro}(x), & V_{i^*}(x) \leq d \\ 0, & V_{i^*}(x) > d \end{cases} \tag{62}$$

and

$$u(x) = u_{i^*}(x) + \frac{r}{2}\rho^2(|x|^2)u_{i^*+1}(x) + \bar{e}_{ro}(x) \tag{63}$$

Then, along the solutions of (9), by completing the squares, we have

$$\dot{V}_{i^*}(x)$$

$$\leq -Q(x) + \frac{1}{\rho^2(|x|^2)}(\Delta + \bar{e}_{ro}(x))^2$$

$$= -(Q(x) - \epsilon^2|x|^2) - \frac{4\gamma^2 - (\Delta + \bar{e}_{ro}(x))^2}{\rho^2(|x|^2)}$$

$$\leq -Q_0(x) - 4\frac{\gamma^2 - \max\{\kappa_1^2(|w|), \kappa_2^2(|x|), \bar{e}_{ro}^2(|x|)\}}{\rho^2(|x|^2)}$$

where $Q_0(x) = Q(x) - \epsilon^2|x|^2$ is a positive definite function of x.

Therefore, under Assumptions 3.3, 3.4 and the gain condition (26), we have the following implication:

$$V_{i^*}(x) \geq \bar{\alpha} \circ \gamma^{-1} \circ \kappa_1 \circ \underline{\lambda}^{-1}(W(w))$$

$$\Rightarrow |x| \geq \gamma^{-1} \circ \kappa_1 \circ \underline{\lambda}^{-1}(W(w))$$

$$\Rightarrow \gamma(|x|) \geq \kappa_1(|w|) \tag{64}$$

$$\Rightarrow \gamma(|x|) \geq \max\{\kappa_1(|w|), \kappa_2(|x|), \bar{e}_{ro}(|x|)\}$$

$$\Rightarrow \dot{V}_{i^*}(x) \leq -Q_0(x).$$

Also, under Assumption 3.4, we have

$$W(w) \geq \kappa_3 \circ \underline{\alpha}^{-1}(V_{i^*}(x))$$

$$\Rightarrow W(w) \geq \kappa_3(|x|)$$

$$\Rightarrow \nabla W(w)\Delta_w(w, x) \leq -\kappa_4(|w|). \tag{65}$$

Finally, under the gain condition (26), it follows that

$$\gamma(s) > \kappa_1 \circ \underline{\lambda}^{-1} \circ \kappa_3 \circ \underline{\alpha}^{-1} \circ \bar{\alpha}(s)$$

$$\Rightarrow \gamma \circ \bar{\alpha}^{-1}(s') > \kappa_1 \circ \underline{\lambda}^{-1} \circ \kappa_3 \circ \underline{\alpha}^{-1}(s') \tag{66}$$

$$\Rightarrow s' > \bar{\alpha} \circ \gamma^{-1} \circ \kappa_1 \circ \underline{\lambda}^{-1} \circ \kappa_3 \circ \underline{\alpha}^{-1}(s')$$

where $s' = \bar{\alpha}(s)$. Hence, the following small-gain condition holds:

$$\left[\bar{\alpha} \circ \gamma^{-1} \circ \kappa_1 \circ \underline{\lambda}^{-1}\right] \circ \left[\kappa_3 \circ \underline{\alpha}^{-1}(s)\right] < s, \quad \forall s > 0. \tag{67}$$

By Theorem 3.1 in [13], the system (8), (9), (63) is globally asymptotically stable at the origin.

Next, denote $\chi_1 = \bar{\alpha} \circ \gamma^{-1} \circ \kappa_1 \circ \underline{\lambda}^{-1}$, and $\chi_2 = \kappa_3 \circ \underline{\alpha}^{-1}$. Also, let $\hat{\chi}_1$ be a function of class $\mathcal{K}_{\infty}$ such that
1) $\hat{\chi}_1(s) \leq \chi_1^{-1}(s)$, $\forall s \in [0, \lim_{s \to \infty} \chi_1(s))$,
2) $\chi_2(s) \leq \hat{\chi}_1(s)$, $\forall s \geq 0$.

Then, as shown in [13], there exists a continuously differentiable class $\mathcal{K}_{\infty}$ function $\sigma(s)$ satisfying $\sigma'(s) > 0$ and $\chi_2(s) < \sigma(s) < \hat{\chi}_1(s)$, $\forall s > 0$, such that the set

$$\Omega_{i^*} = \{(w, x) : \max[\sigma(V_{i^*}(x)), W(w)] \leq d\} \tag{68}$$

is an estimate of the region of attraction of the closed-loop system composed of (8), (9), and (23).

The proof is thus complete.

PROOF OF THEOREM 4.2

Define

$$
\bar{e}_{ro1}(X_1) = \begin{cases} e_{ro1}(X_1), & U_{i^*}(X_1) \leq d_1, \\ 0, & U_{i^*}(X_1) > d_1, \end{cases}
$$

$$
\bar{\bar{e}}_{ro}(x) = \begin{cases} e_{ro}(x), & U_{i^*}(X_1) \leq d_1, \\ 0, & U_{i^*}(X_1) > d_1, \end{cases}
$$

Along the solutions of (29)-(31) with the control policy

$$
u = -\bar{f}_1(x,z) + 2r\hat{u}_{i^*+1}(x) - \frac{\bar{g}^2(x)\rho_1^2(|X_1|^2)\zeta}{4}
$$
$$
- \frac{\rho_1^2(|X_1|^2)\zeta}{4} - \frac{\epsilon^2\rho^2(\zeta^2)\zeta}{2\rho^2(|x|^2)} - \epsilon^2\zeta - \bar{e}_{ro1}(X_1),
$$

it follows that

$$
\dot{U}_{i^*}(X_1) \leq -Q_0(x) - \frac{1}{2}\epsilon^2\zeta^2
$$
$$
- \frac{\gamma_1^2(|X_1|) - \max\{\tilde{\kappa}_1^2(|w|), \tilde{\kappa}_2^2(|X_1|), \bar{\bar{e}}_{ro}^2(x)\}}{\frac{1}{4}\rho^2(|x|^2)}
$$
$$
- \frac{\gamma_1^2(|X_1|) - \max\{\tilde{\kappa}_1^2(|w|), \tilde{\kappa}_2^2(|X_1|), \bar{\bar{e}}_{ro}^2(x)\}}{\frac{1}{4}\rho_1^2(|X_1|^2)}
$$
$$
- \frac{\gamma_1^2(|X_1|) - \max\{\tilde{\kappa}_1^2(|w|), \tilde{\kappa}_2^2(|X_1|), \bar{e}_{ro1}^2(X_1)\}}{\frac{1}{4}\rho_1^2(|X_1|^2)}
$$

As a result,

$$
U_{i^*}(X_1) \leq \max\{\bar{\alpha}_1 \circ \gamma_1^{-1} \circ \tilde{\kappa}_1 \circ \underline{\lambda}^{-1}(W(w)), \bar{\alpha}_1 \circ \gamma_1^{-1}(|v|)\}
$$
$$
\Rightarrow \dot{U}_{i^*}(X_1) \leq -Q_0(x) + \frac{1}{2}\epsilon^2|\zeta|^2
$$

The rest of the proof follows the same reasoning as in the proof of Theorem 3.2.

REFERENCES

[1] M. Abu-Khalaf and F. L. Lewis, "Neurodynamic programming and zero-sum games for constrained control systems," *IEEE Trans. Neural Networks*, vol. 19, no. 7, pp. 1243-1252, 2008.

[2] A. Al-Tamimi, F. L. Lewis, and M. Abu-Khalaf, "Model-free Q-learning designs for linear discrete-time zero-sum games with application to H-infinity control," *Automatica*, vol. 43, pp. 473-481, 2007.

[3] C. G. Atkeson and J. M. Hollerbach, "Kinematic features of unrestrained vertical arm movements," *The Journal of Neurosaence*, vol. 5., no. 9, pp. 2318-2330, 1985.

[4] R. E. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton University Press, 1957.

[5] D. P. Bersekas, and J. N. Tsitsiklis, *Neuro-dynamic programming*, Athena Scientific, Nashua, NH, 1996.

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

[7] Y. Jiang and Z. P. Jiang, "Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics," *Automatica*, vol. 48, no. 10, pp. 2699-2704, 2012.

[8] Y. Jiang and Z. P. Jiang, "Robust adaptive dynamic programming with an application to power systems," *IEEE Trans. Neural Networks and Learning Systems*, in press, DOI: 10.1109/TNNLS.2013.2249668

[9] Y. Jiang and Z. P. Jiang, "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems," submitted to *IEEE Trans. Neural Networks and Learning Systems*, 2012.

[10] Y. Jiang and Z. P. Jiang, "Robust approximate dynamic programming and global stabilization with nonlinear dynamic uncertainties," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, FL, USA, pp. 115–120, 2011.

[11] Y. Jiang and Z. P. Jiang, "Robust adaptive dynamic programming," in *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*, F. L. Lewis and D. Liu, Eds, John Wiley and Sons, pp. 281-302, 2012.

[12] Y. Jiang and Z. P. Jiang, Jiang, "Adaptive dynamic programming as a theory of sensorimotor control," in *Proceedings of the 2012 IEEE Signal Processing in Medicine and Biology Symposium*, pp. 1-4, 2012.

[13] Z. P. Jiang, I. Mareels and Y. Wang, "A Lyapunov formulation of the nonlinear small gain theorem for interconnected ISS systems," *Automatica*, vol. 32, no. 8, pp. 1211-1215, 1996.

[14] Z. P. Jiang and I. M. Y. Mareels, "A small-gain control method for nonlinear cascaded systems with dynamic uncertainties," *IEEE Trans. Automatic Control*, vol. 42, no. 3, pp. 292-308, 1997.

[15] Z. P. Jiang, A. R. Teel, and L. Praly, "Small-gain theorem for ISS systems and applications," *Mathematics of Control, Signals, and Systems*, vol. 7, no. 2, pp. 95-120, 1994.

[16] I. Karafyllis and Z. P. Jiang, *Stability and Stabilization of Nonlinear Systems*, Springer, 2011.

[17] H. K. Khalil, *Nonlinear Systems* (3rd edition), Prentice Hall, 2002.

[18] M. Krstic, I. Kanellakopoulos and P. V. Kokotovic, *Nonlinear and Adaptive Control Design*, John Wiley, 1995.

[19] P. Kundur, N. J. Balu, and M. G. Lauby, *Power System Stability and Control*, McGraw-Hill: New York, 1994.

[20] F. L. Lewis and V. L. Syrmos, *Optimal Control*, Wiley, 1995.

[21] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Trans. Circuits and Systems Magazine*, vol. 9, no. 3, pp. 32-50, 2009.

[22] F. L. Lewis, K. G. Vamvoudakis, "Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data," *IEEE Transactions Systems, Man, and Cybernetics, Part B*, vol. 41, no. 1, pp. 14-23, 2011.

[23] F. Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: an introduction," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 39-47, 2009.

[24] J. J. Murray, C. J. Cox, G. G. Lendaris, "Adaptive dynamic programming", *IEEE Trans. Systems, Man, and Cybernetics‡Part C: Applications and Reviews*, vol. 32, no. 2, pp. 140–153, 2002.

[25] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Trans. Neural Networks*, vol. 1, no. 1, pp. 4-27, 1990.

[26] L. Praly and Y. Wang, "Stabilization in spite of matched unmodeled dynamics and an equivalent definition of input-to-state stability," *Mathematics of Control, Signals, and Systems*, vol. 9, pp. 1-33, 1996.

[27] G. N. Saridis and C.-S. G. Lee, "An approximation theory of optimal control for trainable manipulators," *IEEE Trans. System, Man, and Cybernetics*, vol. 9, no. 3, pp. 152-159, 1979.

[28] R. Shadmehr and S. P. Wise, *The Computational Neurobiology of Reaching and Pointing*, MIT Press, 2005.

[29] E. D. Sontag, "Input to state stability: basic concepts and results," in Nonlinear and Optimal Control Theory, Berlin: Springer-Verlag, pp. 163-220, 2007.

[30] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.

[31] K. G. Vamvoudakis, F. L. Lewis, "Multi-player non zero sum games: online adaptive learning solution of coupled hamilton-jacobi equations," *Automatica*, vol. 47, no. 8, pp. 1556-1569, 2011.

[32] D. Vrabie and F. Lewis, "Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems," *Neural Networks*, vol. 22, no. 3, pp. 237–246, 2009.

[33] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477-484, 2009.

[34] P. J. Werbos, *The Elements of Intelligence*, Namur, Belgium: Cybernetica, 1968. No. 3.

[35] P. J. Werbos, *Beyond regression: New tools for prediction and analysis in the behavioural sciences*, Ph.D. Thesis, Harvard University, 1974.

[36] 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: Van Nostrand Reinhold, 1992.

[37] Y. Zhang, P. Y. Peng, and Z. P. Jiang, "Stable neural controller design for unknown nonlinear systems using backstepping," *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1347-1360, 2000.