51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

Robust Adaptive Dynamic Programming for Nonlinear Control Design

Yu Jiang and Zhong-Ping Jiang

Abstract— **This paper presents a robust optimal controller design for unknown nonlinear systems from a perspective of robust adaptive dynamic programming (robust-ADP). The proposed methodology has several novel features. First, the class of nonlinear systems studied in the paper allows for the presence of dynamic uncertainties with unmeasured state and uncertain system order/dynamics. Second, in the absence of the dynamic uncertainty, the online policy iteration technique developed in this paper can be viewed as an extension of the existing ADP method to affine continuous-time nonlinear systems with completely unknown dynamics. Third, the theory of approximate/adaptive dynamic programming (ADP) is integrated for the first time with tools from modern nonlinear control theory, such as the nonlinear small-gain theorem, for robust optimal control design. It is shown that, with appropriate robust redesign, the robust-ADP controller asymptotically stabilizes the overall system. A practical robust-ADP-based online learning algorithm is developed in this paper, and is applied to the robust optimal controller design for a two-machine power system.**

I. Introduction

This paper studies the robust optimal control problem for the following system

$$\dot{w} = q(w,x), \tag{1}$$
$$\dot{x} = f(x) + g(x)[u + h(w,x)] \tag{2}$$

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; $q : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}^p$, $f : \mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}^n$ and $h : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}$ are locally Lipschitz.

In this paper, this problem is studied using the idea of approximate/adaptive dynamic programming (ADP) [28], [29], [30], [31], which is a computational method that directly approximates the optimal control policy without knowing precisely the system dynamics. In recent years, considerable attention has been paid for ADP-based controller design for continuous-time systems (see, for example, [8], [26], [33]). Compared with model predictive control (MPC) [5], [19], which is a receding-horizon method attempting to obtain approximate optimal controllers using online measurements, ADP is a model-free method aiming to approximate, through learning and online measurement, optimal controllers with guaranteed stability for dynamic systems with unknown dynamics.

As a natural extension of ADP to uncertain dynamic systems with incomplete state information and unknown system order, robust adaptive dynamic programming (robust-ADP) has been proposed in [9], [10], and [11]. The framework of robust-ADP integrates tools from nonlinear control theory, such as Lyapunov designs, input-to-state stability theory [21], [22], and nonlinear small-gain techniques [13]. Employing the idea of robust-ADP, the main objective of this paper is to find online a control policy that semi-globally asymptotically stabilizes the system comprised of (1) and (2). In addition, the control policy should preserve certain optimality, with respect to some given cost function, for the x-subsystem in the absence of the dynamic uncertainty (i.e. $h \equiv 0$).

In order to extend the robust-ADP framework from partially linear composite systems [10] to fully nonlinear systems, two efforts are made in this paper: First, instead of tuning the weighting matrices as in [10], we adopt the robust redesign technique [20] to modify the ADP-based control policy such that certain small-gain condition can be satisfied, and the asymptotic stability is guaranteed. Second, in the absence of the dynamic uncertainty, the continuous-time ADP methodology in the previous literature required $g(x)$ to be exactly know [25]. To remove this assumption, we simultaneously perform policy evaluation and policy improvement using two sets of basis functions. As a result, the knowledge of $f(x)$ and $g(x)$ is no longer needed for online learning.

The remainder of the paper is organized as follows. Section 2 introduces the concept of input-to-state stability (ISS) and reviews the nonlinear small-gain theorem. Section 3 develops the robust-ADP technique for nonlinear control design, and a practical algorithm is provided. Two numerical examples, including the controller design for synchronous power generators, are provided in Section 4. Finally, concluding remarks are given in Section 5.

Throughout this paper, we use $\mathbb{R}$ to denote the set of real numbers. Vertical bars $|\cdot|$ represent the Euclidean norm for vectors, or the induced matrix norm for matrices. For any piecewise continuous function u, $\|u\|$ denotes $\sup\{|u(t)|, t \geq 0\}$. We use $\otimes$ to indicate Kronecker product, and $\text{vec}(A)$ is defined to be the mn-vector formed by stacking the columns of $A \in \mathbb{R}^{n \times m}$ on top of one another, i.e., $\text{vec}(A) = [a_1^T \ a_2^T \ \cdots \ a_m^T]^T$, where $a_i \in \mathbb{R}^n$ are the columns of A. A control law is also called a *policy*, and it is said to be stabilizing if under the policy, the closed-loop system is asymptotically stable at the origin.

II. Preliminaries

In this section, we recall some important tools from modern nonlinear control [12], [13], [14], [21]. They will be helpful for developing the robust-ADP methodology in the next section.

This work has been supported in part by NSF grants DMS-0906659 and ECCS-1101401. Y. Jiang and Z. P. Jiang are with the Department of Electrical and Computer Engineering, Polytechnic Insinuate of New York University, Brooklyn, NY 11201, USA. yjiang06@students.poly.edu, zjiang@poly.edu

978-1-4673-2064-1/12/$31.00 ©2012 IEEE

Definition 2.1 ([21]): A dynamic control system $\dot{x} = f(x) + g(x)u$ is said to be *input-to-state stable* (ISS) with gain γ if, for any measurable essentially bounded input u and any initial condition $x(0)$, the solution $x(t)$ exists for every $t \geq 0$ and satisfies

$$|x(t)| \leq \beta(|x(0)|, t) + \gamma(\|u\|) \tag{3}$$

where β is of class $\mathcal{KL}$ and γ is of class $\mathcal{K}$ [14].

Next, consider an interconnected system described by

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, x_2), & (4)\\ \dot{x}_2 &= f_2(x_1, x_2) & (5) \end{aligned}$$

where, for $i = 1, 2$, $x_i \in \mathbb{R}^{n_i}$, and $f_i : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{n_i}$ is locally Lipschitz.

Assumption 2.1: For each $i = 1, 2$, there exists an ISS-Lyapunov function V_i for the x_i subsystem such that the following hold:

1) there exist functions $\underline{\alpha}_i, \bar{\alpha}_i \in \mathcal{K}_\infty$, such that

$$\underline{\alpha}_i(|x_i|) \leq V_i(x_i) \leq \bar{\alpha}_i(|x_i|),\ \forall x_i \in \mathbb{R}^{n_i}; \tag{6}$$

2) there exist a class $\mathcal{K}$ functions χ_i and a positive definite function α_i, such that the following implications hold:

$$\begin{aligned} V_1(x_1) \geq \chi_1(V_2(x_2)) \Rightarrow \nabla V_1 f_1 \leq -\alpha_1(V_2(x_2)), && (7)\\ V_2(x_2) \geq \chi_2(V_1(x_1)) \Rightarrow \nabla V_2 f_2 \leq -\alpha_2(V_1(x_1)). && (8) \end{aligned}$$

The following theorem gives the small-gain condition, under which global asymptotic stability of the interconnected system can be achieved.

Theorem 2.1 ([12]): Under Assumption 2.1, suppose the following small-gain condition holds:

$$\chi_1 \circ \chi_2(s) < s, \quad \forall s > 0. \tag{9}$$

Then, the interconnected system (4), (5) is globally asymptotically stable at the origin.

Under Assumption 2.1 and the small-gain condition (9), Let $\hat{\chi}_1$ be a function of class $\mathcal{K}_\infty$ such that

1) $\hat{\chi}_1(s) \leq \chi_1^{-1}(s)$, $\forall s \in [0, \lim\limits_{s\to\infty} \chi_1(s))$,
2) $\chi_2(s) \leq \hat{\chi}_1(s)$, $\forall s \geq 0$.

Then, as shown in [12], there exists a continuously differentiable class $\mathcal{K}_\infty$ function $\sigma(s)$ satisfying $\dot{\sigma}(s) > 0$ and $\chi_2(s) < \sigma(s) < \hat{\chi}_1(s)$, $\forall s > 0$.

In [12], it is also shown that the function

$$V_{12} = \max\{\sigma(V_1(x_1)), V_2(x_2)\} \tag{10}$$

is proper and positive-definite. In addition, we have

$$\dot{V}_{12}(x_1, x_2) < 0, \tag{11}$$

that holds almost everywhere in the state space.

III. Robust-ADP Design

In this section, we develop the robust-ADP design method for nonlinear systems. First, we introduce the robust optimal design method, which is an integration of the optimal control theory [17] with the gain assignment technique [20]. A gain condition for global asymptotic stability is derived. Second, a result on semi-global stabilization is developed when the feedback control policy can only be defined on compact sets. Third, we propose a learning scheme to find online the robust optimal control policy without knowing the system dynamics. Finally, a practical algorithm will be provided.

A. Robust optimal design

Consider the following quadratic cost

$$V = \int_t^\infty \left[Q(x) + ru^2\right] d\tau \tag{12}$$

where $Q(\cdot)$ is positive definite and $r > 0$ is a constant. Notice that $V = V(x(t))$ when feedback policy $u = u(x)$ is applied to system (2) with $h \equiv 0$.

Assumption 3.1: For system (2) with $h \equiv 0$, there exists a C^1 positive definite and proper function $V^*(x)$, satisfying the following HJB equation:

$$\nabla V^* f + Q - \frac{1}{4r}|\nabla V^* g|^2 = 0. \tag{13}$$

Under Assumption 3.1, according to optimal control theory [17], the optimal control policy is given by

$$u^* = -\frac{1}{2r}\nabla V^* g, \tag{14}$$

and there exist functions α_1 and α_2, of class $\mathcal{K}_\infty$, such that,

$$\alpha_1(|x|) \leq V^*(x) \leq \alpha_2(|x|), \qquad \forall x \in \mathbb{R}^n. \tag{15}$$

Assumption 3.2: For system (1), there exist functions $\kappa_1, \kappa_2 \in \mathcal{K}_\infty$, $\kappa_3, \kappa_4, \kappa_5 \in \mathcal{K}$, and positive definite functions W and κ_6, such that for all $w \in \mathbb{R}^p$ and $x \in \mathbb{R}^n$, we have

$$\begin{aligned} \kappa_1(|w|) \leq W(w) \leq \kappa_2(|w|), && (16)\\ |h(x, w)| \leq \kappa_3(|w|) + \kappa_4(|x|), && (17) \end{aligned}$$

together with the following implication:

$$W(w) \geq \kappa_5(|x|) \Rightarrow \nabla W(w) q(w, x) \leq -\kappa_6(|w|). \tag{18}$$

In order to achieve global asymptotic stability, robust redesign of the control policy is needed. To this end, we introduce the following notations [20]:

$$\begin{aligned} \mathrm{sat}(s) &= \begin{cases} s, & s \in [0, 1],\\ 1, & s > 1, \end{cases} & (19)\\ \theta(x) &= \begin{cases} \mathrm{sat}(\theta_0(x)), & |u^*(x)| \neq 0,\\ 0, & |u^*(x)| = 0, \end{cases} & (20)\\ \theta_0(x) &= \frac{\sqrt{\frac{Q^2(x)}{4} + 3\left(2ru^*(x)\rho(V^*(x))\right)^2} - \frac{Q(x)}{2}}{|2ru^*(x)|\rho(V^*(x))}, & (21)\\ u^{\mathrm{r}}(x) &= \mathrm{sign}(u^*(x))\theta(x)\rho(V^*(x)) & (22) \end{aligned}$$

where ρ is of class $\mathcal{K}_\infty$ and u^{r} is a continuous function [20].

The following proposition is inspired by [20].

Proposition 3.1: The system (1), (2) is globally asymptotically stable under the control policy

$$u(x) = u^*(x) + u^{\mathrm{r}}(x) \tag{23}$$

if ρ is chosen to satisfy:

$$\rho(s) > \kappa_4 \circ \alpha_1^{-1}(s) + \kappa_3 \circ \kappa_1^{-1} \circ \rho_0(s), \quad \forall s > 0 \tag{24}$$

where ρ_0 is a class $\mathcal{K}_\infty$ function such that

$$\rho_0^{-1} \circ \kappa_5 \circ \alpha_1^{-1}(s) < s, \quad \forall s > 0. \tag{25}$$

Proof: For simplicity, we will drop the arguments which are clear from the context. Similar as in [20], we obtain

$$\begin{aligned}\dot{V}^* &\leq -\frac{Q}{2} - |\nabla V^* g| \left[\rho(V^*) - |h(x,w)|\right] \\ &\leq -\frac{Q}{2} - |\nabla V^* g| \left[\rho(V^*) - \kappa_3(|w|) - \kappa_4 \circ \alpha_1^{-1}(V^*)\right] \\ &\leq -\frac{Q}{2} - |\nabla V^* g| \left[\kappa_3 \circ \kappa_1^{-1} \circ \rho_0(V^*) - \kappa_3 \circ \kappa_1^{-1}(W)\right]\end{aligned}$$

Therefore, we have the following implication:

$$V^* \geq {\rho_0}^{-1}(W) \Rightarrow \nabla V^*(f + gu) \leq -\frac{Q}{2}. \tag{26}$$

Also, under Assumption 3.2, we have

$$\begin{aligned} W \geq \kappa_5 \circ \alpha_1^{-1}(V^*) \quad &\Rightarrow \quad W \geq \kappa_5(|x|) \\ &\Rightarrow \quad \nabla W q \leq -\kappa_6(|w|). \end{aligned} \tag{27}$$

In summary, by Theorem 2.1, the system (1), (2), (23) is globally asymptotically stable at the origin. ■

B. Estimating the domain of attraction

When designing the adaptive critic [31], the control policy may not be obtained globally, but can only be solved or approximated on a compact set. Hence, in practice, we need to estimate the domain of attraction, as shown below.

Theorem 3.1: Given any bounded set $X \subset \mathbb{R}^{n \times p}$, we can find a compact set $\Omega \subset \mathbb{R}^n$, such that for any control policy $u_c(x)$ satisfying $u_c(x) \equiv u^*(x) + u^{\mathrm{r}}(x)$, $\forall x \in \Omega$, the closed-loop system comprised of (1), (2) and $u_c(x)$ is asymptotically stable with X contained in its domain of attraction.

Proof: Let us define the following sets:

$$\begin{aligned} B_d &= \{(x,w) : |x| \leq d,\ W(w) \leq \sigma \circ \alpha_2(d)\}, \\ \Omega_d &= \{(x,w) : \max\{\sigma(V^*(x)), W(w)\} \leq \sigma \circ \alpha_2(d)\}, \\ \tilde{B}_d &= \{(x,w) : |x| \leq \alpha_1^{-1} \circ \alpha_2(d),\ W(w) \leq \alpha_2(d)\} \end{aligned}$$

where $d > 0$ is a sufficiently large number such that $X \subseteq B_d$.

Next, we show that

$$B_d \subseteq \Omega_d \subseteq \tilde{B}_d. \tag{28}$$

Indeed, if $(x', w') \in B_d$, we have

$$\sigma \circ V^*(x') \leq \sigma \circ \alpha_2(|x'|) \leq \sigma \circ \alpha_2(d), \tag{29}$$

$$W(w') \leq \sigma \circ \alpha_2(d). \tag{30}$$

Therefore, $(x', w') \in \Omega_d$.

On the other hand, we can show that

$$(x', w') \notin \tilde{B}_d \Rightarrow (x', w') \notin \Omega_d. \tag{31}$$

Actually, by definition, if $(x', w') \notin \tilde{B}_d$, we have

$$\sigma \circ V^*(x') \geq \sigma \circ \alpha_1(|x'|) > \sigma \circ \alpha_2(d) \tag{32}$$

or

$$\sigma \circ W(w') \geq \sigma \circ \alpha_2(d) \tag{33}$$

In either case, it is easy to see $(x', w') \notin \Omega_d$.

Hence, given a bounded set X, we can find $d > 0$, such that

$$X \subseteq \Omega_d \subseteq \tilde{B}_d. \tag{34}$$

Since only x is available for feedback design, we define

$$\Omega = \{x : |x| \leq \alpha_1^{-1} \circ \alpha_2(d)\}. \tag{35}$$

Finally, since Ω_d is an estimate of the domain of attraction [14], the control input defined on the Ω will give the desired asymptotic stability property.

The proof is complete. ■

C. Online policy iteration

Now, we are able to develop the online policy iteration technique. To begin with, let us introduce the concept of admissible control policy [3] and an iterative technique to approximate the optimal control policy [1], [3].

Definition 3.1 ([3]): For system (2) with $h \equiv 0$, a control policy $u(x)$ is defined as *admissible*, if $u(x)$ is continuous on Ω, $u(0) = 0$, $u(x)$ stabilizes the x-system on the compact set Ω defined in (35), and the cost $V(x_0)$ defined in (12) is finite for all $x_0 \in \Omega$.

The following steps are iteratively employed to approximate the optimal control policy [1], [3]:

Policy evaluation: Given an admissible policy u_i with $i \geq 0$, solve for $V_i(x)$ using

$$0 = \nabla V_i(f + gu_i) + Q + r|u_i|^2, \quad V_i(0) = 0. \tag{36}$$

Policy improvement: Update the control policy using

$$u_{i+1} = -\frac{1}{2}\nabla V_i g. \tag{37}$$

It is shown in [1] and [3] that the policy iteration (36) and (37) converges uniformly to the optimal control solution (V^*, u^*) in Ω, i.e. $\forall \epsilon > 0$, there exists $i_0 > 0$, such that for all $i > i_0$ we have

$$\sup_{x \in \Omega} |V_i(x) - V^*(x)| < \epsilon, \ \sup_{x \in \Omega} |u_i(x) - u^*(x)| < \epsilon. \tag{38}$$

Unfortunately, the iterative technique relies on the knowledge of $f(x)$ and $g(x)$. To remove this requirement, we extend the computational adaptive optimal control method [8] to affine nonlinear systems.

To begin with, notice that (2) with $h \equiv 0$ can be rewritten as

$$\dot{x} \ = \ f + gu_i + g(\hat{u} - u_i) \tag{39}$$

where $\hat{u} = u + h$.

Therefore, for each V_i, along the solutions of (39) we have

$$\begin{aligned}\dot{V}_i &= \nabla V_i \left[(f + gu_i) + g(\hat{u} - u_i)\right] \\ &= -Q - r|u_i|^2 + \nabla V_i g(\hat{u} - u_i) \\ &= -Q - r|u_i|^2 - 2ru_{i+1}(\hat{u} - u_i).\end{aligned}$$

Integrating both sides yields

$$\begin{aligned}&V_i(x(t+T)) - V_i(x(t)) \\ &= \int_t^{t+T} \left[-Q - r|u_i|^2 - 2ru_{i+1}(\hat{u} - u_i)\right] d\tau. \quad (40)\end{aligned}$$

In order to solve (40), we define the following two vectors of basis functions:

$$\begin{aligned}\Phi(x) &= \left[\begin{array}{cccc} \phi_1(x) & \phi_2(x) & \cdots & \phi_L(x)\end{array}\right]^T \in \mathbb{R}^L, \\ \Psi(x) &= \left[\begin{array}{cccc} \psi_1(x) & \psi_2(x) & \cdots & \psi_N(x)\end{array}\right]^T \in \mathbb{R}^N\end{aligned}$$

where $\{\phi_j\}$ and $\{\psi_k\}$ are two sets of linearly independent functions on Ω, with $\phi_j(0) = \psi_k(0) = 0$, for all $j = 1, 2, \cdots, L$, and $k = 1, 2, \cdots, N$.

Then, for each $i = 0, 1, \cdots$, the cost function and the control policy are approximated by

$$V_i(x) = \mathbf{p}_i^T \Phi(x) + e_i^{(1)}, \quad (41)$$

$$u_i(x) = \mathbf{k}_i^T \Psi(x) + e_i^{(2)} \quad (42)$$

where, for each $i = 0, 1, \cdots$, $\mathbf{p}_i \in \mathbb{R}^L$, $\mathbf{k}_i \in \mathbb{R}^N$ are two constant vectors, and $e_i^{(1)}$ and $e_i^{(2)}$ stand for the approximation errors.

Further, for any positive integer $l > 0$, define

$$\delta_\phi = \left[\begin{array}{c} \Phi(x(t_1)) - \Phi(x(t_0)) \\ \Phi(x(t_2)) - \Phi(x(t_1)) \\ \vdots \\ \Phi(x(t_l)) - \Phi(x(t_{l-1})) \end{array}\right] \in \mathbb{R}^{l \times L}, \quad (43)$$

$$I_\psi = \left[\begin{array}{c} \int_{t_0}^{t_1} \Psi(x(t)) \otimes \Psi(x(t)) dt \\ \int_{t_1}^{t_2} \Psi(x(t)) \otimes \Psi(x(t)) dt \\ \vdots \\ \int_{t_{l-1}}^{t_l} \Psi(x(t)) \otimes \Psi(x(t)) dt \end{array}\right] \in \mathbb{R}^{l \times N^2}, \quad (44)$$

$$I_{\psi u} = \left[\begin{array}{c} \int_{t_0}^{t_1} \Psi(x(t)) \hat{u}(t) dt \\ \int_{t_1}^{t_2} \Psi(x(t)) \hat{u}(t) dt \\ \vdots \\ \int_{t_{l-1}}^{t_l} \Psi(x(t)) \hat{u}(t) dt \end{array}\right] \in \mathbb{R}^{l \times N}, \quad (45)$$

$$I_Q = \left[\begin{array}{c} \int_{t_0}^{t_1} Q(x(t)) dt \\ \int_{t_1}^{t_2} Q(x(t)) dt \\ \vdots \\ \int_{t_{l-1}}^{t_l} Q(x(t)) dt \end{array}\right] \in \mathbb{R}^l. \quad (46)$$

Hence, (40) implies the following equation

$$\begin{aligned}&[\delta_\phi \quad -2rI_{\psi u} + 2rI_\psi(\mathbf{k}_i \otimes I_N)] \left[\begin{array}{c} \mathbf{p}_i \\ \mathbf{k}_{i+1} \end{array}\right] \\ &= -I_Q - I_\Psi(\mathbf{k}_i \otimes \mathbf{k}_i) + \varepsilon_i \quad (47)\end{aligned}$$

where $\varepsilon_i \in \mathbb{R}^l$ denotes the residual error, which is caused by replacing the actual cost function and control policy with the basis functions. This error can also be regarded as the temporal difference residual error for continuous-time systems [2], [4], [23].

The constant vectors are solved in the least-squares sense as follows:

$$\left[\begin{array}{c} \mathbf{p}_i \\ \mathbf{k}_{i+1} \end{array}\right] = (\Theta_i^T \Theta_i)^{-1} \Theta_i^T \Xi_i \quad (48)$$

where $\Theta_i \in \mathbb{R}^{l \times L+N}$ and $\Xi_i \in \mathbb{R}^l$ are defined as

$$\Theta_i = [\delta_\phi \quad -2rI_{\psi u} + 2rI_\psi(\mathbf{k}_i \otimes I_N)], \quad (49)$$

$$\Xi_i = -I_Q - I_\Psi(\mathbf{k}_i \otimes \mathbf{k}_i). \quad (50)$$

To ensure that the matrix $\Theta_i^T \Theta_i$ is invertible, i.e., Θ_i is full column rank, we give the following assumption:

Assumption 3.3:

$$\text{rank}\left(\left[\begin{array}{ccc} \delta_\phi & I_{\psi u} & I_\psi \end{array}\right]\right) = L + N + \frac{N(N+1)}{2}. \quad (51)$$

Theorem 3.2: Under Assumption 3.3, Θ_i is full column rank, for all $i = 0, 1, \cdots$.

Proof: for each $i = 0, 1, \cdots$, we have

$$\begin{aligned}&\left[\begin{array}{cc} \Theta_i & I_\psi \end{array}\right] \\ &= [\delta_\phi \quad -2rI_{\psi u} + 2rI_\psi(\mathbf{k}_i \otimes I_N) \quad I_\psi] \\ &= [\delta_\phi \quad I_{\psi u} \quad I_\psi] \left[\begin{array}{ccc} I_L & 0 & 0 \\ 0 & -2rI_N & 0 \\ 0 & 2r(\mathbf{k}_i \otimes I_N) & I_{N^2} \end{array}\right].\end{aligned}$$

Let $\bar{I}_\psi \in \mathbb{R}^{l \times \frac{N(N+1)}{2}}$ be the matrix comprised of $\frac{N(N+1)}{2}$ distinct columns of I_ψ. Then, under Assumption 3.3, it follows that

$$\begin{aligned}\text{rank}(\left[\begin{array}{cc} \Theta_i & \bar{I}_\psi \end{array}\right]) &= \text{rank}(\left[\begin{array}{cc} \Theta_i & I_\psi \end{array}\right]) \\ &= \text{rank}(\left[\begin{array}{ccc} \delta_\phi & I_{\psi u} & I_\psi \end{array}\right]) \\ &= L + N + \frac{1}{2}N(N+1).\end{aligned}$$

Therefore, the matrix $\left[\begin{array}{cc} \Theta_i & \bar{I}_\psi \end{array}\right]$ is full column rank. As a result, we have $\text{rank}(\Theta_i) = L + N$.

The proof is complete. ■

Corollary 3.1: Under Assumption 3.3, assume $e_i^{(1)} = e_i^{(2)} = 0$, $\forall i = 0, 1, \cdots$. Then, solving (36) and (37) is equivalent to solving (48).

Proof: Notice that $V_i(x)$ and $u_{i+1}(x)$ uniquely determine $\mathbf{p}_i$ and $\mathbf{k}_{i+1}$, respectively, if $e_i^{(1)}(x) = e_i^{(2)}(x) = 0$, $\forall x \in \Omega$. Next, from (40), it can be seen that $\mathbf{p}_i$ and $\mathbf{k}_{i+1}$ satisfy (48). On the other hand, under Assumption 3.3, there is only one solution satisfying (48). Therefore, $V_i(x)$ and $u_{i+1}(x)$ uniquely determined by $\mathbf{p}_i$ and $\mathbf{k}_{i+1}$ must satisfy (36) and (37). The proof is complete. ■

D. Robust-ADP algorithm

Finally, we are able to give the robust-ADP algorithm. Notice that the algorithm is conducted in the absence of dynamic uncertainties. To employ the online policy iteration, it is assumed that an initial admissible control policy $u^{(0)}(x)$

is known. In order to satisfy Assumption 3.3, we apply the following input for the learning purpose

$$u = u_0(x) + e(t), \quad \forall t \in [0, T] \tag{52}$$

where $e(t)$ is an exploration noise and T is a sufficiently large positive constant.

Notice that finite escape can be voided for the system comprised of (2) and (52) based on the main result of [21].

Robust-ADP Algorithm

1. Employ the input (52) and record δ_ϕ, I_ψ and $I_{\psi u}$. Let $i \leftarrow 0$.
2. Solve $\mathbf{p_i}$ and $\mathbf{k_{i+1}}$ from (47).
3. Let $i \leftarrow i+1$, and go to Step 2, until $|\mathbf{p_i} - \mathbf{p_{i-1}}| \leq \epsilon$, where the constant $\epsilon > 0$ can be any predefined threshold.
4. Let $\hat{V}^*(x) = \mathbf{p}_i\Phi(x)$ and $\hat{u}^*(x) = \mathbf{k}_{i+1}\Psi(x)$, compute $u^{\mathrm{r}}(x)$ according to (19)-(22).
5. Use $u(x) = \mathbf{k}_{i+1}\Psi(x) + u^{\mathrm{r}}(x)$ as the approximated robust optimal control policy.

IV. Application: Synchronous generators

Consider the two synchronous generators with governor controllers described as follows [16]:

$$\dot{\delta}_i = \omega_i - \omega_0, \tag{53}$$

$$\dot{\omega}_i = -\frac{D_i}{2H_i}\omega_i + \frac{\omega_0}{2H_i}\left(P_{mi} - P_{ei}\right), \tag{54}$$

$$\dot{P}_{mi} = \frac{1}{T_i}\left(-P_{mi} + u_{gi}\right) \tag{55}$$

where $i = 1, 2$, $P_{e1} = E_1E_2\left[B_{12}\sin\delta_{12} + G_{12}\cos\delta_{12}\right]$, $P_{e2} = -P_{e1} + E_2\frac{V_s}{x_{ds}}\sin\delta_2$, $\delta_i(t)$ is the angle of the i-th generator, $\delta_{ij} = \delta_i - \delta_j$; ω_i is the relative rotor speed; P_{mi} is the mechanical input power; P_{mi} and P_{ei} are the mechanical power and the electrical power; E'_{qi} is the transient EMF in quadrature axis, and is assumed to be constant under high-gain SCR controllers; D_i, H_i, and T_i are the damping constant, the inertia constant and the governor time constant; B_{12}, G_{12}, V_s, and x_{ds} are constant numbers.

Similarly as in [6], system (53)-(55) can be put into the following form:

$$\Delta\dot{\delta}_i = \Delta\omega_i, \tag{56}$$

$$\Delta\dot{\omega}_i = -\frac{D_i}{2H_i}\Delta\omega_i, + \frac{\omega_0}{2H_i}\Delta P_{mi}, \tag{57}$$

$$\Delta\dot{P}_{mi} = \frac{1}{T_i}\left[-\Delta P_{mi} + n_i + u_i + h_i\right] \tag{58}$$

where $\Delta\delta_i = \delta_i - \delta_{i0}$, $\Delta\omega_i = \omega_i - \omega_{i0}$, $\Delta P_{mi} = P_{mi} - P_{ei}$, $u_i = u_{gi} - P_{ei}$, $n_1 = 0$, $n_2 = -\frac{E_2V_s}{x_{ds}}\cos\delta_2\Delta\omega_2$, $h_1 = -E_1E_2\left[B_{12}\cos\delta_{12} - G_{12}\sin\delta_{12}\right]\omega_{12}$, $h_2 = -h_1$.

Only for simulation purpose, parameters are set to be $\omega_0 = 314.16$, $D_1 = 1$, $H_1 = 2$, $T_1 = 0.3$, $E_1 = 3.5$, $\delta_{10} = 2$, $H_2 = 1$, $D_2 = 1$, $T_2 = 0.5$, $E_2 = 1$, $\delta_{20} = 2.5$, $B_{12} = 1.5$, $G_{12} = 0.4$, $V_s = 1$, $x_{ds} = 1$.

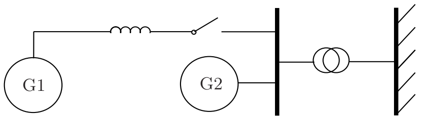

Fig. 1. A two-machine power system

Robust-ADP learning is performed for each of the two generators from $t = 0s$ to $t = 1s$. The cost functions for the first and second generators are set to be

$$\begin{aligned} & V^{(1)}(\Delta\delta_1(t), \Delta\omega_1(t), \Delta P_{m1}(t), u_1(t)) \\ = & \int_t^\infty [2\Delta\delta_1^2 + \Delta\omega_1^2 + \Delta P_{m1}^2) + 0.0001u_1^2]d\tau, \end{aligned}$$

and

$$\begin{aligned} & V^{(2)}(\Delta\delta_2(t), \Delta\omega_2(t), \Delta P_{m2}(t), u_2(t)) \\ = & \int_t^\infty [2\Delta\delta_2^2 + \Delta\omega_2^2 + \Delta P_{m2}^2) + 0.0001u_2^2]d\tau. \end{aligned}$$

For each $i = 1, 2$, we use the following vectors of basis functions, which are polynomials in $\Delta\delta_i$,$\Delta\omega_i$, and ΔP_{mi}.

$$\begin{aligned} \Phi^{(i)} &= [\Delta\delta_i^2, \Delta\omega_i^2, \Delta P_{mi}^2, \Delta\delta_i\Delta\omega_i, \Delta\delta_i\Delta P_{mi}, \Delta P_{mi}\Delta\omega_i]^T, \\ \Psi^{(i)} &= [\Delta\delta_i, \Delta\omega_i, \Delta P_{mi}, \Delta\delta_i^2, \Delta\omega_i^2, \Delta P_{mi}^2, \Delta\delta_i\Delta\omega_i, \\ &\quad \Delta\delta_i\Delta P_{mi}, \Delta P_{mi}\Delta\omega_i]^T. \end{aligned}$$

After 10 iterations, the approximated optimal cost functions are

$$\begin{aligned} \hat{V}^{(1)*} &= [1.4449\ 0.0157\ 0.0036\ 0.0438\ 0.0061\ 0.0085]\Phi^{(1)}, \\ \hat{V}^{(2)*} &= [1.4459\ 0.0094\ 0.0055\ 0.0265\ 0.0080\ 0.0120]\Phi^{(2)} \end{aligned}$$

and the approximated optimal control policies are

$$\begin{aligned} \hat{u}_1^* &= [-141.4214\ \ -101.7815\ \ -120.6443\ 0\ 0\ 0\ 0\ 0\ 0]\Psi^{(1)}, \\ \hat{u}_2^* &= [-142.5598\ \ -101.6468\ \ -144.4849\ \ -0.5219 \\ &\quad -1.4874\ \ -0.1823\ 0.3094\ \ -0.0167\ \ -1.5829]\Psi^{(2)}. \end{aligned}$$

Finally, we redesign $\hat{u}_1^*$ according to (19)-(22) with $\rho(s) = 10s$, and obtain a new control policy $u_1 = \hat{u}_1^* + u^{\mathrm{r}}$ for the first generator.

Trajectories of the control inputs and the angles of the two-machine system are illustrated in Figure 2. The performance of the angles without robust-ADP design is shown in Figure 3. It can be seen that, under the proposed robust-ADP control policies, the oscillation between the two generators is significantly reduced and asymptotic stability is attained.

V. Conclusions

In this paper, a robust optimal controller design for uncertain nonlinear systems has been presented from a perspective of robust-ADP. As a natural extension of the ADP methodology, the proposed robust-ADP framework allows for the presence of dynamic uncertainty with unmeasured state and unknown system order/dynamics. In addition, the theory of ADP has been integrated for the first time with tools from modern nonlinear control theory, such as the nonlinear small-gain theorem [12], [13], for robust optimal control design.

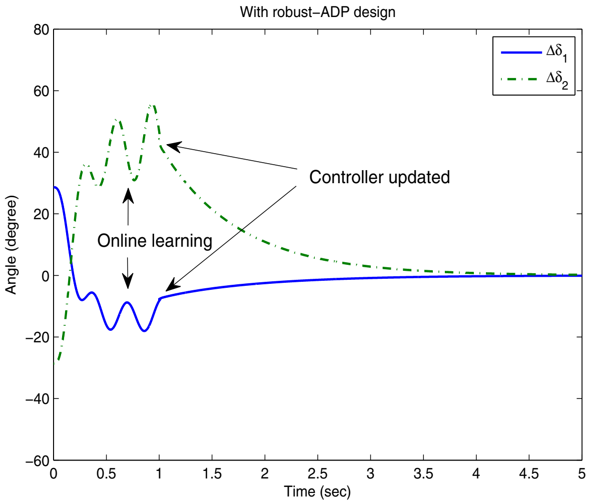

Fig. 2. System performance with robust-ADP.

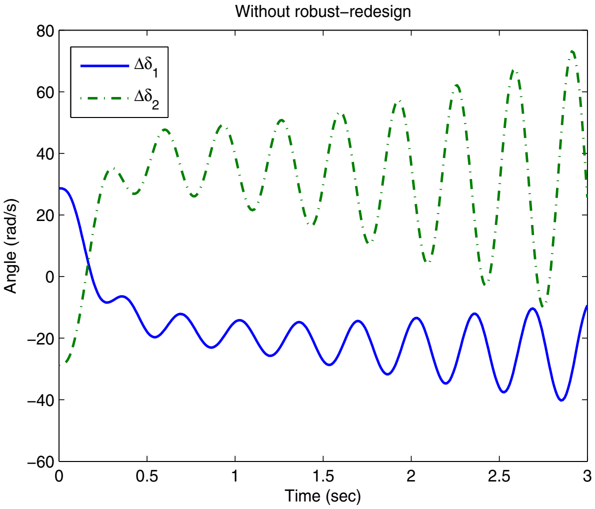

Fig. 3. System performance without robust-redesign.

A practical robust-ADP-based online learning algorithm has been developed and applied to the robust optimal controller design for a two-machine power system.

References

[1] M. Abu-Khalaf, and F. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.

[2] L. C. Baird III, "Reinforcement learning in continuous time: Advantage updating," in *Proceedings of International Conference on Neural Networks*, vol. 4, pp. 2448–2453, 1994.

[3] R. Beard, G. Saridis, and J. Wen, "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, 1997.

[4] K. Doya, "Reinforcement learning in continuous time and space," *Neural Computation*, vol. 12, no. 1, pp. 219–245, 2000.

[5] C. E. Garcia, D. M. Prett, and M. Morari (1989). "Model predictive control: Theory and practice–A survey," *Automatica*, vol. 25, no. 3, pp. 335–348, 1989.

[6] G. Guo, Y. Wang, and D. J. Hill, "Nonlinear output stabilization control for multimachine power systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 47, no. 1, 2000.

[7] A. Isidori, *Nonlinear Control Systems. vol. II*, Springer-Verlag, 1999.

[8] 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.

[9] Y. Jiang and Z. P. Jiang, "Robust adaptive dynamic programming with an application to power systems", submitted to *IEEE Transactions on Neural Networks and Learning Systems*.

[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, IEEE Press Computational Intelligence Series, 2012.

[12] 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.

[13] 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.

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

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

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

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

[18] 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.

[19] D. Q. Mayne and H. Michalska, "Receding horizon control of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 35, no. 7, pp. 814-824, 1990.

[20] 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.

[21] E. D. Sontag, "Smooth stabilization implies coprime factorization," *IEEE Transactions on Automatic Control*, vol. 34, pp. 435-443, 1989.

[22] E. D. Sontag, *Input to state stability: Basic concepts and results,* Nonlinear and Optimal Control Theory, ed. P. Nistri and G. Stefani, Berlin: Springer-Verlag, pp. 163–220, 2007.

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

[24] A. Teel and L. Praly, "Tools for semiglobal stabilization by partial state and output feedback", *SIAM Journal on Control and Optimization*, vol. 33, no. 5, pp. 1443–1488, 1995.

[25] 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.

[26] 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.

[27] C. Watkins, *Learning from delayed rewards*, PhD thesis, King's College of Cambridge, UK, 1989.

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

[29] P. J. Werbos, "Neural networks for control and system identification," *Proceedings of IEEE Conference on Decision and Control*, pp. 260-265, 1989.

[30] P. J. Werbos, "A menu of designs for reinforcement learning over time," *Neural Networks for Control*, pp. 67-95, ed. W. T. Miller, R. S. Sutton, P. J. Werbos, Cambridge: MIT Press, 1991.

[31] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, pp. 493-525, 1992.

[32] P. J. Werbos, "Intelligence in the brain: A theory of how it works and how to build it," *Neural Networks*, vol. 22, no. 3, pp. 200-212, 2009.

[33] K. G. Vamvoudakis, F. L. Lewis, and G. R. Hudas. "Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality,", *Automatica*, vol. 48, no. 8, pp. 1598-1611, 2012.