J Control Theory Appl 2011 9 (3) 361–369
DOI 10.1007/s11768-011-0178-0

Online optimal control of nonlinear discrete-time systems using approximate dynamic programming

Travis DIERKS [1], Sarangapani JAGANNATHAN [2]

1.DRS Sustainment Systems, Inc., 201 Evans Lane, St. Louis, MO 63121, U.S.A.;

2.Department of Electrical and Computer Engineering, Missouri University of Science and Technology, Rolla, MO 65409, U.S.A.

Abstract: In this paper, the optimal control of a class of general affine nonlinear discrete-time (DT) systems is undertaken by solving the Hamilton Jacobi-Bellman (HJB) equation online and forward in time. The proposed approach, referred normally as adaptive or approximate dynamic programming (ADP), uses online approximators (OLAs) to solve the infinite horizon optimal regulation and tracking control problems for affine nonlinear DT systems in the presence of unknown internal dynamics. Both the regulation and tracking controllers are designed using OLAs to obtain the optimal feedback control signal and its associated cost function. Additionally, the tracking controller design entails a feedforward portion that is derived and approximated using an additional OLA for steady state conditions. Novel update laws for tuning the unknown parameters of the OLAs online are derived. Lyapunov techniques are used to show that all signals are uniformly ultimately bounded and that the approximated control signals approach the optimal control inputs with small bounded error. In the absence of OLA reconstruction errors, an optimal control is demonstrated. Simulation results verify that all OLA parameter estimates remain bounded, and the proposed OLA-based optimal control scheme tunes itself to reduce the cost HJB equation.

Keywords: Online nonlinear optimal control; Hamilton Jacobi-Bellman; Online approximators; Discrete-time systems

1 Introduction

Online approximators (OLAs) have been widely used in the controller designs for uncertain discrete-time (DT) nonlinear systems [1]; however, system stability is typically the sole consideration although optimality is generally preferred. While linear systems accompanied by quadratic cost functions can achieve optimal control by solving the Riccati equation [2], the optimal control of nonlinear DT systems often requires solving the nonlinear Hamilton-Jacobi-Bellman (HJB) equation. To extend the results of linear optimal control theory to nonlinear systems, the state-dependent Riccati equation [3] is proposed for suboptimal control under certain tight assumptions including the need for full system dynamics.

To avoid solving the HJB equation, approximate solutions to the HJB equation have been proposed [4–6]. In [4], a Chebyshev series was proposed for approximating the system dynamics, boundary conditions, and cost function. In [5,6], neural networks (NNs) are utilized to solve the DT nonlinear optimal regulation in an offline manner by ignoring NN reconstruction errors and assuming complete system dynamics. Recently, online methods to solve the continuous HJB equation were presented in [7] for linear systems using online policy iterations based on adaptive control.

In addition to the optimal regulation problem, the optimal tracking control problem has also recently been considered. In [8], the authors consider the H_∞ optimal tracking control by linearizing the error equations about the origin yielding a locally optimal control law. The effort in [9] considers the receding horizon optimal tracking control by linearizing the nonlinear error dynamics about the origin. In [10], authors consider the HJB equation and employ similar techniques such as [6] to find an offline solution to the optimal tracking control problem.

In this work, a novel approach to the optimal regulation of nonlinear DT systems is adopted to solve the HJB equation online. Using an initial stabilizing control, an OLA is tuned online at each time step to learn the HJB equation in contrast to normal approaches to ADP [7], which use two indices: a time index for the system dynamics and a policy iteration index for the cost function approximator. Then, a second OLA is utilized that minimizes the HJB function based on the information provided by the first OLA. Knowledge of the internal system dynamics is not required while the control coefficient matrix alone is needed although it can be relaxed by introducing an additional OLA [11].

Subsequently, this approach is extended to include optimal tracking control even with unknown internal dynamics by using a third OLA to approximate the feedforward part of the control input that is normally required for tracking [1,2]. Novel online parameter tuning laws for the OLAs are derived, and Lyapunov theory is utilized to demonstrate the stability of the system while explicitly considering the OLA approximation errors in contrast to the other works [5,6,10].

Received 29 July 2010; revised 11 April 2011.
This work was partly supported by the National Science Foundation (No.ECCS#0621924, ECCS-#0901562), and the Intelligent Systems Center.

2 Background

Consider the affine nonlinear DT system

$$x_{k+1} = f(x_k) + g(x_k)u(x_k), \tag{1}$$

where $x_k \in \mathbb{R}^n$, $f(x_k) \in \mathbb{R}^n$, $g(x_k) \in \mathbb{R}^{n\times m}$ satisfies $\|g(x_k)\|_{\mathrm{F}} \leqslant g_{\mathrm{M}}$, where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm [1], and $u(x_k) \in \mathbb{R}^m$ is the control input. Without loss of generality, assume that $x = 0$ is a unique equilibrium point on a compact set Ω while the states are considered measurable. In order to control (1) in an optimal manner, select the control sequence $u(x_k)$ that minimizes the infinite horizon cost function as [6]

$$\begin{aligned} J(x_k) = {} & Q(x_k) + u(x_k)^{\mathrm{T}} R u(x_k) \\ & + J(f(x_k) + g(x_k)u(x_k)) \end{aligned} \tag{2}$$

for all x_k, where with $Q(x_k) > 0$ and $R \in \mathbb{R}^{m\times m}$ is a symmetric positive definite matrix. Selecting $Q(x_k) > 0$ ensures that variations in any direction of the state affect the cost, which can be linked to the observability condition [2]. In addition, it is required that the control policy $u(x_k)$ be admissible [5] and $J(x_k = 0) = 0$ so that $J(x_k)$ serves as a Lyapunov function.

Using Bellman's principle of optimality [2], the infinite horizon optimal cost function, $J^*(x_k)$, is time invariant and satisfies the DT HJB equation [2]

$$J^*(x_k) = \min_{u(x_k)} \left(r(x_k, u(x_k)) + J(f(x_k) + g(x_k)u(x_k))\right),$$

where $r(x_k, u(x_k)) = Q(x_k) + u(x_k)^{\mathrm{T}} R u(x_k)$. The optimal control $u^*(x_k)$ that minimizes $J^*(x_k)$ is found by applying the stationary condition (see [2])

$$\frac{\partial J^*(x_k)}{\partial u(x_k)} = 2Ru(x_k) + g(x_k)^{\mathrm{T}} \frac{\partial J^*(x_{k+1})}{\partial x_{k+1}} = 0,$$

and is shown to be (see [2])

$$u^*(x_k) = -\frac{1}{2} R^{-1} g(x_k)^{\mathrm{T}} \frac{\partial J^*(x_{k+1})}{\partial x_{k+1}}. \tag{3}$$

The optimal control (3) is generally unavailable for nonlinear DT systems due to its dependence on the future state vector x_{k+1}. To circumvent this deficiency, a new approach to online optimal control is presented next.

3 Near optimal regulation of DT systems

The proposed approach entails two OLAs: one OLA to learn the HJB equation and a second OLA to learn the control signal that minimizes the estimated HJB equation. Using the function approximation property of OLAs [1], the cost function (2) and feedback control policy (3) have OLA representations on a compact set expressed as

$$J(x_k) = \Phi^{\mathrm{T}} \sigma(x_k) + \varepsilon_{Jk}, \tag{4}$$

and

$$u(x_k) = \Theta^{\mathrm{T}} \vartheta(x_k) + \varepsilon_{uk}, \tag{5}$$

respectively, where $\Phi \in \mathbb{R}^{L_1}$ and $\Theta \in \mathbb{R}^{L_2\times m}$ are the constant target OLA parameters, $\varepsilon_{Jk} \in \mathbb{R}$ and $\varepsilon_{uk} \in \mathbb{R}^m$ are the bounded approximation errors, and $\sigma(\cdot) \in \mathbb{R}^{L_1}$ and $\vartheta(\cdot) \in \mathbb{R}^{L_2}$ are the activation functions for the cost and control signal OLA schemes, respectively. Here, L_1 and L_2 define the dimension of Φ and Θ, respectively. The upper bounds for the ideal OLA parameters are taken as $\|\Phi\| \leqslant \Phi_{\mathrm{M}}$ and $\|\Theta\|_{\mathrm{F}} \leqslant \Theta_{\mathrm{M}}$, where Φ_{M} and Θ_{M} are positive constants [1], respectively, and the approximation errors are assumed to be bounded above as $\|\varepsilon_{Jk}\| \leqslant \varepsilon_{J\mathrm{M}}$ and $\|\varepsilon_{uk}\| \leqslant \varepsilon_{u\mathrm{M}}$, where $\varepsilon_{J\mathrm{M}}$ and $\varepsilon_{u\mathrm{M}}$ are positive constants [1]. In addition, the gradient of the approximation error is assumed to be bounded above as $\left\|\dfrac{\partial \varepsilon_{Jk}}{\partial x_{k+1}}\right\|_{\mathrm{F}} \leqslant \varepsilon'_{J\mathrm{M}}$, where $\varepsilon'_{J\mathrm{M}}$ is also a positive constant [12].

To begin, the HJB function OLA is considered next.

3.1 Approximation of the optimal cost function

Now, the cost function (2) is approximated by an OLA and written as

$$\hat{J}(x_k) = \hat{\Phi}_k^{\mathrm{T}} \sigma(x_k), \tag{6}$$

where $\hat{J}(x_k)$ represents an approximated value of the cost function (2), and $\hat{\Phi}_k$ is the estimate of the target OLA parameter vector Φ. The basis function should satisfy $\|\sigma(0)\| = 0$ for $\|x\| = 0$ to ensure $\hat{J}(0) = 0$ can be satisfied [2].

It is observed that (2) represents a nonlinear DT Lyapunov equation that can be rewritten as $J(f(x_k) + g(x_k)u(x_k)) - J(x_k) + r(x_k, u(x_k)) = 0$. However, this relationship is not guaranteed to hold when the estimated cost function in (6) is used. Therefore, the residual or cost-to-go (CTG) error associated with (6) can be written as

$$\hat{J}(x_{k+1}) - \hat{J}(x_k) + r(x_k, u(x_k)) = e_{Jk}$$

or

$$e_{Jk} = r(x_k, u(x_k)) + \hat{\Phi}_k^{\mathrm{T}} \Delta\sigma(x_{k+1}), \tag{7}$$

where $\Delta\sigma(x_{k+1}) = \sigma(x_{k+1}) - \sigma(x_k)$. Next, we define an auxiliary CTG error vector as

$$E_{Jk} = Y_k + \hat{\Phi}_k^{\mathrm{T}} X_k \in \mathbb{R}^{1\times(1+j)}, \tag{8}$$

where

$$\begin{aligned} Y_k = {} & [r(x_k, u(x_k)) \;\; r(x_{k-1}, u(x_{k-1})) \;\cdots \\ & r(x_{k-j}, u(x_{k-j}))], \\ X_k = {} & [\Delta\sigma(x_{k+1}) \;\; \Delta\sigma(x_k) \;\cdots\; \Delta\sigma(x_{k+1-j})], \end{aligned}$$

$0 < j < k - 1 \in \mathbb{N}$ and $\mathbb{N}$ being the set of natural real numbers. It is clear that (8) represents a time history of the previous $j + 1$ CTG errors (7) recalculated using the most recent $\hat{\Phi}_k$.

Moving on, define the OLA parameter update law as

$$\hat{\Phi}_{k+1} = \hat{\Phi}_k - \frac{\alpha_J X_k E_{Jk}^{\mathrm{T}}}{\|X_k X_k^{\mathrm{T}} + I\|_{\mathrm{F}}}, \tag{9}$$

where $0 < \alpha_J < 1$ is a small positive design parameter and I is an identity matrix of appropriate dimension. It is observed that the auxiliary CTG error (8) becomes zero when $x_k = 0$ because the cost functions (4) and (6) are zero at $x_k = 0$. Thus, once the system states have converged to zero, the cost function OLA is no longer updated. This can be viewed as a persistency of excitation (PE) requirement for the inputs to the cost function OLA wherein the system, states must be persistently existing long enough for the optimal cost function to be learned.

To obtain the OLA parameter estimation error dynamics, rewrite (2) using the target OLA representation (6) as

$$r(x_k, u(x_k)) = -\Phi^{\mathrm{T}} \Delta\sigma(x_{k+1}) - \Delta\varepsilon_{Jk}, \tag{10}$$

where $\Delta\varepsilon_{Jk} = \varepsilon_{Jk+1} - \varepsilon_{Jk}$. Substituting (10) into (7) results in

$$e_{Jk} = -\tilde{\Phi}_k^{\mathrm{T}} \Delta\sigma(x_{k+1}) - \Delta\varepsilon_{Jk}, \tag{11}$$

where $\tilde{\Phi}_k = \Phi - \hat{\Phi}_k$ is the cost parameter estimation error. Similarly, (8) can be rewritten as

$$E_{Jk} = -\tilde{\Phi}_k^{\mathrm{T}} X_k - \Psi_k, \tag{12}$$

where $\Psi_k = [\Delta\varepsilon_{Jk} \;\; \Delta\varepsilon_{Jk-1} \;\cdots\; \Delta\varepsilon_{Jk-j}]$ and $\|\Psi_k\|^2 \leqslant \Psi_{\mathrm{M}}^2$. Now, observing $\tilde{\Phi}_{k+1} = \Phi - \hat{\Phi}_{k+1}$ and using (9) and (12) results in the OLA parameter estimation error dynamics to be expressed as

$$\tilde{\Phi}_{k+1} = \tilde{\Phi}_k + \frac{\alpha_J X_k E_{Jk}^{\mathrm{T}}}{\|X_k X_k^{\mathrm{T}} + I\|_{\mathrm{F}}}. \tag{13}$$

Next, the action network that generates the optimal feedback signal (3) is considered.

3.2 Estimation of the optimal feedback control signal

The objective of this section is to find the control policy that minimizes the approximated cost function (6). To begin the development of the feedback control policy, we define the OLA approximation of (5) to be

$$\hat{u}(x_k) = \hat{\Theta}_k^{\mathrm{T}} \vartheta(x_k), \tag{14}$$

where $\hat{\Theta}_k$ is the estimated value of Θ_k.

Next, the action error is defined as the difference between the feedback control applied to (1) and the control signal that minimizes the estimated cost function (6), which is denoted as

$$\tilde{u}(x_k) = \hat{\Theta}_k^{\mathrm{T}} \vartheta(x_k) + R^{-1} g^{\mathrm{T}}(x_k) \Big(\frac{\partial \sigma x_{k+1}}{\partial x_{k+1}}\Big)^{\mathrm{T}} \frac{\hat{\Phi}_k}{2}, \tag{15}$$

and the control OLA parameter update is defined to be

$$\hat{\Theta}_{k+1} = \hat{\Theta}_k - \frac{\alpha_u \vartheta(x_k) \tilde{u}_k^{\mathrm{T}}}{(\vartheta^{\mathrm{T}}(x_k)\vartheta(x_k) + 1)}, \tag{16}$$

where $0 < \alpha_u < 1$ is a positive design parameter. Because the control policy $u(x_k)$ in (5) minimizes the cost function (4), from (3) we can write

$$\begin{aligned} 0 = {} & \varepsilon_{uk} + \Theta^{\mathrm{T}} \vartheta(x_k) + \frac{1}{2} R^{-1} g^{\mathrm{T}}(x_k) \frac{\partial \varepsilon_{Jk+1}}{\partial x_{k+1}} \\ & + \frac{1}{2} R^{-1} g^{\mathrm{T}}(x_k) \frac{\partial \sigma(x_{k+1})}{\partial x_{k+1}} \Phi. \end{aligned} \tag{17}$$

Then, defining the action OLA parameter estimation error as $\tilde{\Theta}_k = \Theta - \hat{\Theta}_k$ and subtracting (16) from (15) yields

$$\begin{aligned} \tilde{u}(x_k) = {} & -\tilde{\Theta}_k^{\mathrm{T}} \vartheta(x_k) - \frac{1}{2} R^{-1} g^{\mathrm{T}}(x_k) \frac{\partial \sigma(x_{k+1})}{\partial x_{k+1}}^{\mathrm{T}} \tilde{\Phi}_k \\ & - \tilde{\varepsilon}_{uk}, \end{aligned} \tag{18}$$

where $\tilde{\varepsilon}_{uk} = \varepsilon_{uk} + \frac{1}{2} R^{-1} g^{\mathrm{T}}(x_k) \frac{\partial \varepsilon_{Jk+1}}{\partial x_{k+1}}$ and satisfies $\|\tilde{\varepsilon}_{uk}\| \leqslant \tilde{\varepsilon}_{u\mathrm{M}}$ for a positive constant $\tilde{\varepsilon}_{u\mathrm{M}}$.

As a final step, we form the parameter estimation error dynamics as

$$\tilde{\Theta}_{k+1} = \tilde{\Theta}_k + \frac{\alpha_u \vartheta(x_k) \tilde{u}_k^{\mathrm{T}}}{(\vartheta^{\mathrm{T}}(x_k)\vartheta(x_k) + 1)}, \tag{19}$$

and the closed loop nonlinear system dynamics (1) can be rewritten in terms of $u^*(x_k)$ and $\tilde{\Theta}_k$ as

$$\begin{aligned} x_{k+1} = {} & f(x_k) + g(x_k)\hat{u}(x_k) \\ = {} & f(x_k) + g(x_k)\hat{\Theta}^{\mathrm{T}} \vartheta(x_k) \\ = {} & f(x_k) + g(x_k) u^*(x_k) - g(x_k)\tilde{\Theta}^{\mathrm{T}} \vartheta(x_k) \\ & - g(x_k)\varepsilon_{uk}. \end{aligned} \tag{20}$$

Remark 1 To calculate the action error (14) and implement the OLA parameter update (15), knowledge of the input transformation matrix $g(x_k)$ is required. However, knowledge of the internal dynamics $f(x_k)$ can be avoided by rewriting (2) in terms of x_{k+1} as noted in [6]. In contrast to [6], this work utilizes online learning and requires an initial stabilizing control to ensure the nonlinear system remains stable while the optimal control signal and cost function are computed because nonlinear systems are known to have finite escape times [13].

The following theorem will show that the cost function and action network OLA parameter estimation errors are uniformly ultimately bounded (UUB) [1]. Furthermore, the estimated control input (14) approaches the optimal control signal with small bounded error, which is a function of the OLA reconstruction errors ε_J and ε_u, and if the OLA approximation errors are considered to be negligible [5, 6], asymptotic convergence is observed.

The initial system states are considered to reside in a compact set $\Omega \subset \mathbb{R}^n$ by using the initial stabilizing control input u_{0k}. Moreover, sufficient conditions for the OLA tuning gains, α_J and α_u, are derived to ensure that all future states never leave the compact set. As a result, in the compact Ω, the cost function and its gradient as well as the basis function of the action network are bounded according to $\|\sigma(x_k)\| \leqslant \sigma_{\mathrm{M}}$, $\Big\|\frac{\partial \sigma(x_k)}{\partial x_k}\Big\| \leqslant \sigma'_{\mathrm{M}}$, and $\|\vartheta(x_k)\| \leqslant \vartheta_{\mathrm{M}}$, respectively. A similar approach has been used in several other well known efforts in control theory [14] to establish bounds on functions of the system states x_k.

Theorem 1 (Convergence of the optimal control signal) Let $u_0(x_k)$ be any initial admissible control policy for the class nonlinear controllable systems in (1). Let the OLA parameter tuning for the cost function estimator and the action network be provided by (9) and (15), respectively. Then, there exist positive constants α_J and α_u such that the system states and the cost and action network OLA parameter estimate errors $\tilde{\Phi}_k$ and $\tilde{\Theta}_k$, respectively, are all UUB for all $k \geqslant k_0 + T$. Furthermore, $\|\hat{u}(x_k) - u^*(x_k)\| \leqslant \varepsilon_{\mathrm{r}}$ for a small positive constant ε_{r}.

Proof of Theorem 1 is shown in Appendix.

Remark 2 The cost and control OLA parameter update laws are both written in terms of the time index k instead of a policy iteration index commonly used in standard ADP approaches [7]. Therefore, the conclusions of Theorem 1 illustrate the convergence of the ADP scheme and the boundedness of the system states simultaneously.

In the next section, the optimal online regulator is extended to consider the optimal tracking problem.

4 Near optimal tracking control

In this section, the optimal tracker is considered as an extension of regulation [2]. In addition to the assumptions on the system dynamics (1) presented in Section 2, in this section it is assumed that $m = n$ so that $g(x_k)$ is invertible. The objective for the optimal tracking problem is to find the optimal control sequence to make the nonlinear system in (1) track a desired trajectory $x_{\mathrm{d}k}$. To begin the development, define the dynamics of the desired trajectory as (see [8])

$$x_{\mathrm{d}k+1} = f(x_{\mathrm{d}k}) + g(x_k) u_{\mathrm{d}}(x_{\mathrm{d}k}, x_k), \tag{21}$$

where $f(x_{\mathrm{d}k})$ is the internal dynamics of the nonlinear system (1) rewritten in terms of the desired state $x_{\mathrm{d}k}$, $g(x_k)$ is the input transformation matrix presented in (1), and $u_{\mathrm{d}}(x_{\mathrm{d}k})$ is the control input to the desired system. It is useful to note that when $x_{\mathrm{d}k}$, $x_{\mathrm{d}k+1}$, and $f(x_{\mathrm{d}k})$ are known, $u_{\mathrm{d}}(x_{\mathrm{d}k}, x_k)$ is obtained by rearranging (20) to give

$$u_{\mathrm{d}}(x_{\mathrm{d}k}, x_k) = g(x_k)^{-1}(x_{\mathrm{d}k+1} - f_{\mathrm{d}}(x_{\mathrm{d}k})). \quad (22)$$

Next, define the tracking error as

$$e_k = x_k - x_{\mathrm{d}k}. \quad (23)$$

Then, using (1) and (20), the tracking error dynamics are

$$\begin{aligned} e_{k+1} &= f(x_k) + g(x_k)u(x_k) - x_{\mathrm{d}k+1} \\ &= f_e(e_k) + g(x_k)u_e(e_k), \end{aligned} \quad (24)$$

where $f_e(e_k) = f(x_k) - f(x_{\mathrm{d}k})$ and

$$u_e(e_k) = u(x_k) - u_{\mathrm{d}}(x_{\mathrm{d}k}, x_k). \quad (25)$$

Considering $u_e(e_k)$ as the control input for (23), $u_e(e_k)$ is an admissible control policy with $e_k = 0$ being an equilibrium point of (23). To convert the nonlinear tracking into a regulation problem, the infinite horizon cost function (2) is rewritten in terms of e_k and $u_e(e_k)$ as

$$J_e(e_k) = Q_e(e_k) + u_e(e_k)^{\mathrm{T}} R_e u_e(e_k) + J_e(e_{k+1}), \quad (26)$$

with $Q_e(e_k) > 0$ and $R_e \in \mathbb{R}^{m\times m}$ is positive definite. Because $u_e(e_k)$ is admissible, (26) is finite. The optimal control input that minimizes (26) is found by solving $\dfrac{\partial J_e(e_k)}{\partial u_e(e_k)} = 0$ as

$$u_e^*(e_k) = -\frac{1}{2}R_e^{-1}g^{\mathrm{T}}(x_k)\frac{(\partial J_e^*(e_{k+1})}{\partial e_{k+1})}, \quad (27)$$

or

$$u^*(x_k) = u_{\mathrm{d}}(x_{\mathrm{d}k}) - \frac{1}{2}R_e^{-1}g^{\mathrm{T}}(x_k)(\frac{\partial J_e^*(e_{k+1})}{\partial e_{k+1}}). \quad (28)$$

It is observed that the optimal tracking control input (28) consists of a predetermined feedforward term, $u_{\mathrm{d}}(x_{\mathrm{d}k}, x_k)$, and an optimal feedback term that is a function of the gradient of the optimal cost function. Additionally, implementation of the feedforward term requires knowledge of the internal dynamics $f(x_k)$ and control coefficient matrix $g(x_k)$. In this effort, $f(x_k)$, $J_e(e_k)$ and $\dfrac{\partial J_e(e_{k+1})}{\partial e_{k+1}}$ are all unknown. To mitigate these deficiencies, the universal approximation properties [1] of OLAs are utilized as described next.

The proposed solution for achieving nonlinear optimal tracking control entails three portions: an HJB function estimator that evaluates the performance of the error system by approximating (26), a feedback system that is designed to produce a nearly optimal portion of the control signal (27), and a feedforward design to produce the feedforward control input (28) by approximating $f(x_{\mathrm{d}k})$. Using the approximation property of OLAs [1], the cost function (26), feedback control policy (27), and desired internal dynamics, $f(x_{\mathrm{d}k})$, in (20) have OLA representations on a compact set expressed as

$$J_e(e_k) = \Phi_e^{\mathrm{T}}\sigma_e(e_k) + \varepsilon_{Je}, \quad (29)$$

$$u_e(e_k) = \Theta_e^{\mathrm{T}}\vartheta_e(e_k) + \varepsilon_{ue}, \quad (30)$$

and

$$f(x_{\mathrm{d}k}) = \Omega_{\mathrm{d}}^{\mathrm{T}}\phi(x_{\mathrm{d}k}) + \varepsilon_{\mathrm{d}}(x_{\mathrm{d}k}), \quad (31)$$

respectively, where Φ_e, Θ_e, and Ω_{d} are the constant target OLA parameters, ε_{Je}, ε_{ue}, and ε_{d} are the bounded approximation errors, and $\sigma_e(\cdot)$, $\vartheta_e(\cdot)$ and $\phi(\cdot)$ are the linearly independent vector activation functions for the cost, feedback, and feedforward control networks, respectively [1]. The upper bounds for the ideal OLA parameters are taken as $\|\Phi_e\| \leqslant \Phi_{e\mathrm{M}}$, $\|\Theta_e\|_{\mathrm{F}} \leqslant \Theta_{e\mathrm{M}}$, and $\|\Omega_{\mathrm{d}}\|_{\mathrm{F}} \leqslant \Omega_{\mathrm{dM}}$ where $\Phi_{e\mathrm{M}}$, $\Theta_{e\mathrm{M}}$, and Ω_{dM} are positive constants [1] while the approximation errors are assumed to be bounded above such that $\|\varepsilon_{Je}\| \leqslant \varepsilon_{Je\mathrm{M}}$, $\|\varepsilon_{ue}\| \leqslant \varepsilon_{ue\mathrm{M}}$, and $\|\varepsilon_{\mathrm{d}}\| \leqslant \varepsilon_{\mathrm{dM}}$ where $\varepsilon_{Je\mathrm{M}}$, $\varepsilon_{ue\mathrm{M}}$ and $\varepsilon_{\mathrm{dM}}$ are positive constants [1]. In addition, the gradient of the cost function approximation error is considered to be bounded according to $\left\|\dfrac{\partial \varepsilon_{Je}}{\partial e_{k+1}}\right\|_{\mathrm{F}} \leqslant \varepsilon'_{Je\mathrm{M}}$ where $\varepsilon'_{Je\mathrm{M}}$ is also a positive constant [12]. As in Section 3, the basis functions and the gradient of the cost function basis vector are considered to be bounded on a compact set [14] according to $\|\sigma_e(\cdot)\| \leqslant \sigma_{e\mathrm{M}}$, $\|\vartheta_e(\cdot)\| \leqslant \vartheta_{e\mathrm{M}}$, $\|\phi(\cdot)\| \leqslant \phi_{\mathrm{M}}$, and $\left\|\dfrac{\partial \sigma_e(\cdot)}{\partial(\cdot)}\right\| \leqslant \sigma'_{e\mathrm{M}}$ for known constants $\sigma_{e\mathrm{M}}$, $\vartheta_{e\mathrm{M}}$, ϕ_{M}, and $\sigma'_{e\mathrm{M}}$, respectively [1]. To begin, the design of the HJB function and optimal feedback approximators will be considered first.

4.1 Cost function and optimal feedback control

The objective of the optimal tracking control law is to stabilize system (23) while minimizing the cost function (26). To begin, (26) is approximated by an OLA as

$$\hat{J}_e(e_k) = \hat{\Phi}_{ek}^{\mathrm{T}}\sigma_e(e_k), \quad (32)$$

where $\hat{\Phi}_{ek}$ is the approximation for the ideal parameters Φ_e. The basis vector $\sigma_e(\cdot)$ is selected to satisfy $\|\sigma_e(0)\| = 0$ to facilitate $J_e(0) = 0$.

Similar to Section 3, it is observed that (26) represents a nonlinear DT Lyapunov equation for tracking; however, when the estimated cost function in (32) is used, a residual or CTG error, e_{Jek}, associated with (32) for tracking is required. The CTG error for tracking is defined identically to the CTG error for regulation, defined in (7) in Section 3, with x_k, $u(x_k)$ and $\hat{\Phi}_k^{\mathrm{T}}\Delta\sigma(x_{k+1})$ replaced by e_k, $u_e(e_k)$, and $\hat{\Phi}_{ek}^{\mathrm{T}}\Delta\sigma_e(e_{k+1})$, respectively. In addition, the auxiliary CTG error for tracking, E_{Jek}, is defined similarly to (8) with the regulation variables replaced by those for tracking as described above.

The OLA parameter update law is now defined as

$$\hat{\Phi}_{ek+1} = \hat{\Phi}_{ek} - \frac{\alpha_{Je}X_{ek}E_{Jek}^{\mathrm{T}}}{\left\|X_{ek}X_{ek}^{\mathrm{T}} + I\right\|_{\mathrm{F}}}, \quad (33)$$

where $0 < \alpha_{Je} < 1$ is a small positive design parameter, X_{ek} is defined similarly to X_k in (8), and I is an identity matrix of appropriate dimension. Following the steps described by (10)–(13) in Section 3, the OLA parameter estimation error dynamics for tracking $\tilde{\Phi}_{ek+1}$ are written similarly to (13).

Next, the action network that generates the optimal feedback signal that minimizes the approximated cost function (32) is considered. To begin, define the OLA approximation of (30) as

$$\hat{u}_e(e_k) = \hat{\Theta}_{ek}^{\mathrm{T}}\vartheta_e(e_k), \quad (34)$$

where $\hat{\Theta}_{ek}$ is the OLA estimate of the target value Θ_e. The

basis function, $\vartheta_e(\cdot)$, is selected to satisfy $\|\vartheta_e(0)\| = 0$ so that $\|\hat{u}_e(0)\| = 0$ can be satisfied as required for admissibility.

Next, the action error for tracking, $\tilde{u}_e(e_k)$, is defined to be the difference between the feedback control applied to (23) and the control signal that minimizes the estimated cost function (32), which is written similarly (14) in Section 3 with the regulation variables replaced by the appropriate variables defined for tracking. The proposed control OLA parameter update is defined to be

$$\hat{\Theta}_{ek+1} = \hat{\Theta}_{ek} - \frac{\alpha_{ue}\vartheta_e(e_k)\tilde{u}_{ek}^{\mathrm{T}}}{(\vartheta_e^{\mathrm{T}}(e_k)\vartheta(e_k)+1)}, \tag{35}$$

where $0 < \alpha_{ue} < 1$ is a positive design parameter. Defining the action OLA parameter estimation error for tracking as $\tilde{\Theta}_{ek} = \Theta_e - \hat{\Theta}_{ek}$ and following similar steps as those presented in (16) and (17) in Section 3, the action error for tracking can be written identically to (17) for regulation, and the parameter estimation error dynamics $\tilde{\Theta}_{ek+1}$, for tracking can be written identically to (18).

Next, the overall nearly optimal control signal is derived and system stability is investigated.

4.2 Nearly optimal control input

Recall that the optimal tracking control input (28) to system (1) is comprised of an optimal feedback term and a predetermined feedforward term, which is a function of the desired yet unknown internal dynamics,

$$f(x_{\mathrm{d}k}) \equiv f(x_k)|_{x_k = x_{x\mathrm{d}}}.$$

Therefore, the desired internal dynamics are approximated online by identifying the actual internal dynamics, and reevaluating the estimate using the desired state, $x_{\mathrm{d}k}$.

To begin the online identifier development, the nonlinear system dynamics (1) are rewritten as

$$x_{k+1} = \Omega_{\mathrm{d}}^{\mathrm{T}}\phi(x_k) + \varepsilon_{\mathrm{d}}(x_k) + g(x_k)u(x_k, x_{\mathrm{d}k}, e_k), \tag{36}$$

where $\Omega_{\mathrm{d}}^{\mathrm{T}}\phi(x_k) + \varepsilon_{\mathrm{d}}(x_k)$ is the OLA estimate defined in (31) reevaluated at x_k. As in the previous cases, it is assumed that there exists a nonzero lower bound such that $\phi_{\mathrm{dm}} \leqslant \|\phi(\cdot)\|$ for a positive constant ϕ_{dm}. The online identifier is then defined as

$$\begin{aligned}\hat{x}_{k+1} = {} & \hat{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k) + g(x_k)u(x_k, x_{\mathrm{d}k}, e_k) \\ & -K_{\mathrm{d}}\tilde{x}_k,\end{aligned} \tag{37}$$

where $\hat{x}_{k+1}$ is the estimate of x_{k+1}, $\hat{\Omega}_{\mathrm{d}k}$ is the parameter estimate of $\Omega_{\mathrm{d}k}$, $\tilde{x}_k = x_k - \hat{x}_k$, and $0 < K_{\mathrm{d}} < 1$ is a constant. Next, subtracting (37) from (36) and defining $\tilde{\Omega}_{\mathrm{d}k} = \Omega_{\mathrm{d}} - \hat{\Omega}_{\mathrm{d}k}$, the identifier error dynamics are

$$\tilde{x}_{k+1} = \tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k) + \varepsilon_{\mathrm{d}}(x_k) + K_{\mathrm{d}}\tilde{x}_k. \tag{38}$$

The parameter update law is now defined as

$$\hat{\Omega}_{\mathrm{d}k+1} = \hat{\Omega}_{\mathrm{d}k} + \alpha_{\mathrm{d}}\phi(x_k)(\tilde{x}_{k+1} + K_{\mathrm{d}}\tilde{x}_k)^{\mathrm{T}}, \tag{39}$$

where $0 < \alpha_{\mathrm{d}} < 1$ is a constant, and the parameter estimation error dynamics, $\tilde{\Omega}_{\mathrm{d}k+1} = \Omega_{\mathrm{d}} - \hat{\Omega}_{\mathrm{d}k+1}$, are found to be

$$\tilde{\Omega}_{\mathrm{d}k+1} = \tilde{\Omega}_{\mathrm{d}k} - \alpha_{\mathrm{d}}\phi(x_k)(\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k) + \varepsilon_{\mathrm{d}}(x_k))^{\mathrm{T}}. \tag{40}$$

Moving on, the feedforward control input can now be defined using the OLA representation (31) as

$$\begin{aligned}& u_{\mathrm{d}}(x_{\mathrm{d}k}, x_k) \\ & = g(x_k)^{-1}(x_{\mathrm{d}k+1} - (\Omega_{\mathrm{d}}^{\mathrm{T}}\phi(x_{\mathrm{d}k}) + \varepsilon_{\mathrm{d}}(x_{\mathrm{d}k}))).\end{aligned} \tag{41}$$

Now, the estimated feedforward control input is defined as

$$u_{\mathrm{d}}(x_{\mathrm{d}k}, x_k) = g(x_k)^{-1}(x_{\mathrm{d}k+1} - \hat{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_{\mathrm{d}k})), \tag{42}$$

where $\hat{f}(x_{\mathrm{d}k}) = \hat{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_{\mathrm{d}k})$ was used while observing the definition of $f(x_{\mathrm{d}k})$ in (20). Now, using (34) and (42), the estimate of the control input (28) is written as

$$u(x_k, x_{\mathrm{d}k}, e_k) = \hat{u}_{\mathrm{d}}(x_{\mathrm{d}k}, x_k) + \hat{u}_e(e_k), \tag{43}$$

and applying (43) to the nonlinear system (1) reveals

$$x_{k+1} = f(x_k) + g(x_k)(\hat{u}_{\mathrm{d}}(x_{\mathrm{d}k}, x_k) + \hat{u}_e(e_k))$$

or

$$e_{k+1} = f(x_k) - \hat{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_{\mathrm{d}k}) + g(x_k)\hat{u}_e(e_k)). \tag{44}$$

Then, adding and subtracting $f(x_{\mathrm{d}k})$ and $g(x_k)u_e(e_k)$ to (44), recalling the OLA representations of $f(x_{\mathrm{d}k})$ and $u_e(e_k)$ in (31) and (30), respectively, (44) is rewritten as

$$\begin{aligned}e_{k+1} = {} & f_e(e_k) + g(x_k)u_e(e_k) + \tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_{\mathrm{d}k}) \\ & -g(x_k)\tilde{\Theta}_{ek}^{\mathrm{T}}\vartheta(e_k) + \varepsilon_{e\mathrm{d}}(x_{\mathrm{d}k}),\end{aligned} \tag{45}$$

where $\varepsilon_{e\mathrm{d}} = \varepsilon_{\mathrm{d}}(x_{\mathrm{d}k}) - g(x_k)\varepsilon_{ue}$.

Next, the convergence of tracking error, online identifier, and the cost function, feedback control signal, and feedforward control signal OLA parameter estimation errors is demonstrated in the following theorem while explicitly considering the OLA reconstruction errors.

Theorem 2 (System stability) Let u_{e0k} be any initial admissible control for the nonlinear system (23) such that (23) is initially asymptotically stable. Let the parameter tuning for the cost OLA and feedforward OLA be provided by (33) and (35), respectively, and let the tuning law for the feedforward estimator be given by (39). Then, there exist positive constants K_{d}, α_{Je}, α_{ue}, and α_{d} such that the tracking error, identification error, and the OLA parameter estimation errors of the cost function, feedback and feedforward terms are all UUB for all $k \geqslant k_0 + T$, and the estimated control input approaches the optimal control input such that $\|u(x_k, x_{\mathrm{d}k}, e_k) - u^*(x_k, x_{\mathrm{d}k}, e_k)\| \leqslant \varepsilon_u$ for a small positive constant ε_u.

Proof of Theorem 2 is shown in Appendix.

5 Simulation results

To demonstrate the effectiveness of the nearly optimal nonlinear controller, a tracking example is presented under two scenarios: when the internal dynamics are known and when they are not. When the internal dynamics are known, only the cost and feedback OLAs are required to regulate the tracking error to zero. Subsequently, the feedforward OLA is added when the internal dynamics are not known. For example, a differentially driven nonholonomic mobile robot is considered whose discretized nonlinear system is described by [15]

$$v_{k+1} = \begin{bmatrix} v_{\mathrm{R}k+1} \\ v_{\mathrm{L}k+1} \end{bmatrix} = \begin{bmatrix} v_{\mathrm{R}k} \\ v_{\mathrm{L}k} \end{bmatrix} + Tf(v_k) + TM^{-1}\tau, \tag{46}$$

where $f(v_k) = -M^{-1}(V(v_k)v_k + F(v_k))$, τ is the control torque, $v_{\mathrm{R}k}$ and $v_{\mathrm{L}k}$ are the velocities of the right and

left wheels of the robot, respectively, and T is the sampling time. Furthermore, M is the inertial matrix, $V(v_k)$ is the nonlinear Coriolis forces matrix, and $F(v_k)$ is the nonlinear friction vector. The robot parameters used in the simulation were chosen to be as those presented in [14] while the sampling time is taken as $T = 0.01$ s. The objective of the mobile robot is to track a virtual reference cart, and the desired right and left wheel velocities, $v_{k\mathrm{d}} = [v_{\mathrm{dR}k} \;\; v_{\mathrm{dL}k}]^{\mathrm{T}}$, which are generated online [14]. For this test, the translational and angular velocities of the reference cart were taken as $v_{rk} = 1$ m/s and $\omega_{rk} = 0.5\sin(0.2\pi kT)$ rad/s, respectively.

To implement the control scheme, two-layer NNs are considered consisting of one layer of randomly assigned constant weights, v_{N}, in the first layer and one layer of tunable weights, Θ_{N}, in the second layer. It has been shown that by randomly selecting the input layer weights v_{N}, the activation function forms a stochastic basis, and thus the approximation property holds for all inputs in a compact set [1]. Additionally, 10 hidden layer neurons were selected for both the cost function and feedback control OLAs while 25 hidden layer neurons were selected to estimate the feedforward signal of the control input. The activation function of the cost function OLA was selected as hyperbolic tangent squared to obtain an even linearly independent basis function. Conversely, the basis function of the feedback action OLA was selected as hyperbolic tangent after taking the partial derivative. Finally, radial basis functions were selected as activation functions for the feedforward control estimator.

The initial stabilizing control law was selected as

$$u_{e0}(e_k) = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} v_{\mathrm{R}k} - v_{\mathrm{dR}k} \\ v_{\mathrm{L}k} - v_{\mathrm{dL}k} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} e_{1k} \\ e_{2k} \end{bmatrix}. \tag{47}$$

The control gains for the OLA-based optimal control scheme were selected as $K_{\mathrm{d}} = 0.1$, $\alpha_{Je} = 0.1$, $\alpha_{ue} = 0.1$ and $\alpha_{\mathrm{d}} = 0.1$, and all tunable NN weights were initialized to zero. The simulation was run for 10 s (1000 time steps), and for the first 6 s, probing noise with mean zero and variance 1.4 was added to the system to ensure the persistency of excitation condition holds (see Section 3).

The resulting robot trajectory when the dynamics are known is shown in Fig. 1, where it is observed that the robot converges to the path of the virtual cart and maintains the desired course for the remainder of the test. The time histories of both the cost function and the feedback control signal parameter estimates are shown in Fig. 2. Examining the figure, it is clear that the parameter estimates converge to constant values and remain bounded consistent with Theorem 1. The cost function and feedback action network errors are shown in Fig. 3 where it is clear that both errors initially incur large values but then converge to a small bounded value near the origin.

Next, knowledge of the internal dynamics is removed, and the feedforward estimator presented in Section 4 is added to the robot control law. The robot trajectories for this case were observed to be similar to the trajectory in Fig. 1 while the final values of the cost and action OLA parameters as well as the values of the cost and action errors were observed to be similar to those presented in Figs. 2 and 3, respectively. In addition, the difference between the actual feedforward control term and the estimated feedforward term is shown in Fig. 4. Here, the estimation error is found to be small and bounded consistent with the theoretical results of Theorem 2.

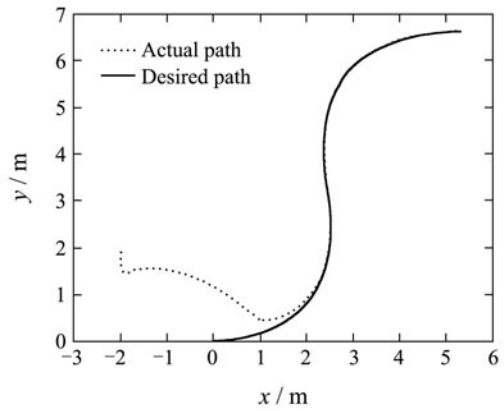

Fig. 1 Robot trajectory.

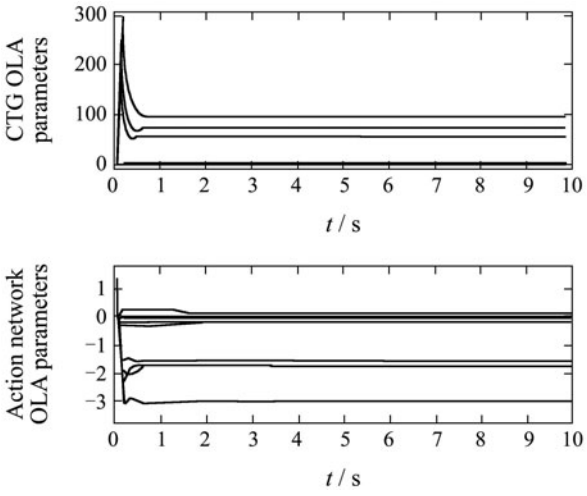

Fig. 2 OLA parameter estimates.

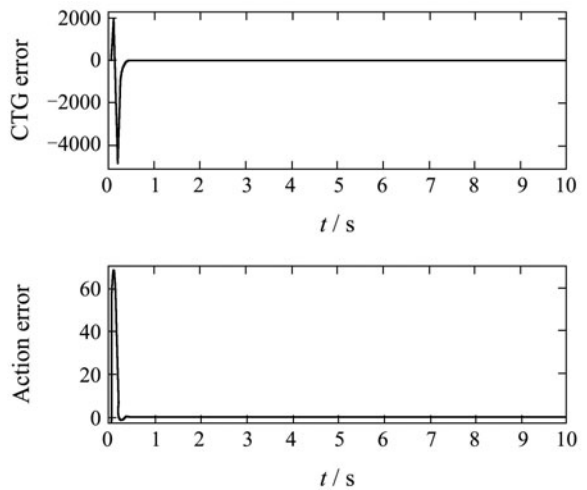

Fig. 3 OLA CTG and action errors.

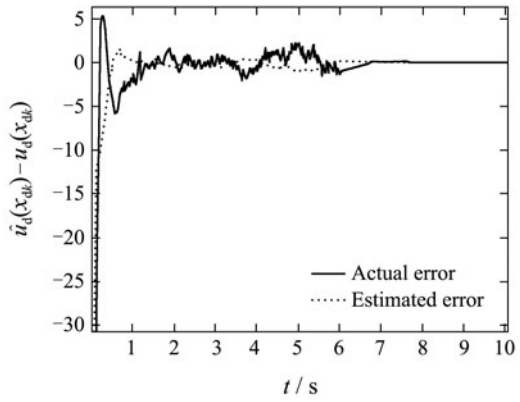

Fig. 4 OLA feedforward control term estimation error.

Closed-form solutions for the discrete-time HJB equation are difficult to obtain, and no bench marking methods currently exist for evaluating discrete-time nonlinear optimal control laws. As a result, a cost comparison with the initial stabilizing control policy is considered where the costs were calculated by summing the cost function (26) from $i = 0$ to $i = 1000$. Examining Table 1, the OLA-based optimal tracking control scheme is observed to incur less cumulative cost than the initial stabilizing control both when the internal dynamics were known and when they were unknown. Thus, the proposed tracker is indeed tuning to decrease the cost function (26) as predicted. Furthermore, the difference in the optimal costs when the dynamics are known and when they are not are observed to be small, which again reinforces the theoretical results of Theorem 2.

Table 1 Comparisons of total cost.

Control policy	Cost with $f_{\rm d}$ known	Cost with $f_{\rm d}$ unknown
u_{e0}	3.104e5	3.320e5
$\hat{u}_e^*$	2.562e5	2.565e5

6 Conclusions

In this work, the HJB equation was solved online for the nearly optimal control of general affine nonlinear DT systems using OLAs to address the regulation and tracking control problems. Knowledge of the system's internal dynamics was not needed while the OLAs generate a novel nearly optimal control law, and an initial admissible control policy guarantees that the system is stable while the OLAs learn. For the tracking problem, a desired feedforward portion of the control input generated by the third OLA rendered an overall stable system with ultimately bounded signals. All OLA parameters were tuned online using novel update laws, and system stability is guarenteed using Lyapunov theory.

References

[1] S. Jagannathan. *Neural Network Control of Nonlinear Discrete-time Systems*. Boca Raton: CRC Press, 2006.

[2] F. L. Lewis, V. L. Syrmos. *Optimal Control*. 2nd ed. New York: John Wiley & Sons, 1995.

[3] J. Shamma, J. Cloutier. Existence of SDRE stabilizing feedback. *IEEE Transactions on Automatic Control*, 2003, 48(3): 513 – 517.

[4] J. Vlassenbroeck, R. Van Dooren. A Chebyshev technique for solving nonlinear optimal control problems. *IEEE Transactions on Automatic Control*, 1988, 33(4): 333 – 340.

[5] Z. Chen, S. Jagannathan. Generalized Hamilton-Jacobi-Bellman formulation based neural network control of affine nonlinear discrete-time systems. *IEEE Transactions on Neural Networks*, 2008, 19(1): 90 – 106.

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

[7] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, et al. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 2009, 45(2): 477 – 484.

[8] G. Toussaint, T. Basar, F. Bullo. H-infinity-optimal tracking control techniques for nonlinear underactuated systems. *Proceedings of IEEE Conference on Decision and Control*, New York: IEEE, 2000: 2078 – 2083.

[9] D. Gu, H. Hu. Receding horizon tracking control of wheeled mobile robots. *IEEE Transactions on Control Systems Technology*, 2006, 14(4): 743 – 49.

[10] H. Zhang, Q. Wei, Y. Luo. A novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy HDP iteration algorithm. *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, 2008, 38(4): 937 – 942.

[11] T. Dierks, B. T. Thumati, S. Jagannathan. Optimal control of unknown affine nonlinear discrete-time systems using offline-trained neural networks with proof of convergence. *Neural Networks*, 2009, 22(5/6): 851 – 860.

[12] K. G. Vamvoudakis, F. L. Lewis. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 2010, 46(5): 878 – 888.

[13] H. K. Khalil. *Nonlinear Systems*, 3rd ed. New York: Prentice Hall, 2002.

[14] D. Wang, J. Huang. Neural network-based adaptive dynamic surface control for a class of uncertain nonlinear systems in strict-feedback form. *IEEE Transactions on Neural Networks*, 2005, 16(1): 195 – 202.

[15] M. K. Bugeja, S. G. Fabri, L. Camilleri. Dual adaptive dynamic control of mobile robots using nearal networks. *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, 2009, 39(1): 129 – 141.

Appendix

Proof of Theorem 1 Consider the positive definite Lyapunov candidate

$$V = \frac{\alpha_u^2 \alpha_J \bar{X}_{\rm m}}{2g_{\rm M}^2 \Omega_{\rm A}(\vartheta_{\rm M}^2+1)} V_{\rm D}(x_k) + \frac{\alpha_J \bar{X}_{\rm m}}{\Pi_{\rm A}} V_u(\tilde{\Theta}_k) + V_J(\tilde{\Phi}_k), \tag{a1}$$

where

$$V_{\rm D}(x_k) = x_k^{\rm T} x_k,\ \ V_u(\tilde{\Theta}_k) = {\rm tr}\{\tilde{\Theta}_k^{\rm T}\tilde{\Theta}_k\},\ \ V_J(\tilde{\Phi}_k) = \Phi_k^{\rm T}\Phi_k,$$
$$\Pi_{\rm A} = (\alpha_u+1)(\sigma'_{\rm M} g_{\rm M} \lambda_{\max}(R^{-1}))^2 + \alpha_u \frac{\left((\sigma'_{\rm M} g_{\rm M}\lambda_{\max}(R^{-1}))^2+2\right)}{4},$$

and $\lambda_{\max}(R^{-1})$ is the maximum singular value of R. The first difference of (a1) is given by

$$\Delta V = \frac{\alpha_u^2 \alpha_J \bar{X}_{\rm m} \Delta V_{\rm D}(x_k)}{2g_{\rm M}^2 \Pi_{\rm A}(\vartheta_{\rm M}^2+1)} + \frac{\alpha_J \bar{X}_{\rm m} \Delta V_u(\tilde{\Theta}_k)}{\Pi_{\rm A}} + \Delta V_J(\tilde{\Phi}_k). \tag{a2}$$

Considering first $\Delta V_{\rm D}(x_k) = x_{k+1}^{\rm T} x_{k+1} - x_k^{\rm T} x_x$, substituting the nonlinear dynamics (19), and applying the Cauchy-Schwartz inequality $((a_1+a_2+\ldots+a_n)^2 \leqslant n(a_1^2+a_2^2+\ldots+a_n^2))$ twice, reveals

$$\Delta V_{\rm D}(x_k) \leqslant 2\|f(x_k)+g(x_k)u^*(x_k)\|^2 - x_k^{\rm T} x_k + 4\|g(x_k)\tilde{\Theta}^{\rm T}\vartheta(x_k)\|^2 + 4\|g(x_k)\varepsilon_{uk}\|^2. \tag{a3}$$

It is observed that the optimal closed loop system, $f(x_k) + g(x_k)u^*(x_k)$, is asymptotically stable on a compact set. Furthermore, it can be shown that on a compact set, the optimal closed

loop system is upper bounded according to

$$\|f(x_k)+g(x_k)u^*(x_k)\|^2 \leqslant k^* \|x_k\|^2,$$

where k^* is a constant. Using this observation while defining $\Xi_{\mathrm{A}k}=\tilde{\Theta}_k^{\mathrm{T}}\vartheta(x_k)$, $\Delta V_{\mathrm{D}}(x_k)$ is upper bounded according to

$$\Delta V_{\mathrm{D}}(x_k) \leqslant -(1-2k^*)\|x_k\|^2 + 4g_{\mathrm{M}}^2\|\Xi_{\mathrm{A}k}\|^2 + 2g_{\mathrm{M}}^2\varepsilon_{u\mathrm{M}}^2. \tag{a4}$$

Moving on, $\Delta V_u(\tilde{\Theta}_k)=\mathrm{tr}\{\tilde{\Theta}_{k+1}^{\mathrm{T}}\tilde{\Theta}_{k+1}\}-\mathrm{tr}\{\tilde{\Theta}_k^{\mathrm{T}}\tilde{\Theta}_k\}$, and substituting (18) and then (17), $\Delta V_u(\tilde{\Theta}_k)$ is upper bounded in terms of $\Xi_{\mathrm{A}k}$ as

$$\begin{aligned}
&\Delta V_u(\tilde{\Theta}_k)\\
&\leqslant -\frac{\alpha_u(2-\alpha_u)}{(\vartheta^{\mathrm{T}}(x_k)\vartheta(x_k)+1)}\|\Xi_{\mathrm{A}k}\|^2\\
&\quad+\frac{\alpha_u(\alpha_u+1)\left(\|\tilde{\Phi}_k\|\sigma'_{\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R^{-1})+2\tilde{\varepsilon}_{u\mathrm{M}}\right)}{(\vartheta^{\mathrm{T}}(x_k)\vartheta(x_k)+1)}\|\Xi_{\mathrm{A}k}\|\\
&\quad+\frac{\alpha_u^2}{(\vartheta^{\mathrm{T}}(x_k)\vartheta(x_k)+1)}\cdot\Big(\frac{(\sigma'_{\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R^{-1}))^2\|\tilde{\Phi}_k\|^2}{4}\\
&\quad+\sigma'_{\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R^{-1})\|\tilde{\Phi}_k\|\tilde{\varepsilon}_{u\mathrm{M}}\Big)+\alpha_u^2\tilde{\varepsilon}_{u\mathrm{M}}^2.
\end{aligned}$$

Now, observing

$$\begin{aligned}
-\alpha_u(2-\alpha_u)\|\Xi_{\mathrm{A}k}\|^2 = &-\frac{\alpha_u(3-3\alpha_u)\|\Xi_{\mathrm{A}k}\|^2}{2}\\
&-\frac{\alpha_u(\alpha_u+1)\|\Xi_{\mathrm{A}k}\|^2}{2}
\end{aligned}$$

and completing the square with respect to

$$\|\Xi_{\mathrm{A}k}\|(|\tilde{\Phi}_k\|\sigma'_{\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R^{-1})+2\tilde{\varepsilon}_{u\mathrm{M}}),$$

and then $\|\tilde{\Phi}_k\|\sigma'_{\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R^{-1})\tilde{\varepsilon}_{u\mathrm{M}}$ reveals the upper bound of $\Delta V_u(\tilde{\Theta}_k)$ to be written as

$$\begin{aligned}
\Delta V_u(\tilde{\Theta}_k) \leqslant &-\frac{\alpha_u}{2(\vartheta_{\mathrm{M}}^2+1)}(3-3\alpha_u)\|\Xi_{\mathrm{A}k}\|^2+\alpha_u\Pi_{\mathrm{A}}\|\tilde{\Phi}_k\|^2\\
&+\Big(\alpha_u(5\alpha_u+4)+\frac{\alpha_u^2(\sigma'_{\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R^{-1}))^2}{2}\Big)\tilde{\varepsilon}_{u\mathrm{M}}^2.
\end{aligned} \tag{a5}$$

Next, considering

$$\Delta V_J(\tilde{\Phi}_k)=\Phi_{k+1}^{\mathrm{T}}\Phi_{k+1}-\Phi_k^{\mathrm{T}}\Phi_k,$$

and substituting (12) and (13) reveals

$$\begin{aligned}
\Delta V_J(\tilde{\Phi}_k) = &-2\alpha_J\frac{\tilde{\Phi}_k^{\mathrm{T}}X_kX_k^{\mathrm{T}}\tilde{\Phi}_k}{\|X_kX_k^{\mathrm{T}}+I\|_{\mathrm{F}}}-\alpha_J\frac{\tilde{\Phi}_k^{\mathrm{T}}X_k\Psi_k^{\mathrm{T}}}{\|X_kX_k^{\mathrm{T}}+I\|_{\mathrm{F}}}\\
&-\alpha_J\frac{\Psi_kX_k^{\mathrm{T}}\tilde{\Phi}_k}{\|X_kX_k^{\mathrm{T}}+I\|_{\mathrm{F}}}+\alpha_J^2\frac{\tilde{\Phi}_k^{\mathrm{T}}X_kX_k^{\mathrm{T}}X_kX_k^{\mathrm{T}}\tilde{\Phi}_k}{\|X_kX_k^{\mathrm{T}}+I\|_{\mathrm{F}}^2}\\
&+\alpha_J^2\frac{\tilde{\Phi}_k^{\mathrm{T}}X_kX_k^{\mathrm{T}}X_k\Psi_k^{\mathrm{T}}}{\|X_kX_k^{\mathrm{T}}+I\|_{\mathrm{F}}^2}+\alpha_J^2\frac{\Psi_kX_k^{\mathrm{T}}X_kX_k^{\mathrm{T}}\tilde{\Phi}_k}{\|X_kX_k^{\mathrm{T}}+I\|_{\mathrm{F}}^2}\\
&+\alpha_J^2\frac{\Psi_kX_k^{\mathrm{T}}X_k\Psi_k^{\mathrm{T}}}{\|X_kX_k^{\mathrm{T}}+I\|_{\mathrm{F}}^2}.
\end{aligned}$$

Observing

$$0<\bar{X}_{\mathrm{m}}\leqslant\frac{\|X_k^{\mathrm{T}}X_k\|}{\|X_kX_k^{\mathrm{T}}+I\|_{\mathrm{F}}}<1,$$

where $\bar{X}_{\mathrm{m}}$ is a constant ensured to exist by the PE condition described in Section 3, $\Delta V_J(\tilde{\Phi}_k)$ can be upper bounded as

$$\Delta V_J(\tilde{\Phi}_k)\leqslant-\alpha_J\bar{X}_{\mathrm{m}}\|\tilde{\Phi}_k\|^2+\alpha_J^2\Psi_{\mathrm{M}}^2. \tag{a6}$$

Then, substituting (a4)–(a6) into (a2) yields the upper bound for ΔV to be

$$\begin{aligned}
\Delta V \leqslant &\frac{-(1-2k^*)\alpha_u^2\alpha_J\bar{X}_{\mathrm{m}}}{2g_{\mathrm{M}}^2\Pi_{\mathrm{A}}(\vartheta_{\mathrm{M}}^2+1)}\|x_k\|^2-\alpha_J\bar{X}_{\mathrm{m}}(1-\alpha_u)\|\tilde{\Phi}_k\|^2\\
&-\frac{\alpha_u\alpha_J\bar{X}_{\mathrm{m}}(3-7\alpha_u)}{2(\vartheta_{\mathrm{M}}^2+1)\Pi_{\mathrm{A}}}\|\Xi_{\mathrm{A}k}\|^2+\varepsilon_{\mathrm{SM}},
\end{aligned} \tag{a7}$$

where

$$\begin{aligned}
\varepsilon_{\mathrm{SM}}=&\alpha_J^2\Psi_{\mathrm{M}}^2+\frac{\varepsilon_{u\mathrm{M}}^2\alpha_u^2\alpha_J\bar{X}_{\mathrm{m}}}{(\Pi_{\mathrm{A}}(\vartheta_{\mathrm{M}}^2+1))}\\
&+\alpha_J\bar{X}_{\mathrm{m}}\Big(\alpha_u(5\alpha_u+4)+\frac{\alpha_u^2(\sigma'_{\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R^{-1}))^2}{2}\Big)\frac{\tilde{\varepsilon}_{u\mathrm{M}}^2}{\Pi_{\mathrm{A}}}.
\end{aligned}$$

Therefore, ΔV is less than zero provided the following inequalities hold

$$\begin{cases}
\|x_k\|>\sqrt{\dfrac{2g_{\mathrm{M}}^2\Pi_{\mathrm{A}}(\vartheta_{\mathrm{M}}^2+1)\varepsilon_{\mathrm{SM}}}{(1-2k^*)\alpha_u^2\alpha_J\bar{X}_{\mathrm{m}}}} \ \text{or}\\
\|\Xi_{\mathrm{A}k}\|>\sqrt{\dfrac{\varepsilon_{\mathrm{SM}}2(\vartheta_{\mathrm{M}}^2+1)\Pi_{\mathrm{A}}}{\alpha_u\alpha_J\bar{X}_{\mathrm{m}}(3-7\alpha_u)}}\equiv b_{\Xi A} \ \text{or}\\
\|\tilde{\Phi}_k\|>\sqrt{\dfrac{\varepsilon_{\mathrm{SM}}}{\alpha_J\bar{X}_{\mathrm{m}}(1-\alpha_u)}},
\end{cases} \tag{a8}$$

and the tuning gains are selected as $\alpha_u<\dfrac{3}{7}$ and $0<\alpha_J<1$ for the class of nonlinear systems that satisfy the optimal closed loop bounds described above with $0<k^*<\dfrac{1}{2}$. Thus, using standard Lyapunov extensions [1], the system states and the cost and control NN weight estimation errors are UUB, and the system states are guaranteed to never leave their initial compact set.

To show $\|\hat{u}(x_k)-u^*(x_k)\|\leqslant\varepsilon_{\mathrm{r}}$, use (5) and (14) to observe $\hat{u}(x_k)-u^*(x_k)=-\tilde{\Theta}_k^{\mathrm{T}}\vartheta(x_k)-\varepsilon_u$. Then, taking the limit as $k\to\infty$ and taking the upper bound of $\hat{u}(x_k)-u^*(x_k)$ shows $\|\hat{u}(x_k)-u^*(x_k)\|\leqslant\|\Xi_{\mathrm{A}k}\|+\varepsilon_{u\mathrm{M}}\leqslant b_{\Xi A}+\varepsilon_{u\mathrm{M}}\equiv\varepsilon_{\mathrm{r}}$, where $b_{\Xi A}$ is defined in (a8).

Proof of Theorem 2 Consider the positive definite Lyapunov candidate

$$\begin{aligned}
V_{\mathrm{T}}=&\frac{\alpha_{ue}\alpha_{Je}\alpha_{\mathrm{d}}\bar{X}_{\mathrm{em}}}{12g_{\mathrm{M}}^2\Pi_{\mathrm{A}e}(\vartheta_{e\mathrm{M}}^2+1)}V_e(e_k)+\alpha_{\mathrm{d}}V_{Je}(\tilde{\Phi}_{ek})\\
&+\frac{\alpha_{Je}\alpha_{\mathrm{d}}\bar{X}_{\mathrm{em}}}{\Pi_{\mathrm{A}e}}V_{eu}(\tilde{\Theta}_{ek})\\
&+\frac{\alpha_{Je}\alpha_{ue}\bar{X}_{\mathrm{em}}}{2g_{\mathrm{M}}^2\Pi_{\mathrm{A}e}(\vartheta_{e\mathrm{M}}^2+1)}V_{\mathrm{d}}(\tilde{\Omega}_{\mathrm{d}k},\tilde{x}_k),
\end{aligned} \tag{a9}$$

where

$$\begin{aligned}
&V_e(e_k)=e_k^{\mathrm{T}}e_k,\ \ V_{Je}(\tilde{\Phi}_{ek})=\tilde{\Phi}_{ek}^{\mathrm{T}}\tilde{\Phi}_{ek},\\
&V_{ue}(\tilde{\Theta}_{ek})=\mathrm{tr}\{\tilde{\Theta}_{ek}^{\mathrm{T}}\tilde{\Theta}_{ek}\},\\
&V_{\mathrm{d}}(\tilde{\Omega}_{\mathrm{d}k},\tilde{x}_h)=\mathrm{tr}\{\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\tilde{\Omega}_{\mathrm{d}k}\}+\frac{\alpha_{\mathrm{d}}^2\tilde{x}_k^{\mathrm{T}}\tilde{x}_k}{6},
\end{aligned}$$

and

$$\begin{aligned}
\Pi_{\mathrm{A}e}=&(\alpha_{ue}+1)(\sigma'_{e\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R_e^{-1}))^2\\
&+\frac{\alpha_{ue}\left((\sigma'_{e\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R_e^{-1}))^2+2\right)}{4},
\end{aligned}$$

where $\lambda_{\max}(R_e^{-1})$ is the maximum eigenvalue of R_e^{-1}. The first difference of V_{T} is given by

$$\begin{aligned}
\Delta V_{\mathrm{T}}=&\frac{\alpha_{ue}\alpha_{Je}\alpha_{\mathrm{d}}\bar{X}_{\mathrm{em}}\Delta V_e(e_k)}{12g_{\mathrm{M}}^2\Pi_{\mathrm{A}e}(\vartheta_{e\mathrm{M}}^2+1)}+\alpha_{\mathrm{d}}\Delta V_{Je}(\tilde{\Phi}_{ek})\\
&+\frac{\alpha_{Je}\alpha_{\mathrm{d}}\bar{X}_{\mathrm{em}}}{\Pi_{\mathrm{A}e}}\Delta V_{eu}(\tilde{\Theta}_{ek})\\
&+\frac{\alpha_{Je}\alpha_{ue}\bar{X}_{\mathrm{em}}\Delta V_{\mathrm{d}}(\tilde{\Omega}_{\mathrm{d}k},\tilde{x}_k)}{2g_{\mathrm{M}}^2\Pi_{\mathrm{A}e}(\vartheta_{e\mathrm{M}}^2+1)}.
\end{aligned} \tag{a10}$$

First, considering $\Delta V_e(e_k)$, substituting the error dynamics (45), and applying the Cauchy-Schwartz inequality twice yeilds

$$\begin{aligned}
\Delta V_e(e_k)\leqslant&\ 2\|f_e(e_k)+g(x_k)u_e(e_k)\|^2+6\|\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_{\mathrm{d}})\|^2\\
&+6g_{\mathrm{M}}^2\|\tilde{\Theta}_{ek}^{\mathrm{T}}\vartheta(e_k)\|^2+6\varepsilon_{\mathrm{dM}}^2-e_k^{\mathrm{T}}e_k.
\end{aligned} \tag{a11}$$

Observing the optimal error dynamics are asymptotically stable, it can be shown that $\|f_e(e_k)+g(x_k)u_e(e_k)\|^2\leqslant k_e^*\|e_k\|^2$ where k_e^* is a constant. Applying this relation and taking the upper

bound of (a11) is written as

$$\Delta V_e(e_k) \leqslant -(1-2k_e^*)\|e_k\|^2 + 6\|\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_{\mathrm{d}})\|^2 + 6g_{\mathrm{M}}^2\|\tilde{\Theta}_{ek}^{\mathrm{T}}\vartheta(e_k)\|^2 + 6\varepsilon_{\mathrm{dM}}^2. \quad \text{(a12)}$$

Next, examining the similarities of $V_{Je}(\tilde{\Phi}_{ek})$ and $V_{eu}(\tilde{\Theta}_{ek})$ to $V_J(\tilde{\Phi}_k)$ and $V_u(\tilde{\Theta}_k)$, respectively, defined in Proof of Theorem 1, the first differences of $\Delta V_{Je}(\tilde{\Phi}_{ek})$ and $\Delta V_{eu}(\tilde{\Theta}_{ek})$ are calculated identically to (a5) and (a6), respectively, after substituting the appropriate variables defined for tracking in Section 4 in place of the regulation variables defined in Section 3.

Now, considering $\Delta V_{\mathrm{d}}(\tilde{\Omega}_{\mathrm{d}k}, \tilde{x}_k)$, using (38) and (40) and taking the upper bound gives

$$\begin{aligned}&\Delta V_{\mathrm{d}}(\tilde{\Omega}_{\mathrm{d}k}, \tilde{x}_k)\\ &\leqslant -\alpha_{\mathrm{d}}(2-\alpha_{\mathrm{d}})\|\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k)\|^2\\ &\quad + \frac{\alpha_{\mathrm{d}}^2\|\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k)\|^2}{2} + 2\alpha_{\mathrm{d}}(1+\alpha_{\mathrm{d}})\|\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k)\|\varepsilon_{\mathrm{dM}}\\ &\quad + \varepsilon_{\mathrm{dM}}^2 + \frac{\alpha_{\mathrm{d}}^2\varepsilon_{\mathrm{dM}}^2}{2} + \frac{\alpha_{\mathrm{d}}^2K_{\mathrm{d}}^2\|\tilde{x}_k\|^2}{2} - \frac{\alpha_{\mathrm{d}}^2\|\tilde{x}_k\|^2}{6}.\end{aligned}$$

Then, completing with respect to $\|\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k)\|\varepsilon_{\mathrm{dM}}$ yields

$$\begin{aligned}&\Delta V_{\mathrm{d}}(\tilde{\Omega}_{\mathrm{d}k}, \tilde{x}_k)\\ &\leqslant -\alpha_{\mathrm{d}}(\frac{3}{2} - 2\alpha_{\mathrm{d}})\|\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k)\|^2\\ &\quad -(\frac{\alpha_{\mathrm{d}}^2}{2})(\frac{1}{3} - K_{\mathrm{d}}^2)\|\tilde{x}_k\|^2 + (1 + 2\alpha_{\mathrm{d}}(\alpha_{\mathrm{d}}+1) + \frac{\alpha_{\mathrm{d}}^2}{2})\varepsilon_{\mathrm{dM}}^2.\end{aligned} \quad \text{(a13)}$$

Next, using (a12), (a13), and $V_{Je}(\tilde{\Phi}_{ek})$ and $V_{eu}(\tilde{\Theta}_{ek})$ written in the form of (a5) and (a6), respectively, to form (a10) gives

$$\begin{aligned}\Delta V_{\mathrm{T}} = &-\frac{\alpha_{ue}\alpha_{Je}\alpha_{\mathrm{d}}\bar{X}_{\mathrm{em}}(1-2k_e^*)}{12g_{\mathrm{M}}^2\Pi_{\mathrm{Ae}}(\vartheta_{e\mathrm{M}}^2+1)}\|e_k\|^2\\ &-\alpha_{\mathrm{d}}\alpha_{Je}\bar{X}_{\mathrm{em}}(1-\alpha_{ue})\|\tilde{\Phi}_{ek}\|^2\\ &-\frac{\alpha_{Je}\alpha_{ue}\alpha_{\mathrm{d}}^2\bar{X}_{\mathrm{em}}(\frac{1}{3}-K_{\mathrm{d}}^2)}{4g_{\mathrm{M}}^2\Pi_{\mathrm{Ae}}(\vartheta_{e\mathrm{M}}^2+1)}\|\tilde{x}_k\|^2\\ &-\frac{\alpha_{Je}\alpha_{ue}\alpha_{\mathrm{d}}\bar{X}_{\mathrm{em}}}{2g_{\mathrm{M}}^2\Pi_{\mathrm{Ae}}(\vartheta_{e\mathrm{M}}^2+1)}(\frac{1}{2}-2\alpha_{\mathrm{d}})\|\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k)\|^2\\ &-\frac{\alpha_{ue}\alpha_{Je}\alpha_{\mathrm{d}}\bar{X}_{\mathrm{em}}}{2(\vartheta_{e\mathrm{M}}^2+1)\Pi_{\mathrm{Ae}}}(2-3\alpha_{eu})\|\tilde{\Theta}_{ek}^{\mathrm{T}}\vartheta(e_k)\|^2 + \varepsilon_{\mathrm{Se}},\end{aligned}$$

where

$$\begin{aligned}\varepsilon_{\mathrm{Se}} \equiv &\frac{\varepsilon_{\mathrm{dM}}^2\alpha_{Je}\alpha_{ue}\bar{X}_{\mathrm{em}}}{2g_{\mathrm{M}}^2\Pi_{\mathrm{Ae}}(\vartheta_{e\mathrm{M}}^2+1)}(1+2\alpha_{\mathrm{d}}(\frac{5}{4}\alpha_{\mathrm{d}}+\frac{3}{2}))\\ &+\alpha_{\mathrm{d}}\alpha_{Je}^2\Psi_{e\mathrm{M}}^2 + \frac{\alpha_{Je}\alpha_{\mathrm{d}}\bar{X}_{\mathrm{em}}}{\Pi_{\mathrm{Ae}}}(\alpha_{ue}(5\alpha_{ue}+4)\\ &+\frac{\alpha_{ue}^2(\sigma'_{e\mathrm{M}}g_{\mathrm{M}}\lambda_{\max}(R_e^{-1}))^2}{2})\tilde{\varepsilon}_{ue\mathrm{M}}^2,\end{aligned}$$

and $\Delta V_{\mathrm{T}} < 0$ for the class of nonlinear systems that satisfy the optimal closed loop bounds described above with $k_e^* < \frac{1}{2}$ provided the control gains are selected according to

$$K_{\mathrm{d}}^2 < \frac{1}{3},\ \alpha_{Je} < 1,\ \alpha_{eu} < \frac{2}{3},\ \alpha_{\mathrm{d}} < \frac{1}{4},$$

and the following inequalities hold:

$$\begin{aligned}&\|e_k\| > \sqrt{\frac{12g_{\mathrm{M}}^2\Pi_{\mathrm{Ae}}(\vartheta_{e\mathrm{M}}^2+1)\varepsilon_{\mathrm{Se}}}{\alpha_{ue}\alpha_{Je}\alpha_{\mathrm{d}}\bar{X}_{\mathrm{em}}(1-2k_e^*)}} \equiv b_e \ \text{ or}\\ &\|\tilde{x}_k\| > \sqrt{\frac{4g_{\mathrm{M}}^2\Pi_{\mathrm{Ae}}(\vartheta_{e\mathrm{M}}^2+1)\varepsilon_{\mathrm{Se}}}{\alpha_{Je}\alpha_{ue}\alpha_{\mathrm{d}}^2\bar{X}_{\mathrm{em}}(\frac{1}{3}-K_{\mathrm{d}}^2)}} \equiv b_{\tilde{x}} \ \text{ or}\\ &\|\tilde{\Theta}_{ek}^{\mathrm{T}}\vartheta(e_k)\| > \sqrt{\frac{2(\vartheta_{e\mathrm{M}}^2+1)\Pi_{\mathrm{Ae}}\varepsilon_{\mathrm{Se}}}{\alpha_{Je}\alpha_{\mathrm{d}}\alpha_{ue}\bar{X}_{\mathrm{em}}(2-3\alpha_{eu})}} \equiv b_{\Xi e} \ \text{ or}\\ &\|\tilde{\Phi}_{ek}\| > \sqrt{\frac{\varepsilon_{\mathrm{Se}}}{\alpha_{\mathrm{d}}\alpha_{Je}\bar{X}_{\mathrm{em}}(1-\alpha_{ue})}} \equiv b_{\Phi e} \ \text{ or}\\ &\|\tilde{\Omega}_{\mathrm{d}k}^{\mathrm{T}}\phi(x_k)\| > \sqrt{\frac{2g_{\mathrm{M}}^2\Pi_{\mathrm{Ae}}(\vartheta_{e\mathrm{M}}^2+1)\varepsilon_{\mathrm{Se}}}{\alpha_{\mathrm{d}}\alpha_{Je}\alpha_{ue}\bar{X}_{\mathrm{em}}(1/2-2\alpha_{\mathrm{d}})}} \equiv b_{\Omega\mathrm{d}}.\end{aligned}$$

Therefore, using standard Lyapunov extension [1], it can be concluded that ΔV_{T} is less than zero outside of a compact set so that the tracking error (22) and the OLA parameter estimation errors of the cost function, feedback control signal, and feedforward control inputs are all UUB. To show $\|u(x_k, x_{\mathrm{d}k}, e_k) - u^*(x_k, x_{\mathrm{d}k}, e_k)\| \leqslant \varepsilon_u$, we use (28), (30), (34), and (41)–(43) to observe

$$\begin{aligned}&\|u(x_k, x_{\mathrm{d}k}, e_k) - u^*(x_k, x_{\mathrm{d}k}, e_k)\|\\ &= g_{\mathrm{M}}^I\|\tilde{\Omega}_{\mathrm{kd}}^{\mathrm{T}}\phi(x_{\mathrm{d}k})\|_{\mathrm{F}} + g_{\mathrm{M}}^I\varepsilon_{\mathrm{dM}} + \|\tilde{\Theta}_{ek}^{\mathrm{T}}\vartheta_e(e_k)\| + \varepsilon_{ue\mathrm{M}}\\ &\leqslant g_{\mathrm{M}}^I b_{\Omega\mathrm{d}} + g_{\mathrm{M}}^I\varepsilon_{\mathrm{dM}} + b_{\Xi e} + \varepsilon_{ue\mathrm{M}}\\ &\equiv \varepsilon_u.\end{aligned}$$

Travis DIERKS received his B.S. and M.S. degrees in Electrical Engineering and Ph.D. from the Missouri University of Science and Technology (formerly the University of Missouri-Rolla), Rolla, in 2005, 2007, and 2009, respectively. While completing his doctoral degree, he was a GAANN fellow and a Chancellor's fellow. His research interests include nonlinear control, optimal control, neural network control, and the control and coordination of autonomous ground and aerial vehicles. He is currently with DRS Sustainment Systems, Inc., St. Louis, MO. E-mail: tdierks@drs-ssi.com.

Sarangapani JAGANNATHAN received his Ph.D. degree in Electrical Engineering from the University of Texas, Arlington in 1994. He worked at Systems and Controls Research Division, in Caterpillar Inc., Peoria as a consultant during 1994 to 1998, and during 1998 to 2001 he was at the University of Texas at San Antonio as an assistant professor, and since September 2001, he is at the Missouri University of Science and Technology (former University of Missouri-Rolla) where he is currently a Rutledge-Emerson Distinguished Professor and Site Director for the NSF Industry/University Cooperative Research Center on Intelligent Maintenance Systems. He has coauthored over 87 peer reviewed journal articles, 175 refereed IEEE conference articles, several book chapters and three books entitled 'Neural Network Control of Robot Manipulators and Nonlinear Systems', published by Taylor & Francis, London in 1999, 'Discrete-time Neural Network Control of Nonlinear Discrete-time Systems', CRC Press, April 2006, and 'Wireless Ad Hoc and Sensor Networks: Performance, Protocols and Control', CRC Press, April 2007, and holds 18 patents. His research interests include adaptive and neural network control, computer/communication/sensor networks, prognostics, and autonomous systems/robotics.