Generalized Hamilton–Jacobi–Bellman Formulation -Based Neural Network Control of Affine Nonlinear Discrete-Time Systems

Zheng Chen, *Student Member, IEEE*, and Sarangapani Jagannathan, *Senior Member, IEEE*

***Abstract*—In this paper, we consider the use of nonlinear networks towards obtaining nearly optimal solutions to the control of nonlinear discrete-time (DT) systems. The method is based on least squares successive approximation solution of the generalized Hamilton–Jacobi–Bellman (GHJB) equation which appears in optimization problems. Successive approximation using the GHJB has not been applied for nonlinear DT systems. The proposed recursive method solves the GHJB equation in DT on a well-defined region of attraction. The definition of GHJB, pre-Hamiltonian function, HJB equation, and method of updating the control function for the affine nonlinear DT systems under small perturbation assumption are proposed. A neural network (NN) is used to approximate the GHJB solution. It is shown that the result is a closed-loop control based on an NN that has been tuned *a priori* in offline mode. Numerical examples show that, for the linear DT system, the updated control laws will converge to the optimal control, and for nonlinear DT systems, the updated control laws will converge to the suboptimal control.**

***Index Terms*—Generalized Hamilton–Jacobi–Bellman (BHJB) equation, neural network (NN), nonlinear discrete-time (DT) system.**

I. Introduction

In the literature, there are many methods of designing stable control of nonlinear systems. However, stability is only a bare minimum requirement in a system design. Ensuring optimality guarantees the stability of the nonlinear system; however, optimal control of nonlinear systems is a difficult and challenging area. If the system is modeled by linear dynamics and the cost function to be minimized is quadratic in the state and control, then the optimal control is a linear feedback of the states, where the gains are obtained by solving a standard Riccati equation [9]. On the other hand, if the system is modeled by the nonlinear dynamics or the cost function is nonquadratic, the optimal state feedback control will depend upon obtaining the solution to the Hamilton–Jacobi–Bellman (HJB) [10] equation which is generally nonlinear. The HJB equation is difficult to solve directly because it involves solving either nonlinear partial difference or differential equations.

To overcome the difficulty in solving the HJB equation, recursive methods are employed to obtain the solution of HJB equation indirectly. Kleinman [14] pointed out that the solution of the Riccati equation can be obtained by successively solving a sequence of Lyapunov equations, which is linear in the cost function of the system, and thus, it is easier to solve when compared to a Riccati equation, which is nonlinear in the cost function. Saridis [11] extended this idea to the case of nonlinear continuous-time systems where a recursive method is used to obtain the optimal control of continuous system by successively solving the generalized Hamilton–Jacobi–Bellman (GHJB) equation, and then, updating the control if an admissible initial control is given. There has been a great deal of effort to address this problem in the literature in continuous time. Approximate HJB solution has been confronted using many techniques by Saridis [11], Beard [19]–[21], Bernstein [1], Bertsekas and Tsitsiklis [2], Han and Balakrishnan [12], Lyshevski [15], Lewis [6], [7], and others.

Although the GHJB equation is linear and easier to solve than HJB equation, no general solution for GHJB is demonstrated. Galerkin's spectral approximation method is employed in [19] to find approximate but close solutions to the GHJB at each iteration step. Beard [20] employed a series of polynomial functions as basic functions to solve the approximate GHJB equation in continuous time but this method requires the computation of a large number of integrals. Park [4] employed interpolating wavelets as the basic functions. On the other hand, Lewis and Abu-Khalaf [8], based on the work of Lyshevski [15], employed nonquadratic performance functional to solve constrained control problems for general affine nonlinear continuous-time systems using neural networks (NNs). In addition, it was also shown how to formulate the associated Hamilton–Jacobi–Isaac (HJI) equation using special nonquadratic supply rates to obtain the nonlinear state feedback H_∞ control. Huang [25], [26] reduced the L_2 gain optimization and nonlinear H_∞ problems to solving a single algebraic Riccati equation (ARE) along with a sequence of linear algebraic equations in discrete time (DT). Here, the value function is expanded by Taylor series consisting of higher order terms into a series of polynomial functions and approximating them but this approach requires significant computations. Additionally, the ARE in DT is still nonlinear which is difficult to solve.

Since NN can effectively extend adaptive control techniques to nonlinearly parameterized systems, Miller [16] proposed

Manuscript received May 28, 2006; revised December 21, 2006; accepted April 17, 2007. This work was supported in part by the National Science Foundation under Grants ECCS #0327877 and ECCS#0621924.

Z. Chen was with the Department of Electrical and Computer Engineering, University of Missouri—Rolla, Rolla, MO 65409-0910 USA. He is now with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: chenzhe1@egr.msu.edu).

S. Jagannathan is with the Department of Electrical and Computer Engineering, University of Missouri—Rolla, Rolla, MO 6540-0910 USA (e-mail: sarangap@umr.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNN.2007.900227

using NN to obtain optimal control laws via the HJB equation. On the other hand, Parisini and Zoppoli [18] used NN to derive optimal control laws for DT stochastic nonlinear systems. Similarly, Lin and Brynes [24] presented H_∞ control of DT nonlinear systems. Although many papers, i.e., [6], [7], [11], [19], and [20], have discussed the GHJB method for continuous-time systems, currently there is very minimal work available on the GHJB method for DT nonlinear systems. DT version of the approximate GHJB-equation-based control is important since all the controllers are typically implemented using embedded digital hardware. Ferrari and Stengel [27] solved DT HJB problem through adaptive critic designs (ACD). The cost function and control is updated through heuristic dynamic programming (HDP), dual heuristic dynamic programming (DHP), global dual heuristic dynamic programming (GDHP), and action-dependent (AD) designs. Recent work on solving HJB for continuous time has appeared in the edited book, where [27] was published.

In this paper, we will apply the idea of GHJB equation in DT and set up the practical method for obtaining the near-optimal control of DT nonlinear systems by using Taylor series extension of the cost function. The higher terms (third order and higher) in the Taylor series expansion of the cost or value functional are ignored by using small signal perturbation assumption around the operating point while keeping a tradeoff between computation and accuracy. With an initial admission control, the cost function can be obtained by solving a so-called GHJB equation in DT. Subsequently, the updated control is obtained by minimizing the pre-Hamiltonian function. It is also demonstrated that the updated control will converge to the optimal control, which renders an approximate solution of the HJB equation in DT. The theory of GHJB in DT has also been applied to the linear DT case which indicates that the optimal control is nothing but the solution of the standard Riccati equation.

We use successive approximation techniques by employing NN in the least squares sense to solve the GHJB in DT and using the quadratic cost function. It is demonstrated that if the activation functions of the NN are linearly independent, the NN weight matrix has a unique solution. It is also shown that the result is a closed-loop control based on an NN that has been tuned *a priori* in offline mode. The theoretical results are verified through extensive rigorous simulation studies performed using linear and nonlinear DT systems and a two-link planar robot arm system. In the linear case, the updated control is shown to converge to the optimal control. In the nonlinear case, as expected, the updated control will converge to the suboptimal control.

It is also important to note that the proposed approach is different than a conventional DT linear quadratic regulator (DTLQR). DTLQR will not render the same solution as that of the one presented in this paper as we have considered several higher order terms in the Taylor series expansion making it nonlinear and yet it is an approximated and sufficiently accurate methodology. Additionally, similarities between dynamical programming (DP) and GHJB theory and the differences between GHJB theory in discrete and continuous time are also highlighted in this paper.

The remainder of this paper is organized as follows. Section II introduces the DT GHJB theory. The method of obtaining the optimal control is discussed and verification for linear DT case is given. The NN method to approximately solve the GHJB equation is described and the Galerkin's spectral approximation method is applied in Section III. The GHJB-based controller design is demonstrated on a linear and nonlinear DT system through simulation in Section IV. Additionally, we apply the GHJB method to obtaining the near-optimal control of a two-link planar robot arm system. Finally, concluding remarks and future works are provided in Section V.

II. Optimal Control and GHJB Equation for Nonlinear DT Systems

Consider an affine in the control nonlinear DT dynamic system of the form

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

where $x(k) \in \Omega \subset \Re^n$, $u : \Re^n \to \Re^m$, $f : \Re^n \to \Re^n$, and $g : \Re^n \to \Re^{n\times m}$. Assume that $f + gu$ is Lipschitz continuous on a set Ω in $\Re^n$ containing the origin, and that the system (1) is controllable in the sense that there exists a continuous control on Ω that asymptotically stabilizes the system. It is desired to find a control function $u : \Re^n \to \Re^m$, which minimizes the generalized quadratic cost function

$$J(x(0);u) = \sum_{k=0}^{\infty}\Big(x(k)^T Q x(k) + u(x(k))^T \mathrm{R}\,u(x(k))\Big) + \phi(x(\infty)) \tag{2}$$

where Q is a positive–definite matrix, R is a symmetric positive–definite matrix, and $\phi : \Re^n \to \Re$ is a final state punishment function which is positive definite.

A. Control Objective

The objective is to select the feedback control law of $u(x(k))$ in order to minimize the cost-functional value.

Remark 1: It is important to note that the control $u(x(k))$ must both stabilize the system on Ω and make the cost-functional value finite so that the control is admissible [21].

Definition 2.1 (Admissible Controls): Let Ω_u denote the set of admissible controls. A control function $u : \Re^n \to \Re^m$ is defined to be admissible with respect to the state penalty function $(x(k))^T Q(x(k))$ and control energy penalty function $(u(x(k)))^T R(u(x(k)))$ on Ω, denoted as $u \in \Omega_u$, if the following is true:

- u is continuous on Ω;
- $u(x)|_{x=0} = 0$;
- $u(x)$ stabilizes system (1) on Ω;
- $J(x(0);u) = \sum_{k=0}^{\infty}((x(k)^T Q x(k)) + u(x(k))^T \mathrm{R}u(x(k))) + \phi(x(\infty)) < \infty, \forall x(0) \in \Omega$.

Remark 2: The admissible control guarantees that the control converges but, in general, any converged control cannot guarantee that it is admissible. For example, consider the nonlinear DT system

$$x(k+1) = \frac{1}{\sqrt{k+1}} + x(k) + u. \tag{3}$$

A feedback control is given as $u = -x$ and the system solution will be

$$x(k) = \begin{cases} (1/\sqrt{k}), & \text{for } k \geq 1 \\ x(0), & \text{for } k = 0 \end{cases}.$$

As $k \to \infty$, $x(k) \to 0$. This system with this feedback control is considered stable. However, $\|u(x(k))\|^2 = \|x(k)\|^2$ and the sum $\sum_{k=0}^{\infty} \|u(k)\|^2 = x^2(0) + \sum_{k=1}^{\infty}(1/k)$ is infinite. We can conclude that this feedback control is stable but not admissible. Hence, we should restrict the systems that decay sufficiently fast.

Given an admissible control and the state of the system at every instant of time, the performance of this control is evaluated through a cost function. If the solution of the dynamic system $x(k+1) = f(x(k)) + g(x(k))u(x(k))$ is known and given the cost function, the overall cost is the sum of the cost value calculated at each time step k. However, for complex nonlinear DT systems, the closed-form solution $x(k)$ is difficult to determine and the solution can depend upon the initial conditions. Therefore, another suitable cost function, which is independent of the solution of the nonlinear dynamic system $x(k)$, is needed. In general, it is very difficult to select the cost function; however, Theorem 2.1 will prove that there exists a positive–definite function $V(x(k))$, denoted in this paper for simplicity as $V(x)$, referred to as the value function, whose initial value $V(x(0))$ is equal to the cost-functional value of J given an admissible control and the state of the system.

Theorem 2.1: Assume $u(x) \in \Omega_u$ is an admissible control law arbitrarily selected for the nonlinear DT system. If there exists a positive–definite, uniformly convex, and continuously differentiable value function $V(x)$ on Ω satisfying the following:

$$\begin{aligned} &\frac{1}{2}(f(x)+g(x)u(x)-x)^T \cdot \nabla^2 V(x) \cdot (f(x)+g(x)u(x)-x) \\ &\quad + \nabla V(x)^T (f(x)+g(x)u(x)-x) \\ &\quad + x^T Q x + u(x)^T \mathrm{R} u(x) = 0 \end{aligned} \tag{4}$$

$$V(x(\infty)) = \phi(x(\infty)) \tag{5}$$

where $\nabla V(x)$ and $\nabla^2 V(x)$ are the gradient vector and Hessian matrix of $V(x)$, then $V(x(j))$ is the value function of the system defined in (1) for all $j = 0, \ldots \infty$ when the feedback control $u(x)$ is applied and

$$V(x(0)) = J(x(0); u). \tag{6}$$

Proof: Assume that $V(x(k)) > 0$ exists and is continuously differentiable. Then

$$V(x(\infty)) - V(x(j)) = \sum_{k=j}^{\infty} \Delta V(x(k)) \tag{7}$$

where $\Delta V(x(k)) = V(x(k+1)) - V(x(k))$ is the first difference. Since $V(x(k))$ is a continuously differentiable function, expanding the function $V(x)$ using Taylor series about the operating point of $x(k)$ renders

$$\begin{aligned} V(x(k+1)) = &V(x(k)) + \nabla V_k^T (x(k+1) - x(k)) \\ &+ \frac{1}{2}(x(k+1) - x(k))^T \\ &\cdot \nabla^2 V_k (x(k+1) - x(k)) + \cdots \end{aligned} \tag{8}$$

where ∇V_k is the gradient vector defined as

$$\begin{aligned} \nabla V_k &= \left.\frac{\partial V(x)}{\partial x}\right|_{x=x(k)} \\ &= \left.\left[\frac{\partial}{\partial x_1}V(x), \frac{\partial}{\partial x_2}V(x), \ldots, \frac{\partial}{\partial x_n}V(x)\right]^T\right|_{x=x(k)} \end{aligned} \tag{9}$$

and $\nabla^2 V_k$ is the Hessian matrix defined as

$$\nabla^2 V_k = \left.\begin{bmatrix} \frac{\partial^2 V(x)}{\partial x_1^2} & \frac{\partial^2 V(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 V(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 V(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 V(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 V(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 V(x)}{\partial x_n \partial x_1} & \frac{\partial^2 V(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 V(x)}{\partial x_n^2} \end{bmatrix}\right|_{x=x(k)}. \tag{10}$$

By assuming small perturbation about the operating point $\Delta x(k) = x(k+1) - x(k)$, the first three terms of Taylor series expansion can be considered and we can ignore terms higher than second order to receive

$$\begin{aligned} \Delta V(x(k)) \approx &\nabla V_k^T (x(k+1) - x(k)) \\ &+ \frac{1}{2}(x(k+1) - x(k))^T \\ &\cdot \nabla^2 V_k (x(k+1) - x(k)). \end{aligned} \tag{11}$$

From (7) and (11), using system dynamics (1), we can get

$$\begin{aligned} V(x(\infty)) - V(x(j)) = \sum_{k=j}^{\infty} \Big[&\nabla V_k^T (f_k + g_k u_k - x(k)) \\ &+ \frac{1}{2}(f_k + g_k u_k - x(k))^T \\ &\cdot \nabla^2 V_k (f_k + g_k u_k - x(k))\Big] \end{aligned} \tag{12}$$

where $f_k = f(x(k))$, $g_k = g(x(k))$, and $u_k = u(x(k))$. For convenience, we denote

$$\Delta x(k) = f_k + g_k u_k - x(k). \tag{13}$$

Then, we rewrite (12) to get

$$\begin{aligned} &V(x(\infty)) - V(x(j)) \\ &\quad = \sum_{k=j}^{\infty} \left[\nabla V_k^T \Delta x(k) + \frac{1}{2}\Delta x(k)^T \nabla^2 V_k \Delta x(k)\right]. \end{aligned} \tag{14}$$

Similarly, we rewrite (2) as

$$J(x(j); u) = \sum_{k=j}^{\infty} \left(x(k)^T Q x(k) + u_k^T \mathrm{R} u_k\right) + \phi(x(\infty)). \tag{15}$$

We add (14) on both sides of (15) and rewrite (14) as

$$J(x(j); u) - V(x(j))$$

$$= \sum_{k=j}^{\infty} \left[\frac{1}{2} \Delta x(k)^T \nabla^2 V_k \Delta x(k) + \nabla V_k^T \Delta x(k) + (x(k))^T Q(x(k)) + u_k^T \mathrm{R} u_k \right] + [\phi(x(\infty)) - V(x(\infty))]. \quad (16)$$

Because $x(k) \in \Omega$, from (4) and (5), we also have

$$\frac{1}{2} \Delta x(k)^T \cdot \nabla^2 V_k \cdot \Delta x(k) + \nabla V_k^T \Delta x(k) + x(k)^T Q x(k) + u_k^T \mathrm{R} u_k(x) = 0 \quad (17)$$

$$V(x(\infty)) = \phi(x(\infty)). \quad (18)$$

Applying (17) and (18) into (16) renders

$$V(x(j)) = J(x(j); u), \qquad \text{for } j = 0, \ldots, \infty. \quad (19)$$

More specifically, for $j = 0$, $V(x(0)) = J(x(0); u_1)$.

Remark 3: An optimal control function $u^*(x)$ for a nonlinear DT system is the one that minimizes the value function $V(x(0))$.

Remark 4: If $V(x)$ is quadratic function of x, since $\nabla^m V(x)|_{m \geq 3} = 0$, then Theorem 2.1 can be applicable to nonlinear DT systems without making the small perturbation assumption.

Definition 2.2 (GHJB Equation for Nonlinear DT System):

$$\frac{1}{2} \Delta x^T \nabla^2 V(x) \Delta x + \nabla V(x)^T \Delta x + x^T Q x + u(x)^T \mathrm{R} u(x) = 0 \quad (20)$$

$$V(x)|_{x=0} = 0 \quad (21)$$

where $\Delta x = f(x) + g(x)u(x) - x$.

In this paper, the infinite-horizon optimal control problem for the nonlinear DT system (1) is attempted. The cost function of the infinite-horizon problem for the DT system is defined as

$$J(x(0); u) = \sum_{k=0}^{\infty} \left(x(k)^T Q x(k) + u_k^T \mathrm{R} u_k \right). \quad (22)$$

The GHJB (20) with the boundary condition (21) can be used as (4) and (5) for the infinite-horizon problems, because, as $N \to \infty$, $x(\infty) = 0$ and $V(x)|_{x=0} = V(x(\infty)) = \phi(x(\infty)) = 0$; so if an admissible control is specified, for any infinite-horizon problem, we can solve the GHJB equation to obtain the value function $V(x)$ which in turn can be used in the cost function J along with $V(x(0))$ to calculate the cost of the admissible control.

We already know how to evaluate the performance of the current admissible control, but this is not our final goal. Our objective is to improve the performance of the system over time by updating the control so that a near-optimal controller can be obtained. Besides deriving an updated control law, it is required that the updated control functions render admission control inputs to the nonlinear system while ensuring that the performance is enhanced over time. The updated control function is obtained by minimizing a pre-Hamiltonian function. In fact, Theorem 2.2 demonstrates that if the control function is updated by minimizing the pre-Hamiltonian function defined in (23), then the system performance can be enhanced over time while guaranteeing that the updated control function is admissible for the original nonlinear system (1). Next, the pre-Hamiltonian function for the DT system is introduced.

Definition 2.3 (Pre-Hamiltonian Function for the Nonlinear DT System): A suitable pre-Hamiltonian function for the nonlinear system (1) is defined as

$$H(x, V(x), u(x)) = \frac{1}{2} \Delta x^T \nabla^2 V(x) \Delta x + \nabla V(x)^T \Delta x + x^T Q x + u(x)^T \mathrm{R} u(x) \quad (23)$$

where $x \in \Omega$. It is important to note that the pre-Hamiltonian is a nonlinear function of the state and cost value function the control functions. If a control function and cost value function $u^{(i)} \in \Omega$ satisfy $V^{(i)}$, an updated control function $\mathrm{GHJB}\left(V^{(i)}, u^{(i)}\right) = 0$ can be obtained by differentiating the pre-Hamiltonian function (23) associated with the value function u^{i+1}. In other words, the updated control function can be obtained by solving

$$\frac{\partial H\left(x, V^{(i)}(x), u^{(i+1)}\right)}{\partial u^{(i+1)}} = 0 \quad (24)$$

so that

$$\left(g^T(x) \nabla^2 V^{(i)}(x) g(x) + 2R\right) u^{(i+1)}(x) = -g^T(k) \left(\nabla V^{(i)}(x) + \nabla^2 V^{(i)}(x)(f(x) - x)\right). \quad (25)$$

In Theorem 2.1, since the positive–definite function $V(x)$ is uniformly convex on Ω, $\nabla^2 V^{(i)}(x)$ is a positive–definite function on Ω and the matrix R is positive definite; so it can be concluded that $g^T(x) \nabla^2 V^{(i)}(x) g(x) + 2R$ is a positive–definite matrix on Ω. We can rewrite (25) as

$$u^{(i+1)}(x) = -\left[g(x)^T \nabla^2 V^{(i)}(x) g(x) + 2R\right]^{-1} \cdot g(x)^T \left(\nabla V^{(i)}(x) + \nabla^2 V^{(i)}(x)(f(x) - x)\right). \quad (26)$$

Theorem 2.2 demonstrates that the updated control function is not only an admissible control but also improved control for the nonlinear DT system described by (1).

Theorem 2.2 (Improved Control): If $u^{(i)}(x) \in \Omega_u$ and $x(0) \in \Omega$ and the positive–definite and convex function $V^{(i)}$ satisfies $\mathrm{GHJB}\left(V^{(i)}, u^{(i)}\right) = 0$ with the boundary condition $V^{(i)}(0) = 0$, then the updated control function derived in (26) by using the pre-Hamiltonian results in an admissible control for the system (1) on Ω. Moreover, if $V^{(i+1)}$ is the unique positive–definite function satisfying $\mathrm{GHJB}\left(V^{(i+1)}, u^{(i+1)}\right) = 0$, then

$$V^{(i+1)}(x(0)) \leq V^{(i)}(x(0)). \quad (27)$$

Proof of Admissibility: First, we should investigate the stability of the system with the control $u^{(i+1)}$. We take the difference of $V^{(i)}(x)$ along the system $\left(f, g, u^{(i+1)}\right)$ trajectories to obtain

$$\begin{aligned}\Delta V^{(i)}(x(k)) &= V^{(i)}(x(k+1)) - V^{(i)}(x(k)) \\ &\approx \nabla V_k^{(i)^T}\left(f_k + g_k u_k^{(i+1)} - x(k)\right) \\ &\quad + \frac{1}{2}\left(f_k + g_k u_k^{(i+1)} - x(k)\right)^T \\ &\quad \cdot \nabla^2 V_k^i \left(f_k + g_k u_k^{(i+1)} - x(k)\right). \quad (28)\end{aligned}$$

Rewriting the GHJB equation GHJB$(V^i, U^i) = 0$ for $x = x(k) \in \Omega$, we have

$$\begin{aligned}&\frac{1}{2}\left(f_k + g_k u_k^{(i)} - x(k)\right)^T \cdot \nabla^2 V_k^i \left(f_k + g_k u_k^{(i)} - x(k)\right) \\ &\quad + \nabla V_k^{(i)^T}\left(f_k + g_k u_k^{(i)} - x(k)\right) \\ &\quad + x(k)^T Q x(k) + u_k^{(i)^T} \mathrm{R} u_k^{(i)} = 0. \quad (29)\end{aligned}$$

Substituting (29) into (28), (28) can be rewritten as

$$\begin{aligned}&\Delta V^{(i)}(x(k)) \\ &= -\frac{1}{2} u_k^{(i)^T}\left(g_k^T \nabla^2 V_k^i g_k + 2R\right) u_k^{(i)} - x(k)^T Q x(k) \\ &\quad + \left(\nabla V_k^{(i)^T} + (f_k - x(k))^T \nabla^2 V_k^i\right) \\ &\quad \cdot g_k \left(u_k^{(i+1)} - u_k^{(i)}\right) + \frac{1}{2} u_k^{(i+1)^T} \cdot \left(g_k^T \nabla^2 V_k^i g_k\right) u_k^{(i+1)}. \quad (30)\end{aligned}$$

Substituting (26) into (30), the difference can be obtained as

$$\begin{aligned}\Delta V^{(i)}(x(k)) &= -x(k)^T Q x(k) - u_k^{(i+1)^T} \mathrm{R} u_k^{(i+1)} \\ &\quad - \frac{1}{2}\left(u_k^{(i+1)} - u_k^{(i)}\right)^T \\ &\quad \cdot \left(g_k^T \nabla^2 V_k^i g_k + 2R\right)\left(u_k^{(i+1)} - u_k^{(i)}\right). \quad (31)\end{aligned}$$

Since $g_k^T \nabla^2 V_k^i g_k + 2R$ and R and Q are positive–definite matrixes, we get

$$\Delta V^{(i)}(x(k)) \le -x(k)^T Q x(k) \le -\lambda_{\min}(Q) \left\| x((k)) \right\|^2. \quad (32)$$

This implies that the difference of $V^{(i)}(x)$ along the system $\left(f, g, u^{(i+1)}\right)$ trajectories is negative for $x(k) \neq 0$. Thus, $V^{(i)}(x)$ is a Lyapunov function for $u^{(i+1)}$ on Ω and the system with feedback control $u^{(i+1)}$ is locally asymptotically stable.

Second, we need to prove that the cost function of the system with the updated control $u^{(i+1)}$ is finite. Since $u^{(i)}$ is an admissible control, from Definition 2.1 and (4), we have

$$V^{(i)}(x(0)) = J\left(x(0); u^{(i)}\right) < \infty, \qquad \text{for } x(0) \in \Omega. \quad (33)$$

The cost function for $u^{(i+1)}$ can be written as

$$J(x(0); u^{(i+1)}) = \sum_{k=0}^{\infty}\left((x(k)^T Q x(k)) + u_k^{(i+1)^T} \mathrm{R} u_k^{(i+1)}\right) \quad (34)$$

where $x(k)$ is the trajectory of system with admission control $u^{(i+1)}$. From (31) and (34), we have

$$\begin{aligned}&V^{(i)}(x(\infty)) - V^{(i)}(x(0)) \\ &= \sum_{k=0}^{\infty} \Delta V^{(i)}(k) = \sum_{k=0}^{\infty} -x(k)^T Q x(k) - u_k^{(i+1)^T} \mathrm{R} u_k^{(i+1)} \\ &\quad - \frac{1}{2}\left(u_k^{(i+1)} - u_k^{(i)}\right)^T \left(g_k^T \nabla^2 V_k^i g_k + 2R\right) \cdot \left(u_k^{(i+1)} - u_k^{(i)}\right) \\ &= -J\left(x(0); u^{(i+1)}\right) - \frac{1}{2}\sum_{k=0}^{\infty}\left(u_k^{(i+1)} - u_k^{(i)}\right)^T \\ &\quad \cdot \left(g_k^T \nabla^2 V_k^i g_k + 2R\right)\left(u_k^{(i+1)} - u_k^{(i)}\right). \quad (35)\end{aligned}$$

Since $x(\infty) = 0$ and $V^{(i)}(x)|_{x=0} = 0$, we get $V^{(i)}(x(\infty)) = 0$. Rewriting (35), we have

$$\begin{aligned}J\left(x(0); u^{(i+1)}\right) &= V^{(i)}(x(0)) - \frac{1}{2}\sum_{k=0}^{\infty}\left(u_k^{(i+1)} - u_k^{(i)}\right)^T \\ &\quad \cdot \left(g_k^T \nabla^2 V_k^i g_k + 2R\right)\left(u_k^{(i+1)} - u_k^{(i)}\right). \quad (36)\end{aligned}$$

From (33) and (36), and considering that $g_k^T \nabla^2 V_k^i g_k + 2R$ is a positive–definite matrix function, we obtain

$$J\left(x(0); u^{(i+1)}\right) \le V^{(i)}(x(0)) = J\left(x\left(0; u^{(i)}\right)\right) < \infty. \quad (37)$$

Third, since $V^{(i)}$ is continuously differentiable and $g : \Re^m \to \Re^n$ is a Lipschitz continuous function on the set Ω in $\Re^n$, the new control law $u^{(i+1)}$ is continuous. Since $V^{(i)}$ is a positive–definite function, it attains a minimum at the origin, and thus, $\nabla V^{(i)}(x)$ and $\nabla^2 V^{(i)}(x)$ must vanish at the origin. This implies that $u^{(i+1)}(x)|_{x=0} = 0$.

Finally, following the Definition 1.1, one can conclude that the updated control function $u^{(i+1)}$ is admissible on Ω.

Proof of the Improved Control: To show the second part of the Theorem 2.2, we need to prove that $V^{(i)}(x(0)) \ge V^{(i+1)}(x(0))$ which means the cost function will be reduced by updating the feedback control. Because $u^{(i+1)}$ is an admissible control, there exists a positive–definite function $V^{(i+1)}$ such that $GHJB\left(V^{(i+1)}, u^{(i+1)}\right) = 0$ on $x \in \Omega$. According to the Theorem 2.1, we can get

$$V^{(i+1)}(x(0)) = J\left(x(0); u^{(i+1)}\right). \quad (38)$$

From (36) and (38), we know that

$$\begin{aligned}&V^{(i+1)}(x(0)) - V^{(i)}(x(0)) \\ &= -\frac{1}{2}\sum_{k=0}^{\infty}\left(u_k^{(i+1)} - u_k^{(i)}\right)^T \\ &\quad \cdot \left(g_k^T \nabla^2 V_k^i g_k + 2R\right)\left(u_k^{(i+1)} - u_k^{(i)}\right) \\ &\le 0. \quad (39)\end{aligned}$$

Theorem 2.2 suggests that after solving the GHJB equation and updating the control function by using (26), the system perfor-

mance can be improved. If the control function is iterated successively, the updated control will converge close to the solution of HJB, which then renders the optimal control function. The GHJB becomes the Hamilton–Jacobi–Bellman (HJB) equation on substitution of the optimal control function $u^*(x)$. The HJB equation can now be defined in DT as follows.

Definition 2.4 (HJB Equation for the Nonlinear DT): The HJB equation in DT in this framework can be expressed as

$$\begin{aligned}&\frac{1}{2}(f(x)+g(x)u^*(x)-x)^T\cdot\nabla^2V^*(x)\,(f(x)+g(x)u^*(x)-x)\\&\quad+\nabla V^*(x)^T\,(f(x)+g(x)u^*(x)-x)\\&\quad+x^TQx+u^*(x)^T\mathrm{R}u^*(x)=0 \qquad (40)\\&V^*(x)\,|_{x=0}=0 \qquad (41)\end{aligned}$$

where the optimal control function for the DT system is given by

$$\begin{aligned}u^*(x)=&-\left[g(x)^T\nabla^2V^*(x)g\,(x)+2R\right]^{-1}\\&\cdot g(x)^T\left(\nabla V^*(x)+\nabla^2V^*(x)\,(f(x)-x)\right). \qquad (42)\end{aligned}$$

Note $V^*(x)$ is the optimal solution to the HJB (40). It is important to note that the GHJB is linear in the value function derivative while the HJB equation is nonlinear in the value function derivative. Solving the GHJB equation requires solving linear partial difference equations, while the HJB equation solution involves nonlinear partial difference equations, which may be difficult to solve. This is the reason for introducing the successive approximation technique using GHJB. In the successive approximation method, one solves (20) for $V(x)$ given a stabilizing control $u(x)$, and then, finds an improved control based on $V(x)$ using (26). In the following, Corollary 2.1 indicates that if the initial control function is admissible, then repetitive application of (20) and (26) is a contraction map, and the sequence of solutions $V^{(i)}(k)$ converges to the optimal HJB solution $V^*(k)$.

Corollary 2.1 (Convergence of Successive Approximations): Given an initial admissible control $u^0(x)\in\Omega_u$ by iteratively solving GHJB (20) and updating the control function using (26), the sequence of solutions $V^i(x)$ will converge to the optimal HJB solution $V^*(x)$.

Proof: From the proof of Theorem 2.2, it is clear that after iteratively solving the GHJB equation and updating the control, the sequence of solutions $V^{(i)}$ is a decreasing sequence with a lower bound. Since $V^{(i)}$ is a positive–definite function, $V^{(i)}>0$, and $V^{(i+1)}-V^{(i)}\le 0$, the sequence of solutions $V^{(i)}$ will converge to a positive–definite function $V^{(i)}=V^{(i+1)}=V^d$, when $i\to\infty$. Due to the uniqueness of solutions of the HJB equation [11], now it is necessary to show that $V^d=V^*$. When $V^{(i)}=V^{(i+1)}=V^d$, from (39), we can only obtain $u^{(i)}=u^{(i+1)}$. Using (26) and taking $u^{(i)}=u^{(i+1)}$, we obtain

$$\begin{aligned}u^{(i)}(x)=&u^{(i+1)}(x)=-\left[g(x)^T\nabla^2V^{(i)}(x)g(x)+2R\right]^{-1}g(x)^T\\&\cdot\left(\nabla V^{(i)}(x)+\nabla^2V^{(i)}(x)\,(f(x)-x)\right). \qquad (43)\end{aligned}$$

The GHJB equation for $u^{(i)}$ can now be expressed as

$$\begin{aligned}&\frac{1}{2}\left(f(x)+g(x)u^{(i)}(x)-x\right)^T\\&\quad\cdot\nabla^2V^{(i)}(x)\left(f(x)+g(x)u^{(i)}(x)-x\right)\\&\quad+\nabla V^{(it)}(x)^T\left(f(x)+g(x)u^{(i)}(x)-x\right)\\&\quad+x^TQx+u^{(i)}(x)^T\mathrm{R}u^{(i)}(x)=0 \qquad (44)\\&V^{(i)}(x)\,|_{x=0}=0. \qquad (45)\end{aligned}$$

From (43)–(45), we can conclude that these equations are nothing but the well-known HJB equation, which is presented in Definition 2.4. This implies that $V^{(i)}$ converges to V^* and $u^{(i)}$ converges to $u^{(*)}$.

Note that (40)–(42) are the HJB equations under the small perturbation assumption. The more general and ideal HJB equations are, then

$$\begin{aligned}&V^*\left(f(x)+g(x)u^*(x)\right)-V^*(x)\\&\quad+x^TQx+u^*(x)^T\mathrm{R}u^*(x)=0 \qquad (46)\\&V^*(x)\,|_{x=0}=0 \qquad (47)\end{aligned}$$

where $u^*(x)$ is the solution of

$$\left.\frac{\partial V^*(\alpha)}{\partial(\alpha)}\right|_{\alpha=f(x)+g(x)u^*(x)}\cdot g(x)+2u^*(x)^TR=0. \qquad (48)$$

The ideal GHJB equations are given by

$$\begin{aligned}&V\left(f(x)+g(x)u(x)\right)-V(x)\\&\quad+x^TQx+u(x)^T\mathrm{R}u(x)=0 \qquad (49)\\&V(x)\,|_{x=0}=0. \qquad (50)\end{aligned}$$

Although for the given admissible control $u^{(i)}(x)$, the ideal GHJB (46) can be solved using an NN to get $V^{(i)}(x)$. However, without the small perturbation assumption, the updated control law $u^{(i+1)}(x)$ cannot be easily solved from

$$\left.\frac{\partial V^{(i)}(\alpha)}{\partial(\alpha)}\right|_{\alpha=f(x)+g(x)u^{(i+1)}(x)}\cdot g(x)+2u^{(i+1)}(x)^TR=0. \qquad (51)$$

Additionally, it is quite difficult to show $u^{(i+1)}(x)$ as an admissible and improved control.

Next, we show the consistency between proposed GHJB and DP using small perturbation assumption.

Remark 5: Consistency Between GHJB and DP: From the DP principle [2], the optimal controller can be given as

$$u^*(x)=-\frac{1}{2}R^{-1}gx^T\frac{dV^*(x(k+1))}{dx(k+1)}. \qquad (52)$$

However, the optimal controller for a general nonlinear DT system is difficult to design and only for the special case of linear systems, when $u^*(k)$ can be solved in terms of $x(k)$ and not in terms of $x\,(k+1)$. But consider the derivative of function V^* expressed as

$$\frac{dV^*\left(x\,(k+1)\right)}{dx\,(k+1)}=\nabla V^*\left(x\,(k+1)\right)=\nabla V^*_{k+1}. \qquad (53)$$

Since the small perturbation assumption is considered, the high-order terms in Taylor expansion of ∇V_{k+1}^* can be ignored to get

$$\nabla V_{k+1}^* - \nabla V_k^* = \nabla^2 V_k^* \left(x\left(k+1\right) - x(k)\right). \tag{54}$$

Considering system (1), (54) can be rewritten as

$$\nabla V_{k+1}^* = \nabla V_k^* + \nabla^2 V_k^* \left(f_k + g_k u_k - x(k)\right). \tag{55}$$

Using (55), (52) can be written as

$$u_k^* = -\frac{1}{2} R^{-1} g_k^T \left(\nabla V_k^* + \nabla^2 V_k^* \left(f_k + g_k u_k^* - x(k)\right)\right). \tag{56}$$

By solving u_k^* in (56), we can obtain

$$u_k^* = -\left[g_k^T \nabla^2 V_k^* g_k + 2R\right]^{-1} \cdot g_k^T \left(\nabla V_k^* + \nabla^2 V_k^* \left(f_k - x(k)\right)\right). \tag{57}$$

Equation (57) shows that u_k^* can be solved only in terms of $x(k)$ under the assumption that higher than second-order terms in the Taylor series expansion can be ignored. Equation (52) is consistent with (42). It is important to note that nonlinear approximation theory will be utilized later to approximate the value function which provides a tradeoff between computation and accuracy. In summary, the value function in the proposed method is approximated and iterated until convergence. Then, the policy iteration is performed using the optimal value function. The value and policy iterations are quite similar to the case of approximate DP [16].

In order to verify HJB for a linear DT system, the proposed approach is utilized next.

Remark 6: The ARE associated with the optimal control of linear DT system can be derived from the DT HJB equation. Consider the following linear DT system and cost function defined in (22) as:

$$x\left(k+1\right) = Ax(k) + Bu(k) \tag{58}$$

where $V^*(x) = x^T P x$ and P is a symmetric positive–definite matrix. The gradient vector and Hessian matrix of $V^*(x)$ can be derived as $\nabla^2 V^*(x) = 2P$ and $\nabla V^*(x) = 2Px$. The HJB (40) and (42) can be rewritten as

$$\begin{aligned}&\left(Ax + Bu^*(x) - x\right)^T P \left(Ax + Bu^*(x) - x\right)\\ &\quad + 2x^T P \left(Ax + Bu^*(x) - x\right)\\ &\quad + x^T Q x + u^*(x)^T \mathrm{R} u^*(x) = 0\end{aligned} \tag{59}$$

$$u^*(x) = -\left[2B^T PB + 2R\right]^{-1} B^T \left(2Px + 2P\left(Ax - x\right)\right). \tag{60}$$

After simplifying (59) and (60), we obtain

$$P = Q + A^T PA - A^T PB \left(R + B^T PB\right)^{-1} \cdot B^T PA \tag{61}$$

$$u^*(x) = -\left[B^T PB + R\right]^{-1} B^T PAx. \tag{62}$$

Equation (61) is nothing but ARE [9] for linear DT system and (62) is the optimal control of linear DT system. Next, we show the difference between GHJB in continuous and DT.

Remark 7: Difference Between GHJB in Continuous- and Discrete-Time Systems With Small Perturbation: When the first-order term in Taylor extension of cost function $V(x)$ is considered alone, (8) can be rewritten as

$$V\left(x\left(k+1\right)\right) = V\left(x(k)\right) + \nabla V_k^T \left(x\left(k+1\right) - x(k)\right). \tag{63}$$

By following the same steps in Theorems 2.1 and 2.2, we can obtain GHJB equation for this case

$$\begin{aligned}&\frac{\partial V(k)}{\partial x(k)}^T \left(f(k) + g(k)u(k) - x(k)\right)\\ &\quad + Q(k) + u(k)^T \mathrm{R} u(k) = 0\end{aligned} \tag{64}$$

$$V(x)\,|_{x=0} = 0 \tag{65}$$

and the updated control law

$$u^{(i+1)}(x) = -\frac{1}{2} R^{-1} g^T(k) \nabla V^{(i)}(x). \tag{66}$$

These equations are nothing but the GHJB equations in continuous time [21]. If the second-order terms from the Taylor series expansion of the cost function are considered, the GHJB equations in DT derived in this paper show improvements in approximating the cost function provided the perturbation is sufficiently small. In many cases, the cost function $V(x)$ is quadratic function in x. Then, cost function $V^*(x)$ and also the optimal control $u^*(x)$ can be exactly calculated by the proposed GHJB method. Therefore, the proposed GHJB in DT appears to be more accurate than directly applying the continuous-time GHJB method to a nonlinear DT system.

By considering the higher order terms, approximation accuracy can be improved but a tradeoff exists between accuracy and computational complexity for practical realization of optimal control [11]. Therefore, for practical design considerations, cost or value function should be approximated using the aforementioned approach.

III. NN Least Squares Approach

In Section II, we described that by recursively solving the GHJB equation and by updating the control function, we could improve the closed-loop performance of control laws that are known to be admissible. Furthermore, we can get arbitrarily close to the optimal control by iterating the GHJB solution enough number of times. Although the GHJB equation is in theory easier to solve than the HJB equation, there is no general closed-form solution available to this equation. In [19], Beard used Galerkin's spectral method to get approximate solution to GHJB in continuous time at each iterating step and the convergence is shown in the overall run. This technique does not set the GHJB equation to zero at each iterating step, but to a residual error instead. The Galerkin spectral method requires the computation of a large number of integrals in order to minimize this residual error.

The purpose of this section is to show how we approximate the solution of the GHJB equation in DT using NNs such that the controls which result from the solution are in feedback form. It is well known that NNs can be used to approximate smooth

functions on prescribed compact set [6]. We approximate $V(x)$ with an NN

$$V_L(x) = \sum_{j=1}^{L} w_j \sigma_j(x) = W_L^T \bar{\sigma}_L(x) \tag{67}$$

where the activation function vector $\sigma_j(x) : \Omega \to \Re$ is continuous, $\sigma_j(x)|_{x=0} = 0$, the NN weights are w_j, and L is the number of hidden layer neurons. The vectors $\bar{\sigma}_L(x) \equiv [\sigma_1(x), \sigma_2(x), \ldots, \sigma_L(x)]^T$ and $W_L \equiv [w_1, w_2, \ldots, w_L]^T$ are the vector of activation function and NN weight matrix, respectively. The NN weights will be tuned to minimize the residual error in a least squares sense over a set of points within the stability region of the initial stabilizing control. Least squares solution [5] attains the lowest possible residual error with respect to the NN weights.

For the GHJB $(V, u) = 0$, V is replaced by V_L having a residual error as

$$\text{GHJB}\left(V_L = \sum_{j=1}^{L} w_j \sigma_j, u\right) = e_L(x). \tag{68}$$

To find the least squares solution, the method of weighted residuals is used [5]. The weights w_j are determined by projecting the residual error onto $(\partial(e_L(x))/\partial W_L)$ and setting the result to zero $\forall\, x \in \Omega$, i.e.,

$$\left\langle \frac{\partial(e_L(x))}{\partial W_L}, e_L(x) \right\rangle = 0. \tag{69}$$

When expanded, (69) becomes

$$\begin{aligned}&\left\langle \nabla\bar{\sigma}_L \Delta x + \frac{1}{2}\Delta x^T \nabla^2 \bar{\sigma}_L \Delta x, \nabla\bar{\sigma}_L \Delta x + \frac{1}{2}\Delta x^T \nabla^2 \bar{\sigma}_L \Delta x \right\rangle \\ &\quad \cdot W_L + \left\langle x^T Q x + u^T \mathrm{R} u, \nabla\bar{\sigma}_L \Delta x + \frac{1}{2}\Delta x^T \nabla^2 \bar{\sigma}_L \Delta x \right\rangle \\ &\quad = 0\end{aligned} \tag{70}$$

where

$$\begin{aligned}\nabla\bar{\sigma}_L &= [(\partial\sigma_1(x)/\partial x), (\partial\sigma_2(x)/\partial x), \ldots, (\partial\sigma_L(x)/\partial x)]^T \\ \nabla^2\bar{\sigma}_L &= [(\partial^2\sigma_1(x)/\partial x^2), (\partial^2\sigma_2(x)/\partial x^2), \ldots, (\partial^2\sigma_L(x)/\partial x^2)]^T \\ \Delta x &= f + gu - x.\end{aligned}$$

In order to proceed, the following technical results are needed.

Lemma 3.1: If the set $\{\sigma_j(x)\}_1^L$ is linearly independent and $u \in \Omega_u$, then the set

$$\left\{ \nabla\sigma_j^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 \sigma_j \Delta x \right\}_1^L \tag{71}$$

is also linearly independent.

Proof: Calculating the $\sigma(x(j))$ along the system trajectories (f, g, u) for $x(0) \in \Omega$ by using the similar formulation of (7) and (11), we have

$$\begin{aligned}&\sigma_j(x(\infty)) - \sigma_j(x(0)) \\ &\quad = \sum_{k=0}^{\infty}\left[\nabla\sigma_j(k)^T \Delta x(k) + \frac{1}{2}\Delta x(k)^T \nabla^2 \sigma_j(k) \Delta x(k)\right].\end{aligned} \tag{72}$$

Since $u \in \Omega_u$ is an admissible control, the system (f, gu) is stable and $x(\infty) = 0$. With the condition on the active function $\sigma(x)|_{x=0} = 0$, we have $\sigma(x(\infty)) = 0$. Rewriting (72) with the previous results, we have

$$\begin{aligned}&\sigma(x(0)) \\ &\quad = -\sum_{k=0}^{\infty}\left[\nabla\sigma_j(k)^T \Delta x(k) + \frac{1}{2}\Delta x(k)^T \nabla^2 \sigma_j(k) \Delta x(k)\right].\end{aligned} \tag{73}$$

Extending (73) into the vector formulation gives

$$\begin{aligned}&\bar{\sigma}_L(x(0)) \\ &\quad = -\sum_{k=0}^{\infty}\left[\nabla\bar{\sigma}_L(k)^T \Delta x(k) + \frac{1}{2}\Delta x(k)^T \nabla^2 \bar{\sigma}_L(k) \Delta x(k)\right].\end{aligned} \tag{74}$$

Now, suppose that the Lemma 3.1 is not true. Then, there exists a nonzero $C \in \Re^L$ such that

$$C^T\left(\nabla\bar{\sigma}_L \Delta x + \frac{1}{2}\Delta x^T \nabla^2 \bar{\sigma}_L(k) \Delta x\right) \equiv 0, \qquad \text{for } x \in \Omega. \tag{75}$$

From (74) and (75), we have

$$\begin{aligned}C^T \bar{\sigma}_L(x(0)) = -\sum_{k=0}^{\infty}\Bigg[&C^T\Bigg(\nabla\bar{\sigma}_L(k)\Delta x(k) \\ &+ \frac{1}{2}\Delta x(k)^T \nabla^2 \bar{\sigma}_L(k) \Delta x(k)\Bigg)\Bigg] \\ \equiv 0, \qquad &\text{for } x(0) \in \Omega\end{aligned} \tag{76}$$

which contradicts the linear independence of $\{\sigma_j(x)\}_1^L$; so the set (71) must be linearly independent. Equation (76) can be rewritten, after defining $\left\{\nabla\sigma_j^T \Delta x + 1/2\Delta x^T \nabla^2 \sigma_j \Delta x\right\}_1^L \triangleq \{\theta_j\}_1^L \triangleq \bar{\theta}$ as

$$W_L = -\left\langle \bar{\theta}, \bar{\theta} \right\rangle^{-1} \left\langle Q + u^T \mathrm{R} u, \bar{\theta} \right\rangle. \tag{77}$$

Because of Lemma 3.1, the term $\left\langle \bar{\theta}, \bar{\theta} \right\rangle$ is full rank, and thus, is invertible. Therefore, a unique solution for W_L exists. From (77), we need to calculate the inner product of $\langle f(x), g(x) \rangle$. In Hilbert space, we define the inner product as

$$\langle f(x), g(x) \rangle = \int_{\Omega} f(x) g(x) dx. \tag{78}$$

Executing the integration in (78) is computationally expensive. However, the integration can be approximated to a suitable degree using the Riemann definition of integration so that the inner product can be obtained. This in turn results in a nearly optimal, computationally tractable solution algorithm.

Lemma 3.2 (Riemann Approximation of Integrals): An integral can be approximated as

$$\int_a^b f(x) dx = \lim_{\|\delta x\| \to 0} \sum_{i=1}^{n} f(\bar{x}_i) \cdot \delta x \tag{79}$$

where $\delta x = x_i - x_{i-1}$ and f is bounded on $[a, b]$ [3].

Introducing a mesh on Ω, with mesh size equal to δx, which is taken very small, we can rewrite some terms in (77) as (80) and

(81), shown at the bottom of the page, where p in x_p represents the number of points of the mesh. This number increases as the mesh size is reduced. Using Lemma 3.2, we can rewrite (70) as

$$XW_L + Y = 0. \tag{82}$$

This implies that we can calculate

$$W_L = -\left(X^T X\right)^{-1}(XY). \tag{83}$$

An interesting observation is that (83) is the standard least squares method of estimation for a mesh on Ω. Note that the mesh size δx should be such that the number of points p is greater or equal to the order of the approximation L and the activation functions should be linearly independent. These conditions guarantee a full rank for $(X^T X)$.

The optimal control of nonlinear DT system can be obtained offline by going through the following six steps.

1) Define an NN as $V = \sum_{j=1}^{L} w_j \sigma_j(x)$ to approximate smooth function of $V(x)$.
2) Select an admissible feedback control law $u^{(1)}$.
3) Find $V^{(1)}$ associated with $u^{(1)}$ to satisfy GHJB by applying least square method (LSM) to obtain the NN weights W^1;
4) Update the control as

$$u^{(2)}(x) = -\left[g^T(x)\nabla^2 V^{(1)}(x) g(x) + 2R\right]^{-1} \cdot g^T(k)\left(\nabla V^{(1)}(x) + \nabla^2 V^{(1)}(x)\left(f(x) - x\right)\right). \tag{84}$$

5) Find $V^{(2)}$ associated with u_2 to satisfy GHJB by using LSM to obtain W^2.
6) If $V^1(0) - V^2(0) \le \varepsilon$, where ε is a small positive constant, then $V^* = V^{(1)}$ and stop. Otherwise, go back to step 4) by increasing the index by one.

After we get V^*, the optimal state feedback control, which can be implemented online, can be described as

$$u^*(x) = -\left[g^T(x)\nabla^2 V^*(x) g(x) + 2R\right]^{-1} g^T(x)\left(\nabla V^*(x) + \nabla^2 V^*(x)\left(f(x) - x\right)\right). \tag{85}$$

IV. Numerical Examples

The power of the technique is demonstrated for the case of HJB by using three examples. First, we take on a linear DT system to compare the performance of the proposed approach to that of the standard solution obtained by solving Riccati equation. This comparison will present that the proposed approach works for a linear system and renders an optimal solution. Second, we will use a general nonlinear practical system and a real-world two-link planar revolute–revolute (RR) robot arm system to demonstrate that the proposed approach renders a suboptimal solution for nonlinear DT systems.

In all of the examples that we present in this section, the basis functions required will be obtained from even polynomials so that the NN can approximate the positive–definite function or value function. If the dimension of the system is n and the order of approximation is M, then we use all of the terms in expansion of the polynomial [21]

$$\sum_{j=1}^{M/2}\left(\sum_{k=1}^{n} x_k\right)^{2j}. \tag{86}$$

The resulting basis functions for a 2-D system is

$$\left\{x_1^2, x_1x_2, x_2^2, x_1^4, x_1^3x_2, \ldots, x_2^M\right\}. \tag{87}$$

1) Example 1 (Linear DT System): Consider the linear DT system (52), where

$$A = \begin{bmatrix} 0 & -0.8 \\ 0.8 & 1.8 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ -1 \end{bmatrix}. \tag{88}$$

Define the cost function

$$J(u; x(0)) = \sum_{k=0}^{N}\left(x_1(k)^2 + x_2(k)^2 + u(k)^2\right). \tag{89}$$

Define the NN with the activation functions containing polynomial functions up to the sixth order of approximation by using $n = 2$ and $M = 6$. From (86), the NN can be constructed as

$$\begin{aligned} V(x_1, x_2) = {} & w_1x_1^2 + w_2x_2^2 + w_3x_1x_2 + w_4x_1^4 + w_5x_2^4 \\ & + w_6x_1^3x_2 + w_7x_1^2x_2^2 + w_8x_1x_2^3 + w_9x_1^6 \\ & + w_{10}x_2^6 + w_{11}x_1^5x_2 + w_{12}x_1^4x_2^2 + w_{13}x_1^3x_2^3 \\ & + w_{14}x_1^2x_2^4 + w_{15}x_1x_2^5. \end{aligned} \tag{90}$$

Select the initial control law $u_1(k) = x_2(k)$, which is admissible. Update the control with

$$u_{i+1} = -\left[B^T\nabla^2 V^{(i)}(k) B + 2\right]^{-1} \cdot B^T\left(\nabla V^{(i)}(k) + \nabla^2 V^{(i)}(k)\left(A - I\right)x(k)\right) \tag{91}$$

where $u^{(i)}$ and $V^{(i)}$ satisfy the GHJB equation

$$\begin{aligned} & \frac{1}{2}\left(Ax + Bu^{(i)}(x) - x\right)^T \cdot \nabla^2 V^{(i)}(x)\left(Ax + Bu^{(i)}(x) - x\right) \\ & + \nabla V^{(i)}(x)^T\left(Ax + Bu^{(i)}(x) - x\right) \\ & + x^Tx + u^{(i)}(x)^2 = 0. \end{aligned} \tag{92}$$

In the simulation, the mesh size δx is selected as 0.01 and the asymptotic stability region is chosen for the states as $-0.5 \le$

$$X = \left[\nabla\bar{\sigma}_L\Delta x + \frac{1}{2}\Delta x^T\nabla^2\bar{\sigma}_L\Delta x|_{x=x_1} \ldots \nabla\bar{\sigma}_L\Delta x + \frac{1}{2}\Delta x^T\nabla^2\bar{\sigma}_L\Delta x|_{x=x_p}\right]^T \tag{80}$$

$$Y = \begin{bmatrix} x^TQx + u(x)^T\mathrm{R}u(x)|_{x=x_1} \\ \vdots \\ x^TQx + u^T(x)\mathrm{R}u(x)|_{x=x_p} \end{bmatrix} \tag{81}$$

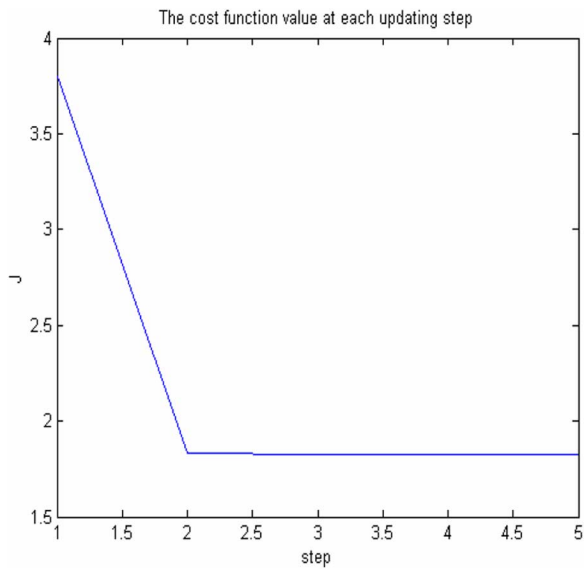

Fig. 1. Cost function at each updating step.

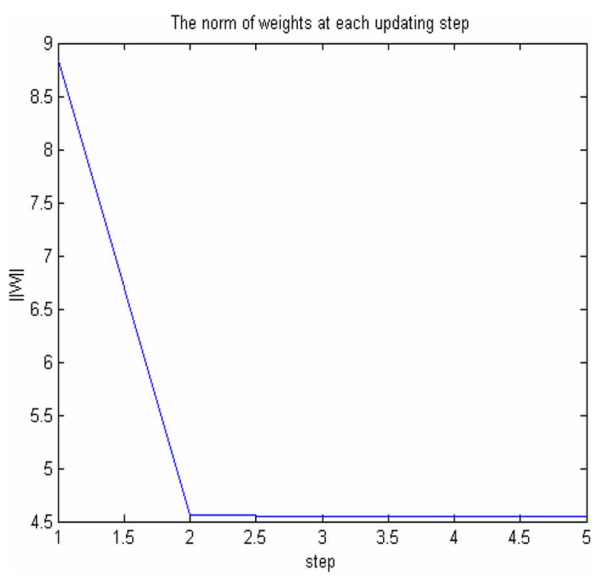

Fig. 2. Norm of NN weights at each updating step.

$x_1 \leq 0.5$ and $-0.5 \leq x_2 \leq 0.5$. The small positive approximation error constant is selected as $\varepsilon = 0.00001$. The initial states are selected as $x_1(0) = x_2(0) = 0.5$. The simulation step is selected as $N = 100$. After updating five times, the optimal value function and V^* the optimal control u^* are obtained. Fig. 1 shows the cost-functional value and Fig. 2 shows the norm of NN weights at each updating step. From these plots, it is clear that the cost-functional value continues to decrease until it reaches a minimum and, afterwards, it remains constant.

After we obtain the optimal control based on the GHJB method, we implement the initial admissible control and the optimal control on the system, respectively. Fig. 3 shows the (x_1, x_2) trajectory with an initial admissible control, whereas Fig. 4 illustrates the (x_1, x_2) trajectory with the GHJB-based optimal control. From these figures, we can conclude that the updated control is not only an admissible control but it also converges to the optimal control.Table I presents this with different initial admissible controls we arbitrarily selected; the final NN weights, the optimal cost-functional values, and the updated control function will converge to the unique optimal control. This method is independent on the selection of the initial admissible control for the linear DT systems.

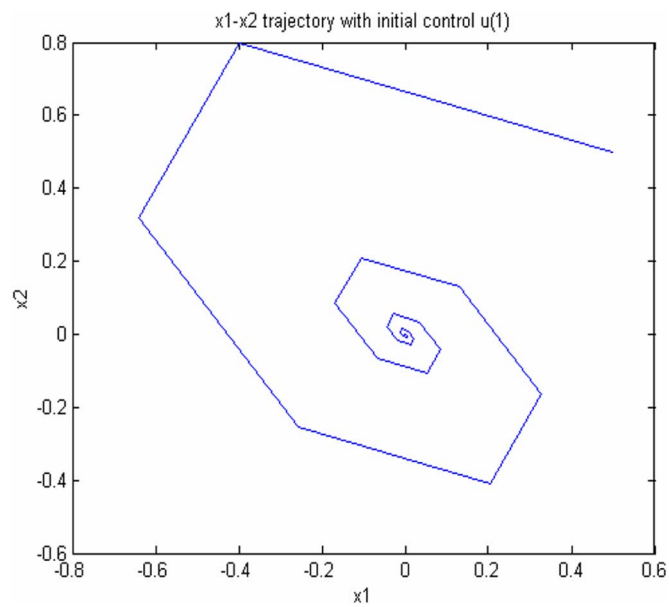

Fig. 3. State trajectory (x_1, x_2) with the initial control.

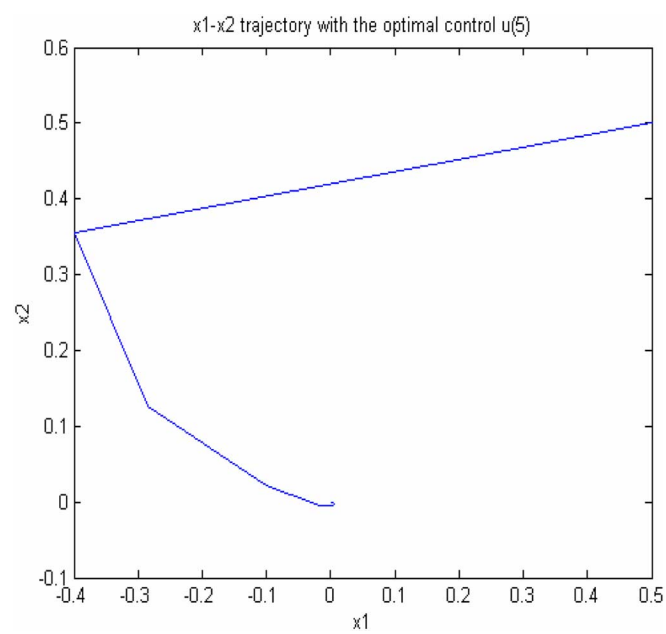

Fig. 4. State trajectory (x_1, x_2) with the GHJB-based optimal control.

TABLE I
COST VALUE WITH ADMISSIBLE CONTROLS

Initial control u_1	Optimal Cost Value	$\|W^5\|$
x_2	1.826	4.5449
$x_1 + x_2$	1.826	4.5449
$x_1 + 2x_2$	1.826	4.5449

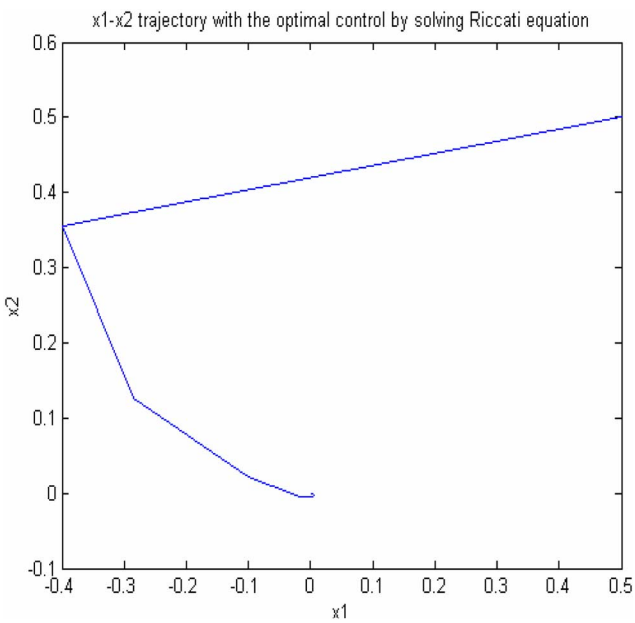

Fig. 5. State trajectory (x_1, x_2) with Riccati-based optimal control.

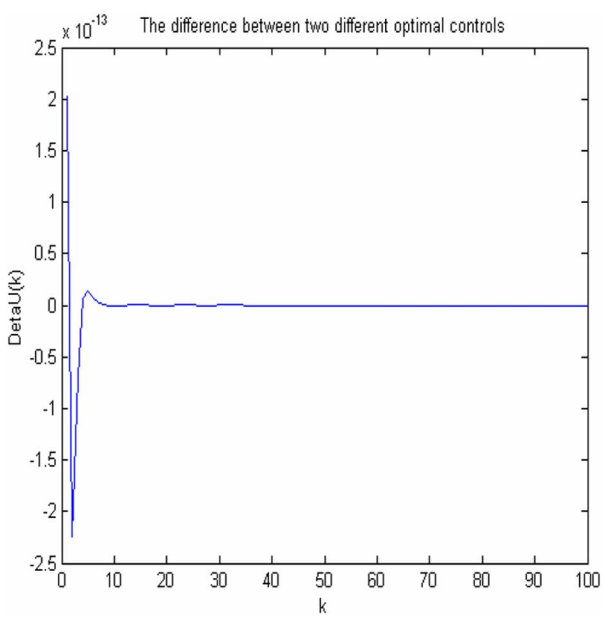

Fig. 6. Difference between the two optimal controls.

In order to evaluate whether the proposed method converges to the optimal control obtained from classical optimal control methods, we use the Riccati equation in DT to solve the LQR optimal control problem for this system [9]. Riccati equation in DT is given by [9]

$$P(i) = A^T P(i+1) \cdot \left(I + BR^{-1}B^T P(i+1)\right)^{-1} A + Q \tag{93}$$

$$P(N) = Q(N) = I \times 10^5 \tag{94}$$

$$u(i) = -\left(R + B^T P(i+1) B\right)^{-1} \cdot B^T P(i+1) Ax. \tag{95}$$

Fig. 5 displays that the optimal (x_1, x_2) trajectory generated by solving Riccati equation whereas Fig. 6 depicts the error between the control inputs obtained from the proposed and the Riccati methods. Table II shows the optimal cost-functional value obtained from the two methods. Comparing Fig. 5 with Fig. 3, and from Fig. 6 and Table II, we can observe that the trajectories and the optimal control inputs are the same. We can conclude that for linear DT system, the updated control associated with GHJB equation will converge to the optimal control.

TABLE II
COMPARISON OF CONTROL METHODS

Control Method	Optimal Cost Value
Successively Updating	1.826
Solving Riccati Equation	1.826

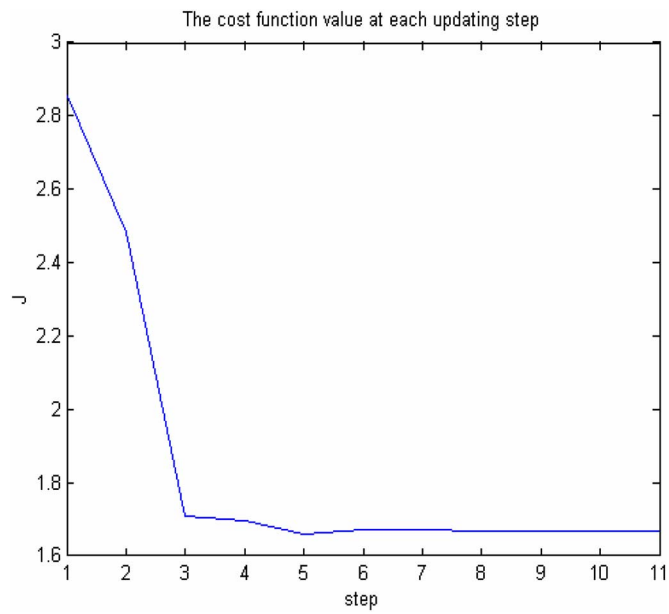

Fig. 7. Cost-functional value of at each updating step.

2) Example 2 (Nonlinear DT System): Consider the nonlinear DT system given by

$$x(k+1) = f(x(k)) + g(x(k))\, u(k) \tag{96}$$

where

$$f(x(k)) = \begin{bmatrix} -0.8x_2(k) \\ \sin(0.8x_1(k) - x_2(k)) + 1.8x_2(k) \end{bmatrix}$$

$$g(x(k)) = \begin{bmatrix} 0 \\ -x_2(k) \end{bmatrix}. \tag{97}$$

We select the initial control law as $u_1 = x_1 + 1.5x_2$ and the NN is also selected from (90). The simulation parameters and cost function are defined the same as in the Example 1. Fig. 7 shows the cost-functional value at each updating time and Fig. 8 shows the norm of NN weights. After updating 11 times, we get the optimal control u^* offline, and then, the optimal control is implemented with several initial conditions. Fig. 9 shows the state trajectory (x_1, x_2) with initial admissible control. By contrast, Fig. 10 shows the state trajectory (x_1, x_2) by solving the GHJB-based control with successive approximation. Different values of initial admissible controls are used to obtain the near-optimal control result. Table III shows, with different initial admissible controls, that the final norm of NN weights and

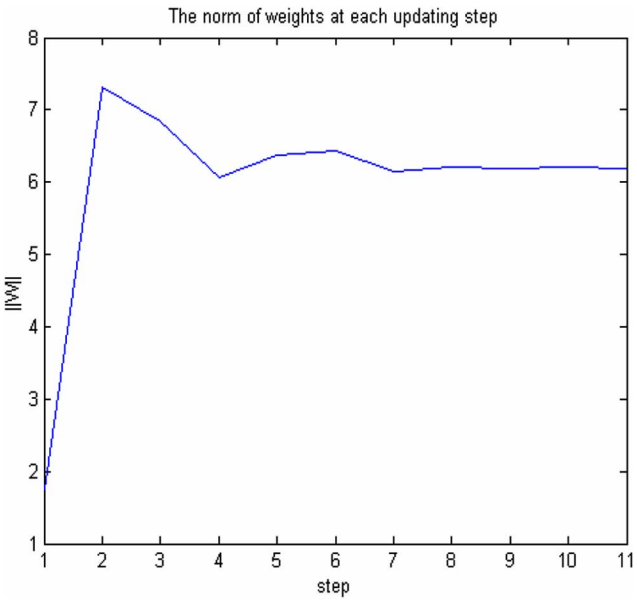

Fig. 8. Norm of NN weights at each updating step.

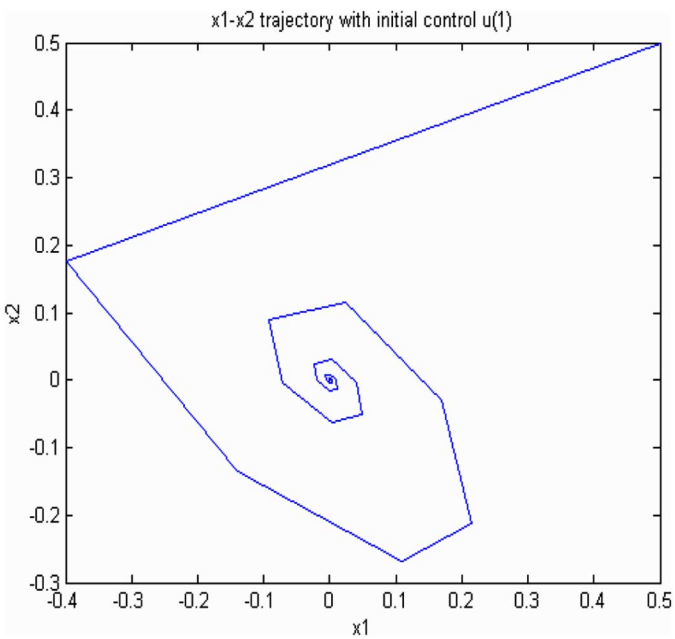

Fig. 9. State trajectory with initial admissible control.

the optimal cost-functional value are almost the same demonstrating the validity of the proposed GHJB-based solution.

3) Example 3 (Two-Link Planar RR Robot Arm System): A two-link planar RR robot arm used extensively for simulation in the literature is shown in Fig. 11. This arm is simple enough to simulate yet has all the nonlinear effects common to general robot manipulators. The DT dynamics of the two-link robot arm system is obtained by discretizing the continuous-time dynamics. In simulation, we apply the GHJB-based near-optimal control method to solve the nonlinear quadratic regulator problem. In other words, we seek a suboptimal control to move the arm to the desired position while minimizing the cost-functional value.

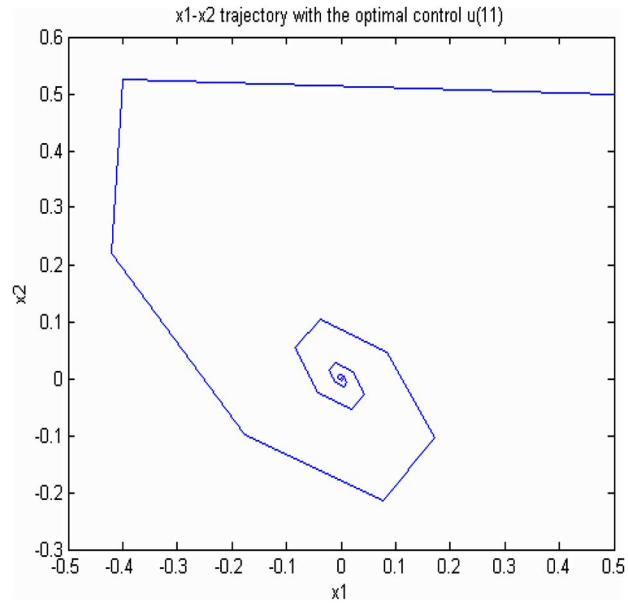

Fig. 10. State trajectory with GHJB-based optimal control.

TABLE III
GHJB-BASED NEAR-OPTIMAL CONTROL WITH INITIAL ADMISSIBLE CONTROL

Initial control u_1	Optimal Cost Value	$\|W^{11}\|$
$x_1+1.5x_2$	1.6692	6.1985
x_1	1.6692	6.1984
x_2	1.6691	6.1993
$0.3x_1+x_2$	1.6691	6.1993

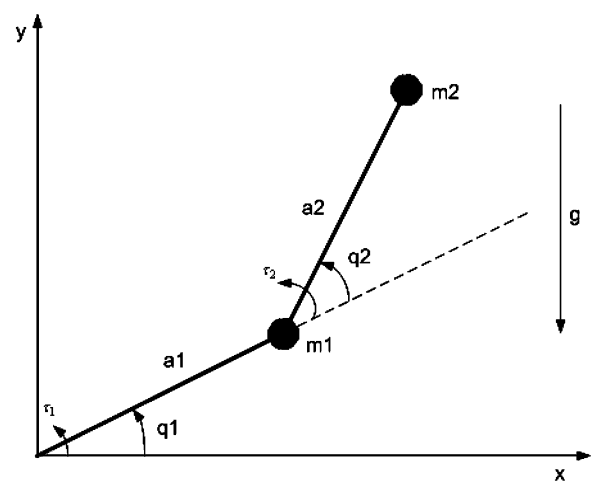

Fig. 11. Two-link planar robot arm.

The continuous-time dynamics model of two-link planar RR robot is given [6] as

$$\begin{bmatrix} \alpha+\beta+2\eta\cos q_2 & \beta+\eta\cos q_2 \\ \beta+\eta\cos q_2 & \beta \end{bmatrix}\begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} + \begin{bmatrix} -\eta\left(2\dot{q}_1\dot{q}_2+\dot{q}_2^2\right)\sin q_2 \\ \eta\dot{q}_1^2\sin q_2 \end{bmatrix} + \begin{bmatrix} \alpha e_1\cos q_1+\eta e_1\cos\left(q_1+q_2\right) \\ \eta e_1\cos\left(q_1+q_2\right) \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \quad (98)$$

where $\alpha = (m_1+m_2)\,a_1^2$, $\beta = m_2a_2^2$, $\eta = m_2a_1a_2$, and $e_1 = (g/a_1)$. We define the state and control variables as $x_1 =$

$q_1\ x_2 = q_2\ x_3 = \dot{q}_1\ x_4 = \dot{q}_2$ and $u = [\tau_1 \quad \tau_2]^T$. For simulation purposes, the parameters are selected as $m_1 = m_2 = 1$ kg, $a_1 = a_2 = 1$ m, and $g = 10$ m/s^2; then, $\alpha = 2$, $\beta = 1$, $\eta = 1$, and $e_1 = 10$. Rewriting the continuous-time dynamics as state equation, we get

$$\dot{x} = f(x) + g(x)u \tag{99}$$

where (100) and (101), shown at the bottom of the page, hold. The control objective is moving the arm from an initial state $x(0) = [(pi/3) \quad (pi/6) \quad 0 \quad 0]$ to the final state $x_d = [(pi/2) \quad 0 \quad 0 \quad 0]$ with the cost function defined as

$$J = \int_{t=0}^{\infty} \left(\|x(t) - x_d\|^2 + \|u(t)\|^2 \right) dt. \tag{102}$$

First, we will convert the continuous-time dynamics system and cost function into DT. Let us consider a DT system with a sampling period Δt and denote a time function $f(t)$ at $t = k\Delta t$ as $f(k)$, where k is a sampling number. If the sampling period Δt is sufficiently small compared to the time constant of the system, the response evaluated by DT methods will be reasonably accurate [9]. Therefore, we use the following approximation for the derivative of $f(k)$:

$$\dot{f}(k) \cong \frac{1}{\Delta t} \left(f(k+1) - f(k) \right). \tag{103}$$

Using this relation with the sampling interval of $\Delta t = 1$ ms, the continuous-time dynamics system can be converted to an equivalent DT nonlinear system as

$$x(k+1) = f'(x(k)) + g'(x(k))\, u \tag{104}$$

where (105) and (106), shown at the bottom of the page, hold, with cost-functional value in DT chosen as

$$J = \sum_{k=0}^{N} \Big((x(k) - x_d)^T Q (x(k) - x_d) + u(x(k))^T \mathrm{R} u(x(k)) \Big) \tag{107}$$

where $Q = 0.001 \times I^4$ and $R = 0.001 \times I^4$. The problem solution is almost the same as the linear system example except that we move the original point of axis to $x_d = [(pi/2) \quad 0 \quad 0 \quad 0]$ and use the new axis as $x_1'(k) = x_1(k) - (pi/2)$. The NN to approximate the GHJB equation is selected as polynomial functions for up to the fourth order of approximation, which means that $n = 4$ and $M = 4$. From (86), the NN can be constructed as

$$\begin{aligned} V(x_1', x_2, x_3, x_4) &= w_1 x_1'^2 + w_2 x_2^2 + w_3 x_3^2 + w_4 x_4^2 \\ &+ w_5 x_1' x_2 + w_6 x_1' x_3 + w_7 x_1' x_4 + w_8 x_2 x_3 + w_9 x_2 x_4 \\ &+ w_{10} x_3 x_4 + w_{11} x_1'^4 + w_{12} x_2^4 + w_{13} x_3^4 + w_{14} x_4^4 \\ &+ w_{15} x_1'^3 x_2 + w_{16} x_1'^3 x_3 + w_{17} x_1'^3 x_4 + w_{18} x_2^3 x_1' \\ &+ w_{19} x_2^3 x_3 + w_{20} x_2^3 x_4 + w_{21} x_3^3 x'^1 + w_{22} x_3^3 x_2 \\ &+ w_{23} x_3^3 x_4 + w_{24} x_4^3 x_1' + w_{25} x_4^3 x_2 + w_{26} x_4^3 x_3 \\ &+ w_{27} x_1'^2 x_2 x_3 + w_{28} x_1'^2 x_2 x_4 + w_{29} x_1'^2 x_3 x_4 + w_{30} x_2^2 x_1' x_3 \\ &+ w_{31} x_2^2 x_1' x_4 + w_{32} x_2^2 x_3 x_4 + w_{33} x_3^2 x_1' x_2 + w_{34} x_3^2 x_1' x_4 \\ &+ w_{35} x_3^2 x_2 x_4 + w_{36} x_4^2 x_1' x_2 + w_{37} x_4^2 x_1' x_3 + w_{38} x_4^2 x_2 x_3 \\ &+ w_{39} x_1' x_2 x_3 x_4. \end{aligned} \tag{108}$$

Associated gradient vector and Hessian matrix are derived as

$$\nabla V = = \left[\frac{\partial V}{\partial x_1}, \frac{\partial V}{\partial x_2}, \frac{\partial V}{\partial x_3}, \frac{\partial V}{\partial x_4} \right]^T$$

$$f(x) = \begin{bmatrix} x_3 \\ x_4 \\ \frac{-\left(2x_3x_4 + x_4^2 - x_3^2 - x_3^2 \cos x_2\right) \sin x_2 + 20 \cos x_1 - 10 \cos(x_1 + x_2) \cos x_2}{\cos^2 x_2 - 2} \\ \frac{\left(2x_3x_4 + x_4^2 + 2x_3x_4 \cos x_2 + x_4^2 \cos x_2 + 3x_3^2 + 2x_3^2 \cos x_2 + 20[\cos(x_1+x_2) - \cos x_1](1 + \cos x_2) - 10 \cos x_2 \cos(x_1 + x_2)\right)}{\cos^2 x_2 - 2} \end{bmatrix} \tag{100}$$

$$g(x) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{2 - \cos^2 x_2} & \frac{-1 - \cos x_2}{2 - \cos^2 x_2} \\ \frac{-1 - \cos x_2}{2 - \cos^2 x_2} & \frac{3 + 2\cos x_2}{2 - \cos^2 x_2} \end{bmatrix} \tag{101}$$

$$f'(x(k)) = \begin{bmatrix} 0.001 x_3(k) + x_1(k) \\ 0.001 x_4(k) + x_2(k) \\ \frac{-\left(2x_3x_4 + x_4^2 - x_3^2 - x_3^2 \cos x_2\right) \sin x_2 + 20 \cos x_1 - 10 \cos(x_1 + x_2) \cos x_2}{1000(\cos^2 x_2 - 2)} + x_3(k) \\ \frac{\left(2x_3x_4 + x_4^2 + 2x_3x_4 \cos x_2 + x_4^2 \cos x_2 + 3x_3^2 + 2x_3^2 \cos x_2\right) \sin x_2 + 20[\cos(x_1+x_2) - \cos x_1](1 + \cos x_2) - 10 \cos x_2 \cos(x_1 + x_2)}{1000(\cos^2 x_2 - 2)} + x_4(k) \end{bmatrix} \tag{105}$$

$$g'(x(k)) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{1000(2 - \cos^2 x_2(k))} & \frac{-1 - \cos x_2(k)}{1000(2 - \cos^2 x_2(k))} \\ \frac{-1 - \cos x_2(k)}{1000(2 - \cos^2 x_2(k))} & \frac{3 + 2\cos x_2(k)}{1000(2 - \cos^2 x_2(k))} \end{bmatrix} \tag{106}$$

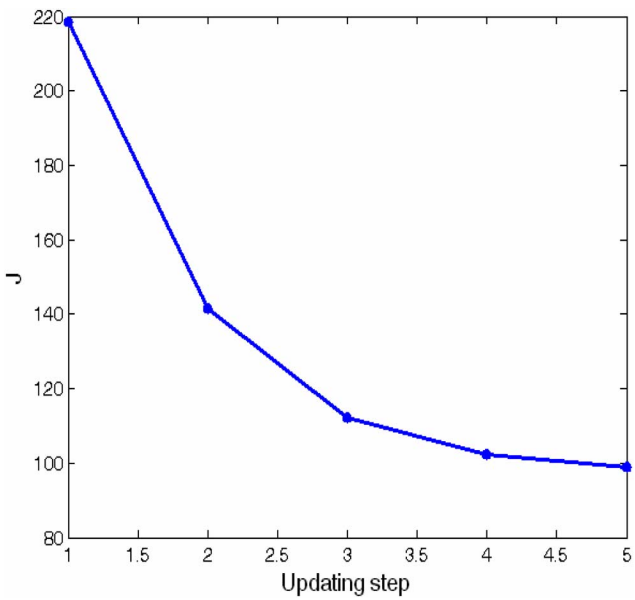

Fig. 12. Cost function at each updating step.

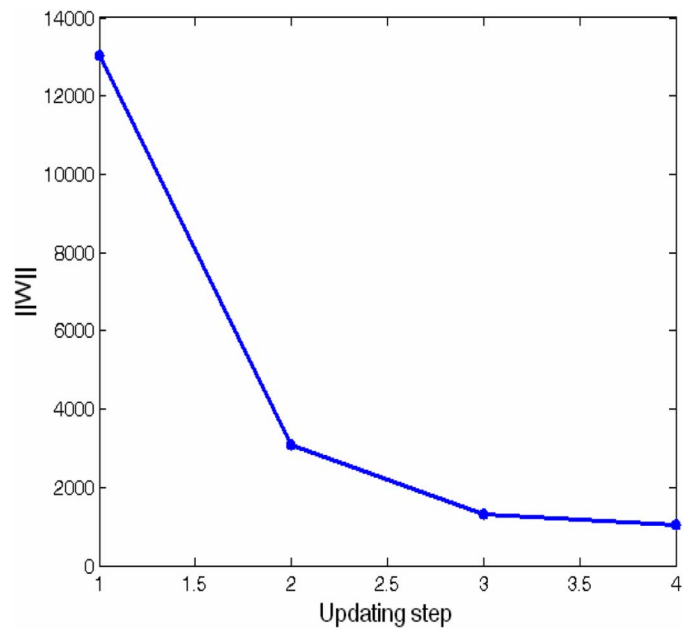

Fig. 13. Norm of the weights at each updating step.

$$\nabla^2 V = \begin{bmatrix} \frac{\partial^2 V}{\partial x_1^2} & \frac{\partial^2 V}{\partial x_1 \partial x_2} & \frac{\partial^2 V}{\partial x_1 \partial x_3} & \frac{\partial^2 V}{\partial x_1 \partial x_4} \\ \frac{\partial^2 V}{\partial x_3 \partial x_1} & \frac{\partial^2 V}{\partial x_2^2} & \frac{\partial^2 V}{\partial x_2 \partial x_3} & \frac{\partial^2 V}{\partial x_2 \partial x_4} \\ \frac{\partial^2 V}{\partial x_3 \partial x_1} & \frac{\partial^2 V}{\partial x_3 \partial x_2} & \frac{\partial^2 V}{\partial x_3^2} & \frac{\partial^2 V}{\partial x_3 \partial x_4} \\ \frac{\partial^2 V}{\partial x_4 \partial x_1} & \frac{\partial^2 V}{\partial x_4 \partial x_2} & \frac{\partial^2 V}{\partial x_4 \partial x_3} & \frac{\partial^2 V}{\partial x_4^2} \end{bmatrix}. \quad (109)$$

We select the initial admissible control law as

$$u_1 = [-500x_1' - 500x_3, \quad -200x_2 - 200x_4]^T. \quad (110)$$

Control function updating rule is taken as

$$u_k^{(i+1)} = -\left[g_k'^T \nabla^2 V_k^{(i)} g' + 2R\right]^{-1} \cdot g_k'^T \left(\nabla V_k^{(i)} + \nabla^2 V_k^{(i)} (f_k' - x(k))\right). \quad (111)$$

The u_i and V^i satisfy the GHJB equation

$$\begin{aligned} &\frac{1}{2}\left(f'(x) + g'(x)u^{(i)}(x) - x\right)^T \\ &\quad \cdot \nabla^2 V^{(i)}(x)\left(f'(x) + g'(x)u^{(i)}(x) - x\right) \\ &\quad + \nabla V^{(i)}(x)^T \left(f'(x) + g'(x)u^{(i)}(x) - x\right) \\ &\quad + x^T Q x + u^{(i)^{T\mathrm{Ru}^{(i)}}} = 0. \end{aligned}$$

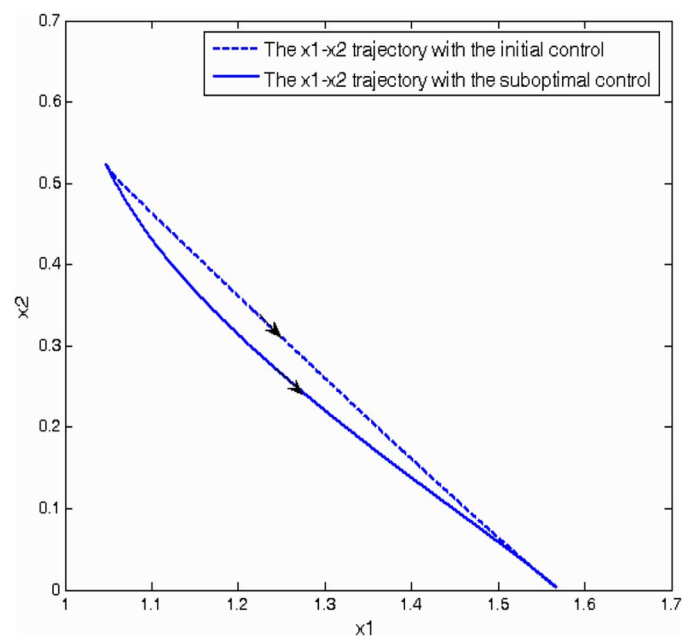

Fig. 14. State trajectory of (x_1, x_2).

In the simulation, the mesh size δx is selected as 0.2, the asymptotic stability region is chosen as $0 \le x_1 \le 2$, $-1 \le x_2 \le 1$, $-1 \le x_3 \le 1$, and $-1 \le x_4 \le 1$. The small positive constant is selected as $\varepsilon = 0.01$. The simulation steps are selected as $N = 2000$. We use the GHJB method to obtain the near-optimal control. After updated five times, the control has converged to the suboptimal control u^*. Fig. 12 shows the cost-functional value over updating step. On the other hand, Fig. 13 shows the norm of the NN weights at each updating step.

After we get the optimal control, we implement the initial admissible and suboptimal controls on the two-link planar robot arm system, respectively. Fig. 14 displays the state trajectory (x_1, x_2) with the initial admissible control and GHJB-based suboptimal control. Similarly, Fig. 15 illustrates the state trajectory (x_3, x_4) with initial admissible and GHJB-based suboptimal control. From these trajectory figures, we know that the robot arm has moved from the starting point to the final goal. On the other hand, Fig. 16 depicts the initial admissible control τ_1 and GHJB-based suboptimal control τ_1^* and Fig. 17 depicts the initial admissible control τ_2 and GHJB-based suboptimal control τ_2^*. Table IV shows that with different initial admissible controls, the converged norm of the NN weights and the suboptimal cost-functional values are almost close to each other. It is important to note that with different admissible control function values, the successive approximation-based updated controls will converge to a unique improved control and the im-

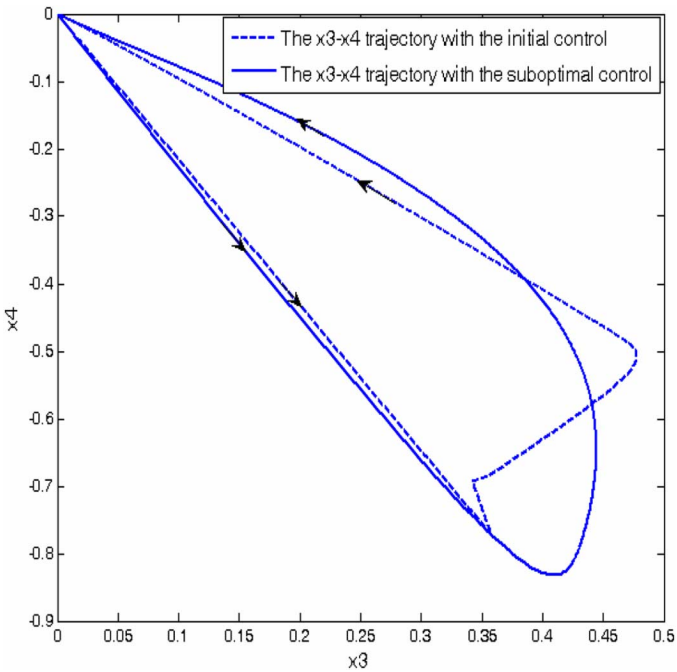

Fig. 15. State trajectory of (x_3, x_4).

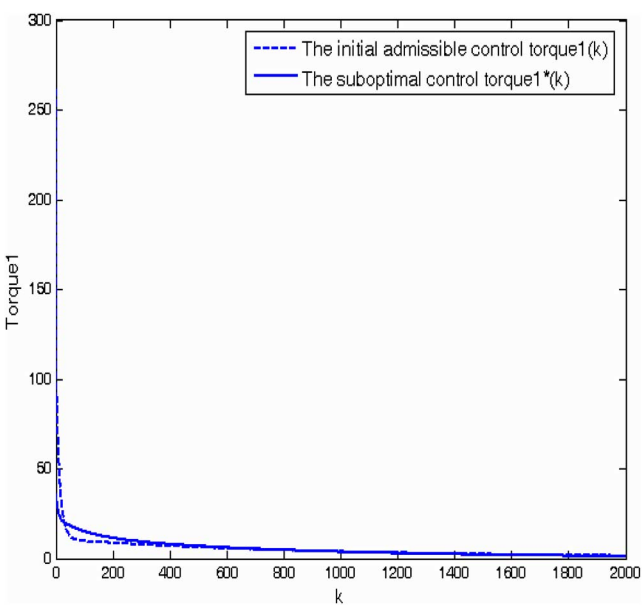

Fig. 16. Initial control τ_1 and suboptimal control τ_1^*.

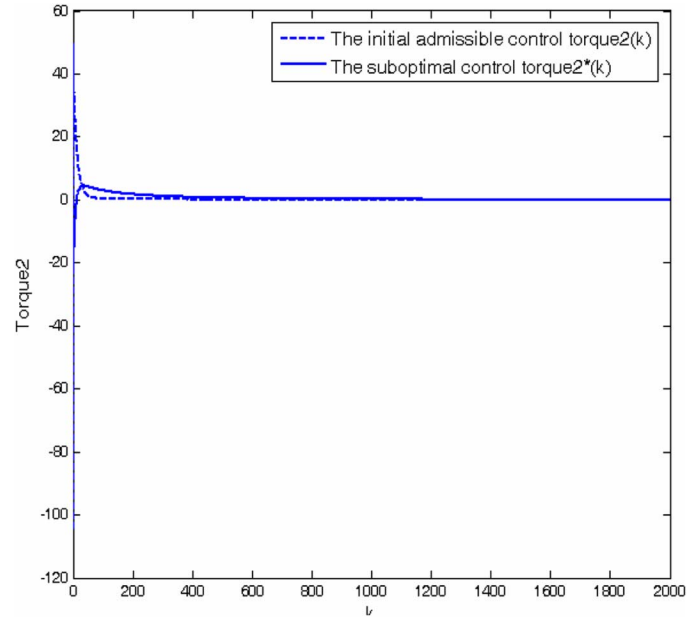

Fig. 17. Initial control τ_2 and suboptimal control τ_2^*.

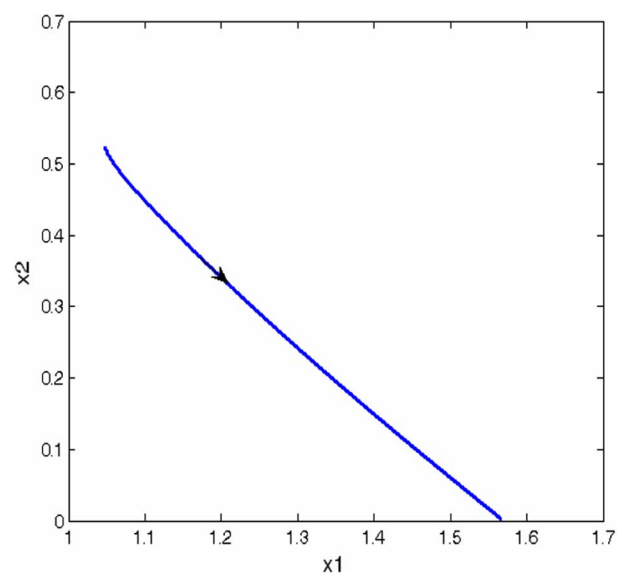

Fig. 18. State trajectory (x_1, x_2) with suboptimal control.

proved cost function values are almost the same. Since a small function approximation error value is used in solving the GHJB equation, the approximation-based GHJB solution renders the suboptimal control, which is quite close to the optimal control solution.

From Fig. 14, the trajectory with suboptimal control is a little longer than the trajectory with initial admissible control even though the cost-functional value with GHJB-based suboptimal control is significantly lower. This is due to the tradeoff observed between the trajectory selection and energy of the control input. The selection of the weighting Q and R matrices will dictate the selection. If we are more interested in perfect trajectory, we can select higher $||Q||$ or reduce $||R||$. If we are more interested in saving control energy, we can select lower $||Q||$ or increase $||R||$. For example, if we select $Q = 1000 \times I^4$ and $R = 0.001 \times I^4$, Figs. 18 and 19 show that the results obtained are different from those of Figs. 14, 16, and 17. It is important to note that the trajectory in Fig. 18 is close to a straight line but at the expense of the control input.

In Table IV, optimal cost values with different initial control are not exactly the same as those of the previous two examples, but are still reasonable due to the selection of the mesh size of 0.2. By decreasing the mesh size, one can increase the accuracy of convergence in the cost function. In the previous second-order system examples, the mesh size is selected as 0.01, which is quite small. However, in the fourth-order robot system,

TABLE IV
GHJB-BASED SOLUTION WITH ADMISSIBLE CONTROL

Initial control u_1	Optimal Cost Value J^*	$\|W^*\|$
$[-500x_1' - 500x_3, \quad -200x_2 - 200x_4]^T$	98.745	912.1
$[-200x_1' - 300x_3, \quad -200x_2 - 200x_4]^T$	97.726	928.32
$[-200x_1' - 400x_3 - 200x_4, \quad -200x_2 - 200x_4]^T$	97.252	999.09
$[-300x_1' - 400x_3, \quad -300x_2 - 300x_4]^T$	97.779	954.77
$[-300x_1' - 200x_3 - 200x_4, \quad -200x_1' - 200x_2 - 200x_3 - 200x_4]^T$	98.294	968.04

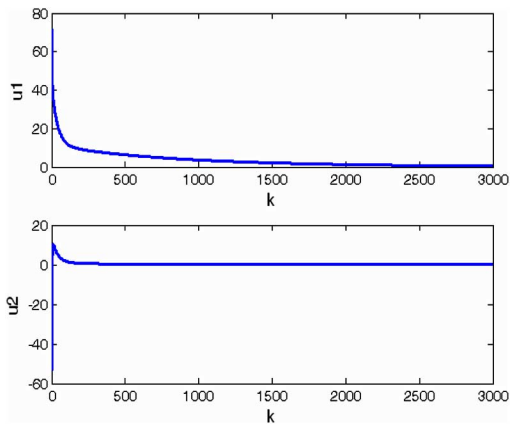

Fig. 19. Suboptimal control input.

a mesh size of 0.2 is chosen as a tradeoff between accuracy and computation. Decreasing the mesh size requires more memory to store the values due to an increase in computation.

V. CONCLUSION

In this paper, HJB, GHJB, and pre-Hamiltonian functions for the nonlinear DT system based on small perturbation assumption are introduced. A systematic method of obtaining the optimal control for general affine nonlinear DT system is proposed. Given an admissible control, the updated control through NN successive approximation of the GHJB equation renders an admissible control. For LQR problem, the updated control will converge to the optimal control. For nonlinear DT system, the updating control law will converge to an improved control, which renders a suboptimal control.

Future works will include improving NN approximation for value function, selecting better active functions, and reducing the computation complexity of NN. Further study will also focus on how to apply GHJB method to solve HJI equation in nonlinear DT system with uncertainties.

REFERENCES

[1] D. S. Bernstein, "Optimal nonlinear, but continuous, feedback control of systems with saturating actuators," *Int. J. Control*, vol. 62, no. 5, pp. 1209–1216, 1995.
[2] D. P. Bertsekas and J. N. Tsitsiklis, "Neuro-dynamic programming," in *Athena Scientific*. Belmont, MA: , 1996.
[3] F. Burk, *Lebesgue Measure and Integration*. New York: Wiley, 1998.
[4] C. Park and P. Tsiotras, "Approximations to optimal feedback control using a successive wavelet collocation algorithm," in *Proc. Amer. Control Conf.*, 2003, vol. 3, pp. 1950–1955.
[5] B. A. Finlayson, *The Method of Weighted Residuals and Variational Principles*. New York: Academic, 1972.
[6] F. L. Lewis, S. Jagannathan, and A. Yesilderek, *Neural Network Control of Robot Manipulator and Nonlinear Systems*. London, U.K.: Taylor & Francis, 1999.
[7] F. L. Lewis and M. Abu-Khalaf, "A Hamilton-Jacobi setup for constrained neural network control," in *Proc. Int. Symp. Intell. Control*, 2003, pp. 1–15.
[8] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal HJB solution for constrained input system using a neural network least-squares approach," in *Proc. 41st IEEE Conf. Decision Control*, 2002, vol. 1, pp. 943–948.
[9] F. L. Lewis, "Applied optimal control and estimation," in *Texas Instruments*. Upper Saddle River, NJ: Prentice-Hall, 1992.
[10] F. L. Lewis and V. L. Syrmos, *Optimal Control*. New York: Wiley, 1995.
[11] G. N. Saridis and C. S. Lee, "An approximation theory of optimal control for trainable manipulators," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 3, pp. 152–159, 1979.
[12] D. Han and S. N. Balakrishnan, "State-constrained agile missile control with adaptive critic based neural networks," *IEEE Trans. Control Syst. Technol.*, vol. 10, no. 4, pp. 481–489, Jul. 2002.
[13] H. Kawasaki and G. Li, "Gain tuning in discrete-time adaptive control for robots," in *Proc. SICE Annu. Conf. Fukui*, Aug. 4–6, 2003, pp. 1286–1291.
[14] D. Kleinman, "On a iterative technique for Riccati equation computations," *IEEE Trans. Autom. Control*, vol. AC-13, no. 1, pp. 114–115, Feb. 1968.
[15] S. E. Lyshevski, *Control Systems Theory with Engineering Applications*. Boston, MA: Birkhauser, 1990.
[16] W. T. Miller, R. Sutton, and P. Werbos, *Neural Networks for Control*. Cambridge, MA: MIT Press, 1990.
[17] M. Xin and S. N. Balakrishnan, "A new method for suboptimal control of a class of nonlinear systems," in *Proc. 41st IEEE Conf. Decision Control*, 2002, vol. 3, pp. 2756–2761.
[18] T. Parisini and R. Zoppoli, "Neural approximations for infinite-horizon optimal control of nonlinear stochastic systems," *IEEE Trans. Neural Netw.*, vol. 9, no. 6, pp. 1388–1408, Nov. 1998.
[19] R. W. Beard, G. N. Saridis, and J. T. Wen, "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, 1997.
[20] R. W. Beard and G. N. Saridis, "Approximate solutions to the time-invariant Hamilton-Jacobi-Bellman equation," *J. Optim. Theory Appl.*, vol. 96, no. 3, pp. 589–626, 1998.
[21] R. W. Beard, "Improving the closed-loop performance of nonlinear systems," Ph.D. dissertation, Electr. Eng. Dept., Rensselaer Polytech. Inst., Troy, NY, 1995.
[22] R. D. Abbott and T. W. McLain, "Validation of a synthesis application of the optimal control of an electro-hydraulic positioning system," in *Proc. Amer. Control Conf.*, 2000, vol. 6, pp. 4119–4123.
[23] R. Munos, L. C. Baird, and A. W. Moor, "Gradient descent approaches to neural-net-based solutions of the Hamilton-Jacobi-Bellman equation," in *Proc. Int. Joint Conf. Neural Netw.*, 1999, vol. 3, pp. 2152–2157.
[24] W. Lin and C. I. Brynes, "H_∞-control of discrete-time nonlinear systems," *IEEE Trans. Autom. Control*, vol. 41, no. 4, pp. , 494–510, Apr. 1996.
[25] J. Huang, "An algorithm to solve the discrete HJI equations arising in the L2 gain optimization problem," *Int. J. Control*, vol. 72, no. 1, pp. 49–57, 1999.
[26] J. Huang and C.-F. Lin, "A numerical approach to computing nonlinear H-infinity control laws," *J. Guid. Control Dyn.*, pp. 989–994, 2000.
[27] S. Ferrari and R. F. Stengel, "Model-based adaptive critic designs," in *Learning and Approximated Dynamic Programming*, J. Si, A. Barto, W. Powell, and D. Wunsch, Eds. New York: Wiley, 2004, ch. 3.

Zheng Chen (SM'04) was born in Hengyang, China, in 1977. He received the B.S. degree in electrical engineering and the M.S. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 1999 and 2004, respectively. Currently, he is working towards the Ph.D. degree at the Department of Electrical and Computer Engineering, Michigan State University, East Lansing.

He was a Senior Electronic Engineer from 2003 to 2004 in Shanghai, China. In 2004, he was a Research Assistant in Electrical and Computer Engineering, University of Missouri—Rolla. Currently, he is a Senior Research Assistant and the Laboratory Manager of Smart Microsystem Lab, Michigan State University. He currently holds two patents in process. His current interests include GHJB-based NN control, adaptive control, dynamic programming, modeling and control of electrical-active polymer, smart sensors and actuators in microelectrical-mechanic system and their biological and biomedical applications, control of dynamical systems with hysteresis, and electroactive polymer-based microrobots.

Mr. Chen received many scholarships and prizes during his undergraduate study. He received a Summer Dissertation Fellowship by the Graduate School at Michigan State University and Microsoft, Inc., in 2005.

Sarangapani Jagannathan (M'94–S'99) received the B.S. degree from College of Engineering, Guindy at Anna University, Madras, India, in 1987, the M.S. degree from the University of Saskatchewan, Saskatoon, Canada, in 1989, and the Ph.D. degree from the University of Texas at Arlington, in 1994, all in electrical engineering.

From 1986 to 1987, he was a Junior Engineer at Engineers India Limited, New Delhi, India. From 1990 to 1991, he was a Research Associate and Instructor at the University of Manitoba, Winnipeg Canada. From 1994 to 1998, he was a Consultant at Systems and Controls Research Division, Caterpillar Inc., Peoria. From 1998 to 2001, he was at the University of Texas at San Antonio. Since September 2001, he has been at the University of Missouri—Rolla, where currently, he is a Professor and Site Director for the National Science Foundation Industry/University Cooperative Research Center on Intelligent Maintenance Systems. He has coauthored more than 180 refereed conference and juried journal articles and several book chapters and three books entitled *Neural Network Control of Robot Manipulators and Nonlinear Systems* (Taylor & Francis: London, U.K., 1999), *Discrete-Time Neural Network Control of Nonlinear Discrete-Time Systems* (CRC Press: Boca Raton, FL, 2006), and *Wireless Ad Hoc and Sensor Networks: Performance, Protocols and Control* (CRC Press: Boca Raton, FL, 2007). He currently holds 17 patents and several are in process. His research interests include adaptive and NN control, computer/communication/sensor networks, prognostics, and autonomous systems/robotics.

Dr. Jagannathan received several gold medals and scholarships during his undergraduate program. He was the recipient of Region 5 IEEE Outstanding Branch Counselor Award in 2006, Faculty Excellence Award in 2006, St. Louis Outstanding Branch Counselor Award in 2005, Teaching Excellence Award in 2005, Caterpillar Research Excellence Award in 2001, Presidential Award for Research Excellence at UTSA in 2001, NSF CAREER award in 2000, Faculty Research Award in 2000, Patent Award in 1996, and Sigma Xi "Doctoral Research Award" in 1994. He has served and currently serving on the program committees of several IEEE conferences. He is an Associate Editor for the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON SYSTEMS ENGINEERING, and on several program committees. He is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi and IEEE Committee on Intelligent Control. He is currently serving as the Program Chair for the 2007 IEEE International Symposium on Intelligent Control, and Publicity Chair for the 2007 International Symposium on Adaptive Dynamic Programming.