Finite-Horizon Neural Network-based Optimal Control Design for Affine Nonlinear Continuous-time Systems

Qiming Zhao, Hao Xu, Travis Dierks and S. Jagannathan

Abstract—In this paper, the finite-horizon optimal control design for affine nonlinear continuous-time systems in the presence of known system dynamics is presented. A neural network (NN) is utilized to learn the time-varying solution of the Hamilton-Jacobi-Bellman (HJB) equation in an online and forward in time manner. To handle the time varying nature of the value function, the NN with constant weights and time-varying activation function is considered. The update law for tuning the NN weights is derived based on normalized gradient descent approach. To satisfy the terminal constraint and ensure stability, additional terms, one corresponding to the terminal constraint, and the other to stabilize the nonlinear system are added to the novel updating law. A uniformly ultimately boundedness of the non-autonomous closed-loop system is verified by using standard Lyapunov theory. The effectiveness of the proposed method is verified by simulation results.

Keywords: finite-horizon; neural network; optimal control; Hamilton-Jacobi-Bellman equation

I. INTRODUCTION

Optimal control problem of nonlinear systems has been one of the key focuses in control theory for decades. Traditional control theory [1] addresses optimal control problems in a backward-in-time manner. In the recent years, adaptive or neural network (NN) based optimal control has been intensely studied for both linear and nonlinear system over infinite-horizon with many applications [2][3][4]. However, the finite-horizon problem still remains an open problem for the control researchers.

In the finite-horizon case, the solution to the Hamilton-Jacobi-Bellman (HJB) equation is inherently time-varying and the system becomes non-autonomous. To properly satisfy the terminal constraint, the traditional techniques address the finite-horizon optimal control problem as obtaining the solution of the HJB equation in a backward-in-time manner [1]. However, forward-in-time finite-horizon optimal control solution poses a great challenge and still remains unresolved.

In the past literature, the author in [5] considered the finite-horizon problem by iteratively solving the so-called generalized-HJB (GHJB) equation using Galerkin method. The solution is obtained by solving an ordinary differential equation from the terminal time $t_f$ such that the terminal constraint is guaranteed to be satisfied. In [6], the authors proposed the fixed-final time optimal control laws using neural networks (NN) to solve the HJB equation for general affine nonlinear system.

In [8], the authors considered the input-constraint finite-horizon optimal control problem using off-line training scheme. The idea presented in [8] was essentially a dual heuristic programming (DHP) based scheme using value/policy iterations while the terminal constraint is satisfied by introducing the augmented vector incorporating the terminal value of the co-state $\lambda(N)$. The authors in [9] considered the discrete-time finite-horizon optimal problem under adaptive dynamic programming (ADP) scheme by using value and policy iterations. However, the terminal time is not specified and final state is fixed at the origin (i.e., $x_N = 0$).

The past literature [5][6][8] and [9] provides some insights into solving the finite-horizon problem whereas an algorithm which can be implemented online and forward-in-time manner is yet to be developed without using value/policy iterations so that the control can be implemented on hardware. Moreover, an initial admissible control requirement needs to be relaxed.

In contrast, in this paper, a new scheme is developed to solve the finite-horizon optimal regulation of an affine nonlinear continuous-time system in an online and forward-in-time manner. Two additional terms, one corresponding to the terminal constraint and the other for stabilizing the system are defined by extending [10] which in turn satisfies the terminal constraint and relaxes the need for an initial admission control. Lyapunov analysis is utilized to show the stability of our proposed scheme. In addition, in contrast to [8] and [9], the time-varying HJB equation is being solved approximately without performing value/policy iterations since the number of iterations needed for stability is not known.

II. BACKGROUND AND PROBLEM FORMULATION

In this paper, the finite-horizon optimal control for general affine nonlinear continuous-time systems is studied. Consider the system

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

where $x \in \mathbb{R}^n$ is the system state vector, $f(x) \in \mathbb{R}^n$, $g(x) \in \mathbb{R}^{m \times n}$ are inherent nonlinear dynamics and $u \in \mathbb{R}^m$ is
the control input vector. Here the input matrix \( g(x) \) is assumed to be bounded such that \( g_{\max} \leq \|g(x)\| \leq g_{\max} \).

The objective of the control design is to determine a feedback control policy that minimizes the value function
\[
V(x(t_0), t_0) = \psi(x(t_f), t_f) + \int_{t_0}^{t_f} L(x, u, t)dt
\]
subjected to the system dynamics (1). The function \( \psi(\cdot) \) penalizes the state \( x(t_f) \) at the terminal time \( t_f \) and \( L(x, u, t) \) is positive definite which penalizes the state \( x \) and the control input \( u \) from \( t_0 \) to \( t_f \). \( L(x, u, t) \) generally takes the form as \( L(x, u, t) = Q(x) + u^T Ru \) with \( Q(x) \geq 0 \), and \( R \) is a symmetric positive definite matrix with appropriate dimension.

Under the assumption that \( V(x, t) \in C^1 \), an infinitesimal equivalent to (2) is given by [1]
\[
-\frac{\partial V(x, t)}{\partial t} = L(x, u, t) + \frac{\partial V(x, t)}{\partial x} \left[ f(x) + g(x)u \right]
\]
(3)

By setting \( t_0 = t_f \), the terminal condition for the value function is given by
\[
V(x(t_f), t_f) = \psi(x(t_f), t_f)
\]
(4)

**Remark 1:** Equation (3) is a time-varying PDE with \( V(x, t) \) being the value function. For a given control input \( u \), \( V(x, t) \) can be solved backward-in-time from the terminal time \( t_f \) with the terminal condition \( V(x(t_f), t_f) \). For the infinite-horizon problem, in the case of linear regulation problem with quadratic cost (LQR), (3) reduces to algebraic Riccati equation while in the case of nonlinear system, (3) becomes a time-invariant PDE and several techniques are developed to solve the problem in a forward-in-time manner [4].

Next, define the Hamiltonian as
\[
H(x, u, V_s, t) = V_s + Q(x) + u^T Ru + V_s^T \left[ f(x) + g(x)u \right]
\]
(5)

where \( V_s = \frac{\partial V(x, t)}{\partial t} \) and \( V_s = -\frac{\partial V(x, t)}{\partial x} \). Note (5) has the time-dependency term \( V_s \) where the infinite-horizon case does not have. By using the traditional control theory [1], the optimal control policy is obtained by using the stationary condition \( \partial H(x, u, V_s, t)/\partial u = 0 \), which yields
\[
u^*(x(t), t) = -\frac{1}{2} R^{-1} g^T(x)V_s^*
\]
(6)

with \( V^*(x, t) \) being the optimal value function.

Substituting (6) into (3) yields the time-varying HJB equation given by
\[
V_s^* + V_s^T f(x) + Q(x) - \frac{1}{4} V_s^T g(x) R^{-1} g^T(x)V_s^* = 0
\]
(7)

It is well-known that the HJB equation (7) is both necessary as well as sufficient condition for optimality, and it provides the solution to fixed-final time optimal control problem for a general nonlinear continuous-time system in affine form. However, it is difficult or even impossible to obtain the analytical solution for the HJB equation. In the past literature, to find a solution to the HJB equation, value and policy iteration-based schemes [7][8][9] are generally utilized. In contrast, in this paper, a solution is found without utilizing the iterative approach as given next.

III. NEURAL NETWORK-BASED FINITE-HORIZON OPTIMAL CONTROL DESIGN

In this section, the finite-horizon optimal control scheme for affine nonlinear continuous-time systems with NN approximation is proposed. Traditionally, the adaptive critic based methodology uses two NNs, one for the value function, referred to as “critic” network, and the second for the control input, referred to “action” network, in order to generate the optimal control inputs. In this paper, the adaptive critic scheme is realized using only a single NN in an online fashion. To satisfy the terminal constraint while ensuring system stability, two additional terms are incorporated in deriving the novel NN weights update law. An error term corresponding to the terminal constraint is defined and minimized over time such that the terminal constraint can be properly satisfied. In addition, a stabilizing term is also added in the update law for guaranteeing the stability of the system so that an initial stabilizing control is not needed. The stability analysis of the closed-loop system is shown based on the Lyapunov stability theory and given in next section.

By approximation property of NNs, the time-varying value function \( V(x, t) \) can be represented by using a NN with time-varying activation function on a compact set \( S \) in the form
\[
V(x, t) = W^T e(t) \sigma(x) + \epsilon(x, t)
\]
(8)

with the terminal constraint represented as
\[
V(x_t, t_f) = W^T e(t_f) \sigma(x_t) + \epsilon(x_t, t_f)
\]
(9)

where \( W \in \mathbb{R}^L \) is the target NN weights vector with \( L \) being the number of hidden-layer neurons, \( e(t) \in \mathbb{R}^{L-1} \) is a bounded time-dependent matrix activation function, \( \sigma(x) \in \mathbb{R}^L \) is the bounded state-dependent vector activation function which is similarly defined in infinite-horizon case [2], \( \epsilon(x, t) \) is the NN reconstruction error. The target NN weights and reconstruction error are assumed to be upper bounded such that \( \|W\| \leq W_M \) and \( \|\epsilon(x, t)\| \leq \epsilon_M \), where \( W_M \) and \( \epsilon_M \) are positive constants [12]. In addition, it is assumed that the gradient of the NN reconstruction error with respect to \( x \) is upper bounded such that \( \|V_s \epsilon(x, t)\| \leq \epsilon_M \), where \( \epsilon_M \) is also a positive constant. \( \phi(t_f), \sigma(x_t) \) and \( e(x_t, t_f) \) have the same meaning but corresponding to the terminal time and state.
From the HJB equation (7), the partial derivative of $V(x,t)$ with respect to $x$ and $t$ are given by

$$V_x = \nabla_x \sigma(x) \phi^T(t) W + \nabla_x \epsilon(x,t)$$

$$V_t = W^T(t) \sigma(x) + \nabla_t \epsilon(x,t)$$

(10)

where $\nabla_x \sigma(x) = \frac{\partial \sigma(x)}{\partial x}$, $\nabla_t \epsilon(x,t) = \frac{\partial \epsilon(x,t)}{\partial t}$ and

$$\nabla_t \epsilon(x,t) = \frac{\partial \epsilon(x,t)}{\partial t}$$

Therefore, the optimal control input is finally given in terms of NN as

$$u^*(x(t),t) = -\frac{1}{2} R^{-1} g^T(x) \nabla_x \sigma(x) \phi^T(t) W - \frac{1}{2} R^{-1} g^T(x) \nabla_x \epsilon(x,t)$$

(11)

To find the Hamiltonian in terms of NN, substitute (11) into (5) to get

$$H(x,u,V_x,t) = W^T(t) \sigma(x) + Q(x) + \nabla_x \epsilon(x,t) D(x) \nabla_x \sigma(x) \phi^T(t) W + \epsilon_{NN}(x,t)$$

(12)

where $D(x) = g(x) R^{-1} g^T(x)$ and

$$\epsilon_{NN}(x,t) = \nabla_x \epsilon(x,t) + \frac{1}{2} W^T(t) \nabla_x \epsilon(x,t) D(x) \nabla_x \epsilon(x,t)$$

$$+ \frac{1}{4} W^T(t) \nabla_x \epsilon(x,t) D(x) \nabla_x \epsilon(x,t)$$

+ $Q(x) + \epsilon_{NN}(x,t)$

(13)

Note that by the assumption $g_{\min} \leq g(x) \leq g_{\max}$, $D(x)$ is also bounded such that we have $D_{\min} \leq ||D(x)|| \leq D_{\max}$.

Therefore, the HJB equation (7) can be represented as

$$HJB(\tilde{V}(x,t)) = W^T(t) \sigma(x) + W^T(t) \nabla_x \epsilon(x,t) f(x)$$

$$-1 \frac{1}{4} W^T(t) \nabla_x \epsilon(x,t) D(x) \nabla_x \epsilon(x,t) \phi^T(t) W$$

$$+ Q(x) + \epsilon_{NN}(x,t)$$

(14)

where

$$\epsilon_{NN}(x,t) = \nabla_x \epsilon(x,t) + \nabla_x \epsilon(x,t) f(x)$$

$$-1 \frac{1}{2} W^T(t) \nabla_x \epsilon(x,t) D(x) \nabla_x \epsilon(x,t) - \frac{1}{4} \nabla_x \epsilon(x,t) D(x) \nabla_x \epsilon(x,t)$$

is the residual error due to the NN reconstruction. The residual error will decrease when the number of neurons are increased such that $\epsilon \to 0$ as $L \to \infty$ [12].

To estimate the value function $V(x,t)$, define

$$\hat{V}(x,t) = \tilde{W}^T(t) \sigma(x)$$

(15)

The terminal constraint becomes

$$\hat{V}(x(t_f), t_f) = \tilde{W}^T(t_f) \sigma(\hat{x}_f)$$

(16)

where $\hat{V}(x,t)$ is the approximated value function, $\tilde{W} \in \mathbb{R}^n$ is the estimate of the target NN weights. $\hat{V}(x(t_f), t_f)$ is the approximated value function at the terminal time $t_f$, and $\sigma(\hat{x}_f)$ is the state-dependent activation function with approximated terminal state $\hat{x}_f$. Note that $\hat{x}_f$ is chosen as random as long as $\hat{x}_f$ is within the region of stability for a stabilizing control.

Using the NN approximation, the approximated Hamiltonian is then given by

$$\hat{H}(x,\hat{\theta},t) = \tilde{W}^T(t) \sigma(x) + Q(x) + \tilde{W}^T(t) \nabla_x \sigma(x) f(x)$$

$$-1 \frac{1}{4} \tilde{W}^T(t) \nabla_x \sigma(x) D(x) \nabla_x \sigma(x) \phi^T(t) \tilde{W}$$

(17)

Finally, the approximated control policy is given by

$$\hat{u}(x(t), t) = -\frac{1}{2} R^{-1} g^T(x) \nabla_x \sigma(x) \phi^T(t) \tilde{W}$$

(18)

Remark 2: Through the above analysis, it can be seen that the time-varying nature of the HJB solution is handled by the time-varying activation function $\phi(t)$ . This yields an essentially non-autonomous system and the proof of convergence will be more involved. In contrast, for infinite-horizon case, $\phi(t)$ is no longer needed, since the solution of the optimal control becomes time-invariant, hence the NN approximation does not have the time-dependency activation function [2] and [10].

Remark 3: It is observed from the definition (15) and the Hamiltonian approximation (17) that both the value function and the Hamiltonian become zero when $\|x\| = 0$. Hence, when the system state converges to zero, the value function approximation is no longer updated. This can be viewed as a persistency of excitation (PE) requirement for the inputs to the NN. Therefore, the system states must be persistently exciting long enough for the NN to learn the optimal value function. The PE condition is well known in adaptive control and can be satisfied by adding exploration noise or destabilizing controls [13]. In this paper, exploration noise is added to satisfy the PE condition. PE condition can be removed once the NN weights converge to their target values and the Hamiltonian converges to a small neighborhood around the origin.

To address the feature of the finite-horizon optimal control problem, not only the time-varying property should be kept in mind but also the terminal constraint needs to be taken care of in a proper manner. With NN approximation, define the terminal constraint error as

$$e_{t_f} = \psi(x(t_f), t_f) - \hat{V}(x(t_f), t_f) = \psi(x(t_f), t_f) - \tilde{W}^T(t_f) \sigma(\hat{x}_f) - \tilde{W}^T(t_f) \sigma(x(t_f))$$

where $\hat{\sigma}(x(t_f)) = \sigma(x(t_f)) = \sigma(\hat{x}_f)$. 

Our objective is to minimize the Hamiltonian (17) and the terminal constraint error (19) along the system trajectory, such
that the optimality can be achieved while satisfying the terminal constraint. Hence, define the total error term as
\[ e_{\text{total}} = \frac{1}{2} \left( H(x, \hat{W}, t) \right)^2 + \frac{1}{2} e_i^2. \] (20)

The update law for tuning the NN weights is found by minimizing (20) using normalized gradient descent algorithm as
\[ \dot{\hat{W}} = -\alpha_1 \frac{\partial}{\partial \hat{W}} H(x, \hat{W}, t) + \alpha_2 \frac{\partial}{\partial \hat{W}} \sum_{j} \gamma_j \sigma(x) D(x) J(x) + \dot{\hat{V}}(x, u) - \dot{\hat{V}}(x, u), \] (21)
where \( \dot{\hat{V}} = \phi(t) \sigma(x), \) with
\[ \dot{\hat{V}} = \phi(t) \sigma(x) + \phi(t) \nabla_{x} \sigma(x) f(x) - \frac{1}{2} \phi(t) \nabla_{x} \sigma(x) D(x) \nabla_{x} \sigma(x) \phi^T(t) \hat{W} \]
and the index term is defined as
\[ \dot{\hat{V}}(x, u) = \begin{cases} 0, & \text{if } \dot{J}(x) = J^T \dot{x} = J^T \dot{x}(f(x) + g(x) \dot{u}) < 0 \leq 1, \text{ otherwise} \end{cases} \]
where \( J(x) \) is the partial derivative of a Lyapunov candidate with respect to \( x \), defined as in Lemma 1 in next section.

**Remark 4:** In the update law (21), the first term aims to minimize the approximated Hamiltonian while the second term ensures that the terminal constraint error is also minimized over time. The last term assures that the system states remain bounded while the NN learns the value function. The index operator \( \Sigma(x, u) \) is defined based on Lyapunov sufficient condition for stability. It can be seen that the last term is removed when the system exhibits stable behavior, and activated when the system becomes unstable. This relaxes the requirement for an initial stabilizing control, in contrast to [2] and [7], where the admissible control is needed which is very difficult to select for an unknown dynamic system.

To complete this section, the flowchart of the proposed algorithm is shown as in Fig. 1.

**IV. STABILITY ANALYSIS**

In this section, the stability of the closed-loop system is analyzed. It will be shown that all the signals are uniformly ultimately bounded (UUB). Due to the time-varying solution to the HJB equation, the closed-loop system becomes non-autonomous [11] which requires special attention. This is fundamentally different from the proofs shown in [5][6][8][9].

Before proceeding, the following lemma is needed.

**Lemma 1**[10]: Consider the system (1) with value function (2) and the optimal control policy(6). Let \( J(x) \) be continuously differentiable, radially unbounded Lyapunov candidate such that \( \dot{J}(x) = J^T \dot{x}(f(x) + g(x) \dot{u}) < 0 \) with \( J(x) \) being the partial derivative of \( J(x) \) with respect to \( x \). In addition, let \( Q(x) \in \mathbb{R}^{n \times n} \) be a positive definite matrix, i.e., \( \forall x \neq 0, x \in \Omega \). \( \|Q(x)\| > 0 \), and \( x = 0 \Rightarrow \|Q(x)\| = 0 \), and \( \bar{Q} \) is bounded. Moreover, let \( Q(x) \) satisfy
\[ \lim_{x \to \infty} Q(x) = \infty \text{ as well as} \]
\[ V^T(x, u) J_s = r(x, u^*) = Q(x) + u^T R u^* \] (22)
Then, the following relation holds:
\[ J^T_s(f(x) + g(x) \dot{u}) = -J^T_s Q(x) J_s \] (23)

**Remark 5:** In [7], the closed-loop dynamics \( f(x) + g(x) \dot{u} \) are required to satisfy a Lipschitz condition such that
\[ \|f(x) + g(x) \dot{u}\| \leq K \text{ for a constant } K \text{ which is a stringent assumption}. \]
In contrast, the optimal closed-loop dynamics are assumed to be upper bounded by a function of the system states in this paper such that
\[ \|f(x) + g(x) \dot{u}\| \leq \delta(x) \] (24)
The general bound \( \delta(x) \) is taken as \( \delta(x) = \sqrt{\|J^T_s\| \|J_s\|} \) in this paper where \( \|J_s\| \) can be selected to satisfy the general bound and \( K^* \) is a constant.

Define \( \dot{W} = W - \hat{W} \), the approximated Hamiltonian can be expressed as
\[ H(x, \hat{W}, t) = -W^T \phi(t) \sigma(x) + W^T \phi(t) \sigma(x) - \hat{W}^T \phi(t) \nabla_{x} \sigma(x) f(x) \]
\[ + \frac{1}{4} W^T \phi(t) \nabla_{x} \sigma(x) D(x) \nabla_{x} \sigma(x) \phi^T(t) W \]
\[ \hat{W} + \frac{1}{2} \hat{W} \nabla_{x} \sigma(x) D(x) \nabla_{x} \sigma(x) \phi^T(t) \hat{W} \] (24)
Next, we will find the error dynamics for \( \dot{W} \). Observe that
\[ \dot{\hat{V}} = \phi(t) \sigma(x) + \phi(t) \nabla_{x} \sigma(x) \hat{u}(x) \]
(25)
where
\[ \dot{\hat{V}} = \frac{1}{2} R^{-1} \sigma(x) \phi(t)^T \hat{W} - \frac{1}{2} R^{-1} g^T(x) \nabla_{x} f(x) \]
Substituting \( \dot{u} \) into (25) yields
\[ \dot{\hat{V}} = \phi(t) \sigma(x) + \phi(t) \nabla_{x} \sigma(x) \phi(t)^T \hat{W} \]
(26)
where \( \psi = \dot{x}^T + \frac{1}{2} D(x) \nabla_x \psi(x,t) \).

Then the error dynamics are given as
\[
\dot{W} = -\alpha_1 \left( x^T \sigma(x) + D(x)^T \sigma(x) \right) + \alpha_2 \left( W^T \phi(t) + \phi^T(t) W + \epsilon_1 \right) + \epsilon_2 \left( x^T \sigma(x) + D(x)^T \sigma(x) \right) \dot{x} \sigma(t) + \dot{\sigma}(x,t) \end{equation}
\[
= -\alpha_1 \left( x^T \sigma(x) + D(x)^T \sigma(x) \right) + \alpha_2 \left( W^T \phi(t) + \phi^T(t) W + \epsilon_1 \right) + \epsilon_2 \left( x^T \sigma(x) + D(x)^T \sigma(x) \right) \dot{x} \sigma(t) + \dot{\sigma}(x,t) \end{equation}
\]

Next, the stability of the NN-based optimal control is examined as below.

**Theorem (NN-based scheme convergence to the HJB and the stability of the system):** Given the nonlinear system (1) with the target HJB equation (7). Let the update law be given by (21). Then, there exist positive constants \( b_{\epsilon_1} \) and \( b_{\epsilon_2} \) such that \( \| J_x \| \) and weights estimation error \( \| W \| \) are UUB.

**Proof:** Define the Lyapunov candidate function as
\[
J = \alpha_3 J(x) + \frac{1}{2} W^T \dot{W}
\]
Taking the derivative on both sides, we have
\[
\dot{J} = \alpha_3 \dot{J}(x) + \frac{1}{2} \dot{W}^T \dot{W}
\]
where \( J(x) \) and \( J_x(x) \) are given in the Lemma 1. By [11], the derivative of Lyapunov candidate function, which is essentially non-autonomous, is upper bounded by a time-invariant function. Therefore, \( J \) is negative definite provided the following conditions hold:
\[
\| J_x(x) \| \geq \max \left\{ \frac{\sigma_{\epsilon_1}^2}{\alpha_3 \hat{x}_{\min}}, \sqrt{\frac{2 \sigma_{\epsilon_1}^2}{\alpha_3 \hat{Q}_{\min}}} \right\} = b_{\epsilon_1}
\]
Or
\[
\| \dot{W} \| \geq \max \left\{ \frac{2(1 + \frac{\sigma_{\epsilon_1}^2}{\alpha_3 \hat{x}_{\min}}) \sigma_{\epsilon_1}^2}{\alpha_3 \hat{Q}_{\min}}, \frac{1 - \frac{\sigma_{\epsilon_1}^2}{\alpha_3 \hat{x}_{\min}}, \frac{1 - \frac{\sigma_{\epsilon_1}^2}{\alpha_3 \hat{Q}_{\min}}, \frac{1 - \frac{\sigma_{\epsilon_1}^2}{\alpha_3 \hat{x}_{\min}}}{\alpha_3 \hat{Q}_{\min}}} \right\} = b_{\epsilon_2}
\]

V. SIMULATION RESULTS

In this section, the proposed finite-horizon optimal regulation scheme is evaluated on a nonlinear system. Consider the second order nonlinear system
\[
f(x) = \begin{bmatrix} x_2 \\ -x_1 \left( \frac{\pi}{2} + \tan^{-1}(5x_1) \right) - \frac{5x_1^2}{2(1 + 25x_1^2)} + 4x_2 \end{bmatrix}
\]
and \( g(x) = [0, 3]^T \).

The performance index to be minimized is defined as
\[
V(x(t_0), t_0) = \psi(x(t_f), t_f) + \int_{t_0}^{t_f} Q(x) + u^T R u dt
\]
where the weighting matrices are selected to be \( Q(x) = x_1^2 \) and \( R = 1 \). The terminal penalty is selected to be \( \psi(x(t_f), t_f) = 2 \).

For the NN setup, the state-dependent basis function is constructed similarly as in the linear example. In our case, \( n = 2 \) and \( M = 6 \). This results in 15 neurons for the NN. The time-dependent activation function \( \phi(t) \in \mathbb{R}^{15 \times 15} \) is selected to be polynomials of time-to-go with saturation.

The design parameter is chosen to be \( \alpha_1 = 0.1, \alpha_2 = 0.15 \), and \( \alpha_3 = 0.001 \). The initial conditions for the system states are selected to be \( x_0 = [0.4, 0.12]^T \) while all the NN weights are initialized to be zeros.

![Fig. 2. System response with proposed controller design](image)

![Fig. 3. History of the NN weights](image)
The simulation results are shown next. First, the system response and the convergence of the NN weights are shown in Figs. 2 and 3, respectively. It can be seen from Figs. 2 and 3 that the system states converge close to the origin in less than 10 seconds and the NN weights converge as desired.

The HJB solution and the terminal constraint error are plotted in Figs. 4 and 5 to test the optimality of the system. It can be seen that the solution of the HJB equation converge close to zero in about 10 seconds. From Fig. 5, it has been shown that the terminal constraint error is minimized. The error for the terminal constraint for nonlinear system does not converge exactly to zero. Such result might be due to the selection of the NN activation function and its reconstruction error. A systematic approach for the selection of activation function might improve the convergence of the terminal constraint error and this will be relegated as part of future work.

VI. CONCLUSIONS

In this paper, the finite-horizon optimal control problem for the affine nonlinear system is proposed in the presence of known system dynamics. The terminal constraint error together with the approximation error is minimized. In addition, an extra stabilizing term is added to the update law to guarantee the stability of the system without an initial stabilizing control. The proposed algorithm yields a forward-in-time and online controller design scheme which enjoys great practical advantages. No iteration-based scheme is used. The boundedness of the closed-loop system is guaranteed by using Lyapunov stability analysis.

REFERENCES