Neural Network-based Finite-Horizon Approximately Optimal Control of Uncertain Affine Nonlinear Continuous-time Systems

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

Abstract—This paper develops a novel neural network (NN) based finite-horizon approximate optimal control of nonlinear continuous-time systems in affine form when the system dynamics are complete unknown. First an online NN identifier is proposed to learn the dynamics of the nonlinear continuous-time system. Subsequently, a second NN is utilized to learn the time-varying solution, or referred to as value function, of the Hamilton-Jacobi-Bellman (HJB) equation in an online and forward in time manner. Then, by using the estimated time-varying value function from the second NN and control coefficient matrix from the NN identifier, an approximate optimal control input is computed. To handle time varying value function, a NN with constant weights and time-varying activation function is considered and a suitable NN update law is derived based on normalized gradient descent approach. Further, in order to satisfy terminal constraint and ensure stability within the fixed final time, two extra terms, one corresponding to terminal constraint, and the other to stabilize the nonlinear system are added to the novel update law of the second NN. No initial stabilizing control is required. A uniformly ultimately boundedness of the closed-loop system is verified by using standard Lyapunov theory.

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

I. INTRODUCTION

Over the past few decades, optimal control of nonlinear systems has been one of the key topics for control researchers. Traditional control theory [10] addressed optimal control for a linear system in a backward-in-time manner by using Riccati equation when the system dynamics are known. Recently, adaptive or neural network (NN) based approximate optimal control has been intensely studied [6][7][14] for both linear and nonlinear system over infinite-horizon when system dynamics are uncertain. However, the finite-horizon problem still remains an open problem.

In the finite-horizon case, the solution to the Hamilton-Jacobi-Bellman (HJB) equation is inherently time-varying and the closed-loop system becomes non-autonomous. The traditional techniques [1] address the finite-horizon or fixed final time optimal control as obtaining the solution to the nonlinear partial differential HJB equation, in an iterative and backward-in-time manner [1], and possibly with offline training [5]. However, an iterative and backward-in-time solution cannot be implemented in real-time [7]. Forward-in-time finite horizon optimal control solution poses a great challenge and still remains unresolved.

In the past literature, the author [1] considered the finite-horizon problem by iteratively solving the so-called generalized-HJB (GHJB) equation using Galerkin method. The solution is attained by solving an ordinary differential equation from fixed final time \( t_f \) such that terminal constraint is guaranteed to be satisfied. In [3], authors proposed fixed final time optimal control laws using neural networks (NN) to solve the HJB equation for general affine nonlinear system. The NNs, which has a structure of time-dependent weights and state-dependent activation function, are used to approximately solve time-varying HJB equation using backward integration. Hence authors in [1] and [3] presented the finite-horizon optimal control problems in a backward-in-time manner.

In contrast, authors in [5] considered the input-constrained finite-horizon optimal control problem using off-line training scheme and dual heuristic programming (DHP) based scheme. The time-varying nature is handled by incorporating a constant NN weights and time-varying activation function. Moreover, value/policy iterations are utilized while the terminal constraint is satisfied by introducing a augmented vector incorporating terminal value of the co-state \( \lambda(N) \). In [4], the authors considered the discrete-time finite-horizon optimal problem under adaptive dynamic programming (ADP) scheme by using value and policy iterations. However, in that paper, the fixed final time is not specified and final state is fixed at the origin (i.e. \( x_f = 0 \) ). Although the past literature [1][3][5] and [5] provided some insights into solving finite-horizon optimal problem in an online and forward-in-time manner using value/policy iterations, an initial admissible control is needed [5-6]. In addition, an inadequate number of value/policy iterations can cause instability [7].

Motivated by the deficiencies aforementioned, in this paper, a new scheme is developed to solve the finite-horizon optimal regulation of a nonlinear continuous-time system in affine form in an online and forward-in-time manner. Two NN are introduced, one for estimating the control coefficient matrix, while the second NN for approximating the time-
varying value function which becomes the solution to the HJB equation. By using both the NNs, the optimal control input is derived.

Novel update laws for tuning the NN weights are derived. For the value function approximation, two extra terms, one corresponding to the terminal constraint and the other for stabilizing the nonlinear system are defined in the NN weight update by extending the work of [8]. The terminal constraint estimation error is minimized over time while the stabilizing term guarantees that the system will be stable during initial online NN learning phase so that an initial admissible control is not required. Moreover, Lyapunov analysis is utilized to show the overall stability of proposed scheme. Also, offline NN training and value/policy iterations are not utilized.

II. BACKGROUND AND PROBLEM FORMULATION

In this section, the optimal control of general nonlinear continuous-time systems in affine form is introduced. Consider the nonlinear system in affine form described by

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

(1)

where \( x \in \mathbb{R}^n \), \( u \in \mathbb{R}^m \) are the system state vector and control input and \( f(x) \in \mathbb{R}^n \), \( g(x) \in \mathbb{R}^{m \times n} \) are nonlinear dynamics which are considered unknown. Here the state vector is considered available for measurement and control coefficient input matrix \( g(x) \) is assumed to be bounded such that \( \|g(x)\| \leq g_{\text{max}} \). Without loss of generality, it is also assumed that the system is controllable in the sense that there exists a continuous control policy that stabilizes the system, with \( x = 0 \) being a unique equilibrium point on a compact set \( \Omega \subseteq \mathbb{R}^n \).

The objective of optimal control design is to determine a feedback control policy that minimizes value function

\[ V(x,t_0) = \psi(x(t_f), t_f) + \int_{t_0}^{t_f} r(x,u,t)dt \]  

(2)

subject to (1). The function \( \psi(\bullet) \) penalizes state \( x(t_f) \) at fixed final time \( t_f \) while \( r(x,u,t) \) penalizes the state \( x \) and the control input \( u \) from \( t_0 \) to \( t_f \). The cost-to-go function \( r(x,u,t) \) generally takes the form \( r(x,u,t) = Q(x,t) + u^T R u \) with \( Q(x,t) \) being a positive definite function of the state, while \( R \) is a symmetric positive definite matrix with appropriate dimension.

Under the assumption that \( V(x,t) \in C^4 \), an infinitesimal equivalent to (2) is given by [10]

\[- \frac{\partial V(x,t)}{\partial t} = r(x,u,t) + \frac{\partial V(x,t)}{\partial x}[f(x) + g(x)u] \]  

(3)

By setting \( t_0 = t_f \), the terminal constraint for the value function is given by

\[ V(x,t_f) = \psi(x(t_f), t_f) \]  

(4)

Remark 1: Equation (3) is a time-varying partial differential equation (PDE) with \( V(x,t) \) being the time-varying solution. For a given control input \( u \), \( V(x,t) \) can be solved backward-in-time from fixed final time \( t_f \) with the terminal constraint \( V(x,t_f) \). For infinite horizon problem, in the case of linear systems, (3) reduces to Algebraic Riccati Equation (ARE) while in the case of nonlinear system, (3) becomes a time-invariant PDE.

Next, define the Hamiltonian as

\[ H(x,u,t) = V_t + Q(x,t) + u^T R u + V^T f(x) + g(x)u \]  

(5)

where \( V_t = \frac{\partial V(x,t)}{\partial t} \) and \( V_x = \frac{\partial V(x,t)}{\partial x} \). Note (5) has the time-dependency term \( V_t \) while its infinite-horizon case does not have. By using [10], optimal control policy is attained by using stationary condition, \( \partial H(x,u,t)/\partial u = 0 \), which yields

\[ u^*(x,t) = -\frac{1}{2} R^{-1} g^T(x)V_x^* \]  

(6)

with \( V^*(x,t) \) being optimal time-varying value function.

Substituting (6) into (3) yields the time-varying HJB equation given by

\[ V_t^* + V_x^* f(x) + Q(x,t) - \frac{1}{4} V_x^* g(x) R^{-1} g^T(x)V_x^* = 0 \]  

(7)

It is well-known that the HJB equation (7) is both necessary and sufficient condition for optimality, and solving the HJB equation provides a solution to the fixed final-time optimal control of a general nonlinear continuous-time system in affine form. However, it is difficult or even impossible to obtain the analytical solution of HJB equation (7) even while system dynamics are known. For a linear system with quadratic cost (LQR) function, (7) becomes Riccati equation (RE) which can be solved by integrating from terminal time \( t_f \) provided the system dynamics are known. Moreover, to find a solution of HJB equation, value and/or policy iteration-based schemes [5-6] are generally utilized when the system dynamics are uncertain. However, since inadequate iterations within a sampling interval can lead to instability [7], in this paper, a solution is developed without utilizing the iterative approach and system dynamics as given next.

III. NEURAL NETWORK-BASED APPROXIMATELY OPTIMAL CONTROLLER DESIGN

In this section, the finite-horizon approximate optimal control scheme for the affine nonlinear continuous-time system with NN approximation is proposed. First, we introduce a novel online NN identifier to learn system dynamics. Then, another NN is utilized to learn the time-varying value function within fixed final time. Then, by using estimated value function and identified system dynamics, finite horizon approximate optimal control of nonlinear continuous-time system is addressed.

A. NN-identifier design

In the recent NDP literatures [1-5], system dynamics (e.g. \( f(\bullet), g(\bullet) \)) are required for developing optimal control of nonlinear continuous-time systems in affine form. However, system dynamics are generally uncertain. To circumvent this issue, a novel NN-identifier is developed as follows.
Recalling the NN universal function approximation property and [10], the actual nonlinear continuous system can be represented on a compact set as

\[ f(x) = W_f \sigma_f(x) + e_f, \quad g(x) = W_g \sigma_g(x) + e_g \]  
(8)

where \( W_f \in \mathbb{R}^{1 \times n_f} \), \( W_g \in \mathbb{R}^{1 \times n_g} \) denote NN target weight matrices, \( \sigma_f(x) \in \mathbb{R}^{n_f} \), \( \sigma_g(x) \in \mathbb{R}^{n_g} \) represent activation functions and \( e_f, e_g \in \mathbb{R}^{1 \times n} \) denote NN reconstruction errors respectively.

Then, the nonlinear continuous-time system (1) in affine form can be represented by using (8) as

\[
\dot{x} = f(x) + g(x)u = \begin{bmatrix} W_f^T & 0 \\ W_g^T & \sigma_f(x) & 0 \\ \sigma_g(x) & 1 \end{bmatrix} + e_f + e_g u
\]  
(9)

where \( W_f = [W_f^T, W_g^T]^T \in \mathbb{R}^{2(n_f+n_g)}, \sigma_f(x) = \text{diag}\{\sigma_f(1), \ldots, \sigma_f(n_f)\} \in \mathbb{R}^{2(n_f+n_g)} \) are NN identifier target weight and activation function, \( \bar{u} = [1 u^T]^T \in \mathbb{R}^{n+1} \) is augment control input, and \( e_i = e_f + e_g u \) is NN identifier reconstruction error. It is important to note that since NN-identifier activation function \( \sigma_f(\bullet), \sigma_g(\bullet) \) and \( \sigma_f(\bullet) \) are known, the nonlinear system dynamics can be approximated provided the NN identifier weight matrix \( W_f \) is tuned properly. Next, a suitable update law for the online NN identifier will be derived.

First, consider the following state estimator or identifier

\[ \dot{x} = \hat{W}_f^T \sigma_f(x)\bar{u} + K\hat{x} \]  
(10)

with \( \hat{W}_f \in \mathbb{R}^{2(n_f+n_g)} \) is the estimated NN identifier weight matrix, \( \hat{x} = x - \hat{x} \in \mathbb{R}^n \) is the state estimation/identification error and \( K \) is a design parameter to maintain stability of the NN identifier.

Recalling equations (9) and (10), the dynamics of state estimation error can be expressed as

\[ \dot{\hat{x}} = \hat{x} = \hat{W}_f^T \sigma_f(x)\bar{u} + e_i - K\hat{x} \]  
(11)

To force estimated NN identifier weight matrix converge close to its target, \( W_f \), within fixed final time, update law for \( \hat{W}_f \) can be selected as

\[ \dot{\hat{W}}_f = -\alpha_i \hat{W}_f + \sigma_f(x)\bar{u} \]  
(12)

where \( \alpha_i \) is tuning parameter of the NN identifier satisfying \( \alpha_i > 0 \). Since \( \hat{W}_f = -\dot{\hat{W}}_f \), the dynamics of NN identifier weight estimation error can be represented as

\[ \dot{\hat{W}}_f = \alpha_i \hat{W}_f - \sigma_f(x)\bar{u} \]  
(13)

Next, the stability of the NN state estimation error (11) and weight estimation error (13) will be demonstrated as follows.

**Theorem 1 (Boundedness of the NN identifier):** Let the initial NN identifier weight \( W_f \) be residing in a compact set, and proposed NN identifier and its NN weight update law be given as (10) and (12) respectively. In the presence of bounded inputs, there exists a positive tuning parameter \( \alpha_i > 0 \) such that the identification (11) and weight estimation errors \( \hat{W}_f \) are all uniformly ultimately bounded (UUB) within fixed final time.

**Proof:** Omitted due to space limitation.

**Remark 2:** In the above theorem, the input to the nonlinear system is considered to be bounded only for identification. However, this assumption is relaxed when identifier is utilized in conjunction with control which is given in the next section.

**B. Adaptive NN approximate optimal control design**

Generally, traditional adaptive critic based schemes utilize two NNs, one for the value function, referred to as “critic” network, and the second for the control input, referred as “actor” network, to solve approximate optimal control inputs [7]. However, in this paper, the adaptive critic scheme is realized using only a single NN in an online fashion. To guarantee the terminal constraint while maintaining system stability, two extra terms are incorporated in the novel NN weight update law. An estimation error term corresponding to terminal constraint is introduced and minimized. Moreover, a stabilizing term is added into the update law for ensuring the stability such that an initial stabilizing control is not needed during the initial NN tuning phase of the value function. The overall stability analysis of the closed-loop system incorporating the identifier and the optimal controller is demonstrated based on the Lyapunov stability theory.

**Lemma 1:** Let the time dependent value function, \( V(x,t) \), be smooth and uniformly continuous in a compact set \( \Omega \subseteq [0,\infty) \times \mathbb{R}^{L \times 1} \). Then, there exists a neural network \([5][14]\) such that smooth and uniformly continuous time dependent value function \( V(x,t) \) can be approximated with constant weights and time-dependent activation function, i.e.

\[ V(x,t) = W_f^T \phi(x,t-t_i) + e_v(x,t) \]  
(14)

where \( W_f \in \mathbb{R}^{L} \) is the target NN weight vector with \( L \) being the number of hidden-layer neurons, \( \phi(x,t-t_i) \in \mathbb{R}^{L \times 1} \) is bounded time-dependent activation function, and \( e_v(x,t) \) is the NN reconstruction error.

**Proof:** Omitted due to space limitation.

Moreover, the terminal constraint can be expressed as

\[ V(x,t_f) = W_f^T \phi(x(t_f),0) + e_v(x,t_f) \]  
(15)

**Remark 3:** The target NN weights \( W_f \) and reconstruction error \( e_v(x,t) \) are assumed to be bounded above such that

\[ \|W_f\| \leq W_{fM} \quad \text{and} \quad \|e_v(x,t)\| \leq e_{vM}, \]  
where \( W_{fM} \) and \( e_{vM} \) are positive constants [11]. Similar to [7], it is assumed that the gradient of the NN reconstruction error with respect to \( x \) is bounded above such that

\[ \|\nabla x e_v(x,t)\| \leq e_{vM}, \]  
where \( e_{vM} \) is also a positive constant. The quantities \( \phi(x(t_f),0) \) and \( e_v(x,t_f) \) correspond to fixed final time and state. Note that
NN representation has a different form when compared to \[3\] and \[5\]. However, the different forms of representation enjoy same time-dependency property, which is essential to approximate time-varying value function \(V(x,t)\).

From the HJB equation (7), the partial derivative of \(V(x,t)\) with respect to \(x\) and \(t\) can be represented as
\[
V_x = V_{x}^{T} \phi_t(x,t) + V_{x} e_t(x,t) + \frac{\partial}{\partial t} V(x,t),
\]
where \(\phi_t(x,t) = \frac{\partial}{\partial t} \phi(x,t)\) and \(e_t(x,t) = \frac{\partial}{\partial t} e_t(x,t)\).

Therefore, the optimal control input is given in terms of value function NN estimate as
\[
u = -\frac{1}{2} R^{-1} g^T \frac{\partial V}{\partial x}(x)
\]
(17)
\[
= -\frac{1}{2} R^{-1} g^T \frac{\partial \phi_t}{\partial x}(x) W_t - \frac{1}{2} R^{-1} g^T \frac{\partial e_t}{\partial x}(x)
\]
To obtain the Hamiltonian in terms of NN, substitute (17) into (5) to get
\[
H(x,u,t) = W_t^T \nabla \phi(x,t) + Q(x,t) + W_t^T \nabla \phi(x,t) - f(x)
\]
where \(D(x) = g(x)R^{-1}g^T(x)\) and
\[
e_t(x,t) = \nabla \cdot e_t(x,t) + \frac{1}{2} W_t^T \phi_t(x,t) \nabla \cdot e_t(x,t) D(x) \nabla \cdot e_t(x,t)
\]
(18)
\[
= \frac{1}{2} \nabla \cdot e_t(x,t) D(x) \nabla \cdot e_t(x,t) + \frac{1}{4} W_t^T \phi_t(x,t) D(x) \nabla \cdot e_t(x,t) + \frac{1}{4} W_t^T \phi_t(x,t) D(x) \nabla \cdot e_t(x,t)
\]
Note that by the assumption \(\|g(x)\| \leq g_{\text{max}}\), \(D(x)\) is also bounded such that we have \(\|D(x)\| \leq D_{\text{max}}\).

Therefore, the HJB equation (7) can be represented as
\[
HJB(V(x,t)) = W_t^T \nabla \phi(x,t) + Q(x,t) + W_t^T \nabla \phi(x,t) - f(x)
\]
where \(e_t(x,t) = \nabla \cdot e_t(x,t) + \frac{1}{2} W_t^T \phi_t(x,t) D(x) \nabla \cdot e_t(x,t)
\)
(19)
\[
x \nabla \cdot e_t(x,t) - \frac{1}{4} W_t^T \phi_t(x,t) D(x) \nabla \cdot e_t(x,t) - \frac{1}{2} W_t^T \nabla \phi(x,t) D(x)
\]
is residual error due to NN reconstruction.

To estimate the value function, \(V(x,t)\), define
\[
\hat{V}(x,t) = \hat{W}_t^T \phi(x,t) - f(x)
\]
(20)
The terminal condition becomes
\[
\hat{V}(x,t_f) = \hat{W}_t^T \phi(x,t_f) - f(x)
\]
(21)
where \(\hat{V}(x,t)\) is the approximated value function, \(\hat{W}_t \in \mathbb{R}^{x \times t}\) is the estimate of the target NN weights, \(\hat{V}(x,t_f)\) is the approximated value function at the terminal time \(t_f\), and \(\phi(\hat{x}(t_f),0) \in \mathbb{R}^{x+1}\) is the activation function with approximated terminal state \(\hat{x}_f\). Note that term \(\hat{x}_f\) can be calculated by using identified system dynamics and current states.

Using the NN approximation, the Hamiltonian can be estimated as
\[
\hat{H}(x,u,t) = \hat{W}_t^T \nabla \phi(x,t) + \hat{W}_t^T \nabla \phi(x,t) - \hat{f}(x)
\]
(22)
\[
-\frac{1}{4} \hat{W}_t^T \nabla \phi(x,t) - \hat{D}(x) \hat{W}_t^T \phi(x,t) - \hat{f}(x)
\]
where \(\hat{D}(x) = g^T(x)R^{-1}g(x)\) is attained from the NN identifier. Finally, the estimated control policy is given by
\[
\hat{u}(x,t) = -\frac{1}{2} R^{-1} g^T(x) \nabla \phi(x,t) - \hat{f}(x)
\]
(23)

**Remark 4:** Observe from the definition (21) and the Hamiltonian approximation (22) that both the value function and the Hamiltonian become zero when \(\|x\| = 0\). Hence, when the system states converge to zero, the value function approximation is no longer updated. This can be viewed as a persistence 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 a perturbation term [12] into control input. In this paper, exploration noise is added to control signal for satisfying the PE condition. The PE condition can be removed once the NN weights converge close to their target values and Hamiltonian converges to a small neighborhood around the origin.

To address finite-horizon optimal control problem, both time-varying nature of value function and terminal constraint need to be taken care of in a proper manner. With NN approximation, define terminal constraint estimation error as
\[
e_t = \psi(x(t_f),t_f) - \hat{W}_t \phi(\hat{x}(t_f),t_f)
\]
\[= W_t \phi(x(t_f),t_f) + \hat{W}_t \phi(\hat{x}(t_f),t_f) + e_t
\]
(24)
where \(\phi(\hat{x}(t_f),t_f) = \phi(x(t_f),t_f) - \phi(\hat{x}(t_f),t_f)\).

Our objective is to minimize the Hamiltonian (22) and the terminal constraint error (24) along the system trajectory, such that the optimality can be achieved while satisfying the terminal constraint. Hence, define the total error as
\[
e_{\text{total}} = \frac{1}{2} \left( H(x,u,t) \right)^2 + \frac{1}{2} e_t^2
\]
(25)
The update law for tuning the NN weights is found by minimizing (25) using normalized gradient descent as
\[
\hat{W}_t = -\alpha_t \frac{\hat{\omega}}{(1 + \hat{\omega}^2)^{1/2}} H(x,u,t) + \alpha_t \frac{\hat{\zeta}}{(1 + \hat{\zeta}^2)^{1/2}} e_t
\]
(26)
\[
+ \frac{\alpha_t}{2} \nabla \phi(x,t) \hat{D}(x) Q(x,t)
\]
where \(\hat{\zeta} = \phi(\hat{x}(t_f),t_f)\), and \(\hat{\omega} = \nabla \phi(x,t) + \nabla \phi(x,t) \hat{f}(x)\)
Remark 5: In the update law (26), the first term aims to minimize the approximated Hamiltonian while the second term ensures that terminal constraint estimation error is also minimized. It is the first time that the second term in (26) has been proposed whereas first term in (26) has been utilized in recent literature [7][14] while the last term is inspired from [7] that assures that the system states remain bounded when the NN learns the value function during the initial phase. It can be seen that the last term relaxes the requirement for an initial stabilizing control, in contrast to [2], where an initial 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.

In addition, let $R \in \mathbb{R}^{m \times m}$ and $Q(x) \in \mathbb{R}$ be a positive definite matrix and function respectively, i.e., $R > 0$, $\forall t \in [0, t_f]$ $\forall x \neq 0, x \in \Omega$, $\|Q(x, t)\| > 0$, and $x = 0 \Rightarrow [Q(x, t)] = 0$. Moreover, let $Q(x, t)$ satisfy $\lim_{t \to \infty} Q(x, t) = \infty$ as well as

$$\nabla V_t(x, t) + \nabla V_t(x, t)(f(x) + g(x)u^*) = -Q(x, t) - u^TRu^*$$

(27)

Further, the following relation holds

$$\nabla V_t(x, t) + \nabla V_t(x, t)(f(x) + g(x)u^*) < -Q(x, t)$$

(28)

Proof: Omitted due to space limitation.

Then, define the weight estimation error of the NN utilized for control as $\hat{W}_e = W - \hat{W}_v$. Note that $\dot{\hat{W}}_e = \dot{\hat{W}}_v$. The approximated Hamiltonian can be expressed as

$$\hat{V}(x, u, t) = -\hat{W}_v^T\nabla_x \phi(x, t) - \hat{W}_v^T\nabla_u \phi(x, t) - \hat{f}(x) - W_t^T \nabla \phi(x, t)$$

(29)

Theorem 2 (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 NNs update law for the identifier and the controller be given by (12), and (26) respectively while the estimated control input is given as (23). Then, there exist positive constants $b_{1g}$, $b_{2g}$ and $b_{nt}$ such that $\|Q(x, t)\|$, identification error $\|\hat{\epsilon}\|$ and NN weights estimation errors for the NN identifier and the controller (i.e., $\|\hat{W}_e\|$ and $\|\hat{W}_v\|$) are UUB respectively.

Proof: Omitted due to space limitation.

V. SIMULATION RESULTS

Example 1: Consider the second order affine nonlinear continuous-time system given by [8]

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

(30)

with internal dynamics $f(x)$ and $g(x)$ are given as

$$f(x) = \begin{bmatrix} -x_1 + \tan^{-1}(5x_1) - 5x_2^2/2 + 4x_2 \\ 2x_1 \end{bmatrix}, g(x) = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

(31)

with $x = \begin{bmatrix} x_1^T x_2^T \end{bmatrix}^T \in \mathbb{R}^2$ and $u \in \mathbb{R}$. Further, the performance index which needs to be minimized is defined as

$$V(x, t) = \psi(x(t), t) + \int_0^t Q(x) + u^TRu \, dt$$

(32)

where $Q(x)$, $R$ are selected as $Q(x) = (x_1^2 + x_2^2)/2$ and $R = 1$, and terminal constraint is given as $\psi(x(t), t) = 2$.

Moreover, the initial system state is selected as $x_0 = [0.4 0.12]^T$, the NN identifier activation function is given as $\begin{bmatrix} 1, x_1^2, x_2 x_2, \ldots, x_1^6, x_2^6, x_1 x_2, \ldots, x_1^6 x_2^6 \end{bmatrix}^T \in \mathbb{R}^6$, state dependent part of basis function for estimating value function is constructed from the expansion of polynomial

$$\sum_{j=0}(\sum_{i=0}^n x_i)^{2j} \in \mathbb{R}^3$$

with $n = 2$ and $M = 6$, and time
dependent part is selected as saturating polynomial time function
\( sat(t_f-t)^{14},(t_f-t)^{13},...,1,(t_f-t)^{14},...,t_f-t \) (i.e.,
(0,1000] denotes saturation range. Note that the saturation for time function is included to ensure the magnitude of time function stays within a reasonable range such that the NN weights are computable.

The tuning parameters are selected to be
\( \alpha_1 = 0.1, \alpha_2 = 0.15, \alpha_3 = 0.001 \) and \( \alpha_4 = 0.5 \) while value function estimation NN weights are initialized to be all zero and NN identifier weights are selected inside \( W_{t0} \in [0,1] \) randomly. Results are shown in Figures 2, 3 and 4.

First, as shown in Figure 2, the proposed finite horizon optimal control can force the nonlinear continuous-time system state vector converge close to zero within finite horizon or the proposed scheme can maintain the UUB even in presence of uncertain system dynamics.

Next, the HJB equation and terminal constraint errors have been analyzed. In Figure 3 and 4, within the finite time (i.e. \( t \in [0,50] \)), not only HJB equation error but also terminal constraint errors converge close to zero. According to Theorem 2, proposed finite horizon optimal design can ensure the UUB of both HJB equation error and terminal constraint error within finite horizon. Moreover, the convergence of the HJB and terminal constraint errors confirm that the approximated control input approaches the finite horizon optimal control input over finite time. According to Figures 2 to 4, the proposed scheme ender nearly the same performance as offline traditional optimal design requiring the knowledge of system dynamics.

VI. CONCLUSIONS

In this paper, the finite-horizon approximate optimal control of general nonlinear continuous-time system in affine form is proposed in the presence of unknown system dynamics. The terminal constraint estimation error together with HJB 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 NN identifier generates the control coefficient matrix which is subsequently utilized in the controller design. The proposed scheme yields a forward-in-time and online controller design scheme which enjoys great practical benefits. No iteration-based scheme is used and offline training is not utilized. The boundedness of the closed-loop system is guaranteed by using Lyapunov stability analysis. Simulation results verify the theoretical claims.

REFERENCES