Value Iteration Based Continuous-time Nonlinear Constrained Optimal Tracking Controller Design

1st Geyang Xiao
Research Center for Industrial Internet
Zhejiang Lab
Hangzhou, China
xgyalan@outlook.com

2nd Boyang Zhou
Research Center for Industrial Internet
Zhejiang Lab
Hangzhou, China
zhouby@zhejianglab.com

3rd Kaiyi Lou
Research Center for Industrial Internet
Zhejiang Lab
Hangzhou, China
loukaiyi@foxmail.com

4th Zhengrong Chen
Research Center for Industrial Internet
Zhejiang Lab
Hangzhou, China
aaronchen95@hotmail.com

2020 Chinese Automation Congress (CAC) | 978-1-7281-7687-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/CAC51589.2020.9327451

***Abstract*—The value iteration mechanism is introduced to design the constrained optimal tracking controller for continuous-time nonlinear system in this paper. Unlike most studies were developed by policy iteration, the proposed VI algorithm is designed to be started with arbitrary semi-definite positive function, and thus no restriction is in the initial condition. The iterative property of the proposed algorithm is discussed and the optimality of the iterated results are obtained. To verify the effectiveness, a nonlinear case is simulated.**

***Index Terms*—Constrained control, adaptive dynamic programming, value iteration**

I. Introduction

In the last decades, a framework called ADP (adaptive dynamic programming) has became an effective way to design nonlinear optimal controller [1] and attracted lots of attentions in recent years [2]–[7]. Generally, policy iteration (PI) and value iteration (VI) were two typical iterative mechanisms in the framework [1]. PI methods suffered from the admissibility of the initial control and attempted to solve the value function according to the admissible control [8]. In contrast, VI methods can seek the optimal solution without the above restriction.

For the nonlinear optimal tracking controller design, most works [2]–[7] were developed by PI. Generally, the optimal tracking controller was designed by combining the feedback part and the steady-state part [2], where the steady-state part was derived by system dynamics and desired trajectory dynamics and the feedback part was obtained by introducing a PI-based method. In [4], the authors introduced discount factor to define the performance function and attempted to design the controller as a whole. In [5], the authors developed a data-based method according to the performance function with discount factor. However, since most works were developed by PI, the initial restriction was generally inevitable. Some researchers attempted to overcome the initial restriction and designed the tuning law by adding a Lyapunov function candidate term [6], [7], but the concrete way to select the Lyapunov function candidate was not proposed, causing difficulty in implementation of the methods. In [9], an algorithm was proposed to derive the optimal tracking controller by using VI, but the control constraints were not considered.

To overcome the control constraints, [8] defined a non-quadratic performance function for constrained systems and proved the iterative convergence for PI. In [10], an algorithm was developed to learn constrained controller with experience replay technique by introducing integral reinforcement learning. [11], [12] studied the off-policy methods for constrained optimal controller design. In [3], the authors proposed the feasibility of PI-based methods while designing the constrained optimal tracking controller. However, the initial condition was restricted to requiring an admissible control policy.

In this paper, the value iteration mechanism is used to design the nonlinear constrained optimal tracking controller. First, we augmented the original system dynamics, and introduce discount factor to define a nonquadratic value function. Then, the algorithm is proposed and the initial condition is given by an assumption, which overcomes the initial restriction in PI. The iterative property is analysed and the iterative property and the optimality of the algorithm are discussed. Finally, a nonlinear case is presented to show the feasibility.

II. Problem statement

Consider the continuous-time nonlinear system as

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

where $x \in \mathbb{R}^n$ represents system state vector, $u \in \mathbb{R}^m$ represents the control input vector. $f(x(t)) \in \mathbb{R}^n$ and $g(x(t)) \in \mathbb{R}^{n\times m}$ represent the drift dynamics and the input dynamics of the system, respectively. The control u is assumed to be constrained as below

$$|u_i| \leq \bar{u}, \tag{2}$$

where $\bar{u} > 0$ represents the bound value and $i = 1, 2, \ldots, m$.

$l(t)$ represents the desired trajectory and is assumed to satisfy

$$\dot{l}(t) = s(l(t)), s(0) = 0. \quad (3)$$

The controller is aimed at tracking the trajectory of $l(t)$ while optimizes a value function and satisfies the bound condition.

Generally, the controller is combined by the steady-state control u_d and the feedback control u_e. The steady-state part u_d satisfies

$$\dot{x} = f(l) + g(l)u_d = s(l), \quad (4)$$

and yields

$$u_d = g^{-1}(l)(s(l) - f(l)). \quad (5)$$

The other part u_e is aimed at minimizing a value function

$$V_e(e) = \int_t^{\infty} [u_e^{\mathrm{T}} R u_e + e^{\mathrm{T}} Q_e e] \mathrm{d}\tau, \quad (6)$$

where $e = x - d$ represents the tracking error, $Q_e(x)$ is a positive definite matrix. Using the HJB equation [1], it has that

$$u_e = -\frac{1}{2} R^{-1} g(x)^{\mathrm{T}} \frac{\partial V_e(e)}{\partial e}. \quad (7)$$

Combine the above two parts, the optimal tracking controller is derived. However, since the control input is constrained, the control policy can not be designed as two independent parts and should be considered as a whole.

III. Main results

A. Augmented system

Inspired by [3], we augment the original system dynamics as

$$\dot{\Lambda} = F(\Lambda) + G(\Lambda)u, \quad (8)$$

where $\Lambda = [e^{\mathrm{T}} l^{\mathrm{T}}]^{\mathrm{T}} \subset \mathbb{R}^{2n}$, and

$$F(\Lambda) = \begin{bmatrix} f(e+l) - s(l) \\ s(l) \end{bmatrix} \quad (9)$$

$$G(\Lambda) = \begin{bmatrix} g(e+l) \\ 0 \end{bmatrix}. \quad (10)$$

Then, to handle the feature of the constrained peculiarity, a nonquadratic function is adopted to design the value function.

$$\begin{aligned} &V(\Lambda(t)) \\ &= \int_t^{\infty} \mathrm{e}^{-\lambda(\tau - t)} \left[\Lambda^{\mathrm{T}} Q \Lambda + 2 \int_0^{u(\tau)} \left[\bar{u} \phi^{-1} \left(\frac{\mu}{\bar{u}} \right) \right]^{\mathrm{T}} R \mathrm{d}\mu \right] \mathrm{d}\tau. \end{aligned} \quad (11)$$

where $Q = [Q_e, 0; 0, 0]$, $\phi(\cdot) = \tanh(\cdot)$, and $R = diag(r_1, r_2, \ldots, r_m) \geq 0$.

It should be noted that the discount factor λ in the above value function (21) is necessary [3], [4]. If $\lambda = 0$, the value function (11) will be unbounded while $d(t)$ keeps oscillate, because u_d is depends on $d(t)$.

Differentiate $V(\Lambda(t))$ along (8), the tracking Bellman equation is derived as

$$\begin{aligned} \dot{V}(\Lambda) =& -\Lambda^{\mathrm{T}} Q \Lambda - 2 \int_0^{u} \left[\bar{u} \phi^{-1} \left(\frac{\mu}{\bar{u}} \right) \right]^{\mathrm{T}} R \mathrm{d}\mu \\ &+ \lambda \int_t^{\infty} \mathrm{e}^{-\lambda(\tau - t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &+ 2 \int_0^{u(\tau)} \left[\bar{u} \phi^{-1} \left(\frac{\mu}{\bar{u}} \right) \right]^{\mathrm{T}} R \mathrm{d}\mu \Big] \mathrm{d}\tau \\ =& \lambda V(\Lambda) - \Lambda^{\mathrm{T}} Q \Lambda - 2 \int_0^{u} \left[\bar{u} \phi^{-1} \left(\frac{\mu}{\bar{u}} \right) \right]^{\mathrm{T}} R \mathrm{d}\mu \end{aligned} \quad (12)$$

Then, the HJB equation is

$$\begin{aligned} &H(\Lambda, u^*, \nabla V^*(\Lambda)) \\ &= -\lambda V^*(\Lambda) + \nabla V^*(\Lambda)^{\mathrm{T}} (F(\Lambda) + G(\Lambda) u^*) \\ &\quad + \Lambda^{\mathrm{T}} Q \Lambda + 2 \int_0^{u^*} \left[\bar{u} \phi^{-1} \left(\frac{\mu}{\bar{u}} \right) \right]^{\mathrm{T}} R \mathrm{d}\mu = 0 \end{aligned} \quad (13)$$

where $\nabla V^*(\Lambda) = \partial V^*(\Lambda) / \partial \Lambda$. The optimal value function $V^*(\Lambda)$ is

$$\begin{aligned} V^*(\Lambda(t)) = \min_u \int_t^{\infty} & \mathrm{e}^{-\lambda(\tau - t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &+ 2 \int_0^{u(\tau)} \left[\bar{u} \phi^{-1} \left(\frac{\mu}{\bar{u}} \right) \right]^{\mathrm{T}} R \mathrm{d}\mu \Big] \mathrm{d}\tau \end{aligned} \quad (14)$$

Using the stationarity condition, the optimal solution $u^*(\Lambda)$ is

$$u^*(\Lambda) = -\bar{u} \phi\left(\frac{1}{2\bar{u}} R^{-1} G(\Lambda)^{\mathrm{T}} \nabla V^*(\Lambda) \right). \quad (15)$$

In [3], the optimal tracking controller is solved by PI method, which suffered from an initial restriction. Next, we will introduce the value iteration mechanism to design the algorithm.

B. VI algorithm

For convenience, define $\Lambda^u(t+T)$ as below

$$\Lambda^u(t+T) = \Lambda(t) + \int_t^{t+T} F(\Lambda(\tau)) + G(\Lambda(\tau)) u(\Lambda(\tau)) \mathrm{d}\tau. \quad (16)$$

Then, the VI algorithm can be proposed as:

Policy improvement

$$u_i(\Lambda) = -\bar{u} \phi\left(\frac{1}{2\bar{u}} R^{-1} G(\Lambda)^{\mathrm{T}} \nabla V_i(\Lambda) \right). \quad (17)$$

Value function update

$$\begin{aligned} V_{i+1}(\Lambda(t)) =& \mathrm{e}^{-\lambda T} V_i(\Lambda^{u_i}(t+T)) \\ &+ \int_t^{t+T} \mathrm{e}^{-\lambda(\tau - t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &+ 2 \int_0^{u_i(\Lambda(\tau))} \left[\bar{u} \phi^{-1} \left(\frac{\mu}{\bar{u}} \right) \right]^{\mathrm{T}} R \mathrm{d}\mu \Big] \mathrm{d}\tau. \end{aligned} \quad (18)$$

Before we show the iterative property of the proposed algorithm, some assumptions are assumed.

Assumption 1. *The initial condition of the proposed algorithm satisfies*

$$\underline{\kappa}V^*(\Lambda) \leq V_0(\Lambda) \leq \overline{\kappa}V^*(\Lambda), \tag{19}$$

where $V_0(\Lambda)$ *represents the initial value function,* $0 \leq \underline{\kappa} \leq \overline{\kappa} < \infty$.

The above assumption is the initial condition of the proposed algorithm. Since $V_0(\Lambda)$ is semi-definite positive and $V^*(\Lambda)$ is definite positive, the above assumption is reasonable.

Assumption 2. *There exist constant* $0 < \beta < \infty$ *satisfies*

$$\begin{aligned} &V^*(\Lambda^\mu(t+T))) \\ &\leq \beta \int_t^{t+T} \mathrm{e}^{-\lambda(\tau-t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &\quad + 2\int_0^{u(\tau)} \left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}} R \mathrm{d}\mu\Big] \mathrm{d}\tau \end{aligned} \tag{20}$$

Next, based on the above definitions and assumptions, the theorems are proposed.

Theorem 1. *The iterations between the proposed algorithm* (17) *and* (18) *satisfies*

$$\underline{\kappa}_i V^*(\Lambda) \leq V_i(\Lambda) \leq \overline{\kappa}_i V^*(\Lambda), \tag{21}$$

$$\underline{\kappa}_{i+1} = 1 + \frac{\underline{\kappa}_i - 1}{1+\beta^{-1}}, \qquad \overline{\kappa}_{i+1} = 1 + \frac{\overline{\kappa}_i - 1}{1+\beta^{-1}} \tag{22}$$

where $\underline{\kappa}_0 = \underline{\kappa}$ *and* $\overline{\kappa}_0 = \overline{\kappa}$.

Proof. The discussion will be divided into two parts.

1.Lower bound of (21)

Using Assumption 1, while $i=1$, it yields

$$\begin{aligned} &V_1(\Lambda(t)) \\ &= \min_u \Big[\mathrm{e}^{-\lambda T} V_0(\Lambda^u(t+T)) \\ &\quad + \int_t^{t+T} \mathrm{e}^{-\lambda(\tau-t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &\quad + 2\int_0^{u} \left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}} R \mathrm{d}\mu\Big] \mathrm{d}\tau\Big] \\ &\geq \min_u \Big[\frac{\underline{\kappa}-1}{1+\beta}\mathrm{e}^{-\lambda T} V^*(\Lambda^u(t+T)) \\ &\quad - \frac{\underline{\kappa}-1}{1+\beta}\mathrm{e}^{-\lambda T} V^*(\Lambda^u(t+T)) \\ &\quad + \int_t^{t+T} \mathrm{e}^{-\lambda(\tau-t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &\quad + 2\int_0^{u} \left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}} R \mathrm{d}\mu\Big] \mathrm{d}\tau \\ &\quad + \underline{\kappa}\mathrm{e}^{-\lambda T} V^*(\Lambda^u(t+T))\Big]. \end{aligned} \tag{23}$$

Note that the following fact always holds for constants $y \geq 0$, $z > 0$

$$y - \frac{y-1}{1+z} = 1 + \frac{y-1}{1+z^{-1}}. \tag{24}$$

According to the Assumption 2, it obtains that

$$\begin{aligned} &V_1(\Lambda(t)) \\ &\geq \min_u \Big[(\underline{\kappa} - \frac{\underline{\kappa}-1}{1+\beta})\mathrm{e}^{-\lambda T} V^*(x^u(t+T)) \\ &\quad + (1 + \frac{\underline{\kappa}-1}{1+\beta^{-1}}) \int_t^{t+T} \mathrm{e}^{-\lambda(\tau-t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &\quad + 2\int_0^{u} \left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}} R \mathrm{d}\mu\Big] \mathrm{d}\tau\Big] \\ &= \left(1 + \frac{\underline{\kappa}-1}{1+\beta^{-1}}\right) \min_u \Big[\mathrm{e}^{-\lambda T} V^*(\Lambda^u(t+T)) \\ &\quad + \int_t^{t+T} \mathrm{e}^{-\lambda(\tau-t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &\quad + 2\int_0^{u} \left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}} R \mathrm{d}\mu\Big] \mathrm{d}\tau + \Big] \\ &= \underline{\kappa}_1 V^*(\Lambda(t)). \end{aligned} \tag{25}$$

If the lower bound of (21) holds for $i = j-1$, it has

$$\begin{aligned} &V_j(\Lambda(t)) \\ &= \min_u \Big[\mathrm{e}^{-\lambda T} V_{j-1}(\Lambda^u(t+T)) \\ &\quad + \int_t^{t+T} \mathrm{e}^{-\lambda(\tau-t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &\quad + 2\int_0^{u} \left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}} R \mathrm{d}\mu\Big] \mathrm{d}\tau\Big] \\ &\geq \min_u \Big[\underline{\kappa}_{j-1}\mathrm{e}^{-\lambda T} V^*(\Lambda^u(t+T)) \\ &\quad + \int_t^{t+T} \mathrm{e}^{-\lambda(\tau-t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &\quad + 2\int_0^{u} \left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}} R \mathrm{d}\mu\Big] \mathrm{d}\tau\Big] \\ &\geq \min_u \Big[(\underline{\kappa}_{j-1} - \frac{\underline{\kappa}_{j-1}-1}{1+\beta})\mathrm{e}^{-\lambda T} V^*(\Lambda^u(t+T)) \\ &\quad + (1 + \frac{\underline{\kappa}_{j-1}-1}{1+\beta^{-1}}) \int_t^{t+T} \mathrm{e}^{-\lambda(\tau-t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &\quad + 2\int_0^{u} \left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}} R \mathrm{d}\mu\Big] \mathrm{d}\tau\Big] \\ &= \left(1 + \frac{\underline{\kappa}_{j-1}-1}{1+\beta^{-1}}\right) \min_u \Big[\mathrm{e}^{-\lambda T} V^*(\Lambda^u(t+T)) \\ &\quad + \int_t^{t+T} \mathrm{e}^{-\lambda(\tau-t)} \Big[\Lambda^{\mathrm{T}} Q \Lambda \\ &\quad + 2\int_0^{u} \left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}} R \mathrm{d}\mu\Big] \mathrm{d}\tau\Big] \\ &= \underline{\kappa}_j V^*(\Lambda(t)). \end{aligned} \tag{26}$$

The proof of the first part is completed.

2.Upper bound of (21)

Using Assumption 1, while $i=1$, it yields

$$\begin{aligned}
&V_1(\Lambda(t))\\
&=\min_u\Bigg[\mathrm{e}^{-\lambda T}V_0(\Lambda^u(t+T))\\
&+\int_t^{t+T}\mathrm{e}^{-\lambda(\tau-t)}\Bigg[\Lambda^{\mathrm{T}}Q\Lambda\\
&+2\int_0^u\left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}}R\mathrm{d}\mu\Bigg]\mathrm{d}\tau\Bigg]\\
&\leq\min_u\Bigg[\overline{\kappa}\mathrm{e}^{-\lambda T}V^*(\Lambda^u(t+T))\\
&\quad+\frac{\overline{\kappa}-1}{1+\beta}\mathrm{e}^{-\lambda T}V^*(\Lambda^u(t+T))\\
&\quad-\frac{\overline{\kappa}-1}{1+\beta}\mathrm{e}^{-\lambda T}V^*(\Lambda^u(t+T))\\
&\quad+\int_t^{t+T}\mathrm{e}^{-\lambda(\tau-t)}\Bigg[\Lambda^{\mathrm{T}}Q\Lambda\\
&\quad+2\int_0^u\left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}}R\mathrm{d}\mu\Bigg]\mathrm{d}\tau\Bigg].
\end{aligned}\tag{27}$$

According to the Assumption 2, it yields

$$\begin{aligned}
&V_1(\Lambda(t))\\
&\leq\min_u\Bigg[(\overline{\kappa}-\frac{\overline{\kappa}-1}{1+\beta})\mathrm{e}^{-\lambda T}V^*(x^u(t+T))\\
&\quad+(1+\frac{\overline{\kappa}-1}{1+\beta^{-1}})\int_t^{t+T}\mathrm{e}^{-\lambda(\tau-t)}\Bigg[\Lambda^{\mathrm{T}}Q\Lambda\\
&\quad+2\int_0^u\left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}}R\mathrm{d}\mu\Bigg]\mathrm{d}\tau\Bigg]\\
&=\left(1+\frac{\overline{\kappa}-1}{1+\beta^{-1}}\right)\min_u\Bigg[\mathrm{e}^{-\lambda T}V^*(\Lambda^u(t+T))\\
&\quad+\int_t^{t+T}\mathrm{e}^{-\lambda(\tau-t)}\Bigg[\Lambda^{\mathrm{T}}Q\Lambda\\
&\quad+2\int_0^u\left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}}R\mathrm{d}\mu\Bigg]\mathrm{d}\tau\Bigg]\\
&=\overline{\kappa}_1V^*(\Lambda(t)).
\end{aligned}\tag{28}$$

If the upper bound of (21) holds for $i=j-1$, it has

$$\begin{aligned}
&V_j(\Lambda(t))\\
&=\min_u\Bigg[\mathrm{e}^{-\lambda T}V_{j-1}(\Lambda^u(t+T))\\
&\quad+\int_t^{t+T}\mathrm{e}^{-\lambda(\tau-t)}\Bigg[\Lambda^{\mathrm{T}}Q\Lambda\\
&\quad+2\int_0^u\left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}}R\mathrm{d}\mu\Bigg]\mathrm{d}\tau\Bigg]\\
&\leq\min_u\Bigg[\overline{\kappa}_{j-1}\mathrm{e}^{-\lambda T}V^*(\Lambda^u(t+T))\\
&\quad+\int_t^{t+T}\mathrm{e}^{-\lambda(\tau-t)}\Bigg[\Lambda^{\mathrm{T}}Q\Lambda\\
&\quad+2\int_0^u\left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}}R\mathrm{d}\mu\Bigg]\mathrm{d}\tau\Bigg]\\
&\leq\min_u\Bigg[(\overline{\kappa}_{j-1}-\frac{\overline{\kappa}_{j-1}-1}{1+\beta})\mathrm{e}^{-\lambda T}V^*(\Lambda^u(t+T))\\
&\quad+(1+\frac{\overline{\kappa}_{j-1}-1}{1+\beta^{-1}})\int_t^{t+T}\mathrm{e}^{-\lambda(\tau-t)}\Bigg[\Lambda^{\mathrm{T}}Q\Lambda\\
&\quad+2\int_0^u\left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}}R\mathrm{d}\mu\Bigg]\mathrm{d}\tau\Bigg]\\
&=\left(1+\frac{\overline{\kappa}_{j-1}-1}{1+\beta^{-1}}\right)\min_u\Bigg[\mathrm{e}^{-\lambda T}V^*(\Lambda^u(t+T))\\
&\quad+\int_t^{t+T}\mathrm{e}^{-\lambda(\tau-t)}\Bigg[\Lambda^{\mathrm{T}}Q\Lambda\\
&\quad+2\int_0^u\left[\bar{u}\phi^{-1}\left(\frac{\mu}{\bar{u}}\right)\right]^{\mathrm{T}}R\mathrm{d}\mu\Bigg]\mathrm{d}\tau\Bigg]\\
&=\overline{\kappa}_jV^*(\Lambda(t)).
\end{aligned}\tag{29}$$

The proof of the second part is completed. □

Theorem 2. *Iterated between the proposed algorithm* (17) *and* (18)*, both sides of* (21) *converge to* $V^*(\Lambda)$ *as* $i\to\infty$*, i.e.,*

$$\lim_{i\to\infty}\overline{\kappa}_iV^*(\Lambda)=\lim_{i\to\infty}\underline{\kappa}_iV^*(\Lambda)=V^*(\Lambda).\tag{30}$$

Proof. From Theorem 1, the lower bound of (21) satisfies

$$V_i(\Lambda)\geq\underline{\kappa}_iV^*(\Lambda),\quad\underline{\kappa}_{i+1}=1+\frac{\underline{\kappa}_i-1}{1+\beta^{-1}},\tag{31}$$

and it has

$$\lim_{i\to\infty}\underline{\kappa}_iV^*(\Lambda)=\lim_{i\to\infty}\left(1+\frac{\underline{\kappa}-1}{(1+\beta^{-1})^i}\right)V^*(\Lambda)=V^*(\Lambda).\tag{32}$$

The upper bound of (21) satisfies

$$V_i(\Lambda)\leq\overline{\kappa}_iV^*(\Lambda),\quad\overline{\kappa}_{i+1}=1+\frac{\overline{\kappa}_i-1}{1+\beta^{-1}},\tag{33}$$

and it has

$$\lim_{i\to\infty}\overline{\kappa}_iV^*(\Lambda)=\lim_{i\to\infty}\left(1+\frac{\overline{\kappa}-1}{(1+\beta^{-1})^i}\right)V^*(\Lambda)=V^*(\Lambda).\tag{34}$$

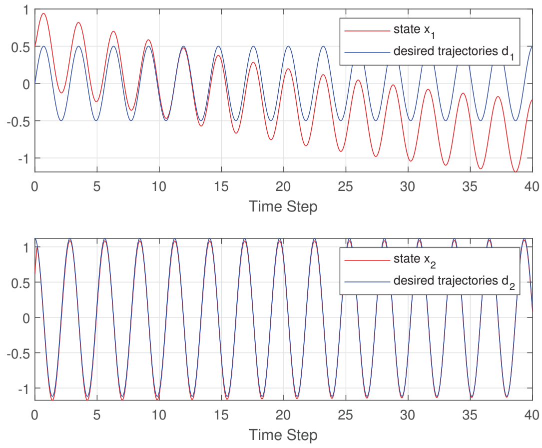

Fig. 1. State trajectories with the first iterative control.

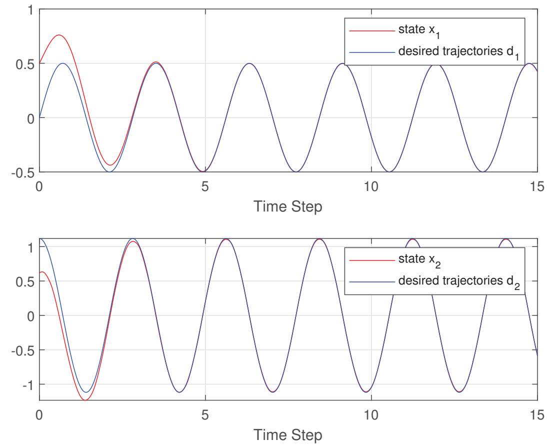

Fig. 2. State trajectories with the obtained control.

The limitation of both sides of (21) can be obtained. The proof is completed. □

Since both sides of (21) converge to $V^*(\Lambda)$, the optimality can be obtained by squeeze theorem.

IV. Simulation

In this part, a nonlinear system [3] is simulated to verify the algorithm, and the system dynamics are as

$$\dot{x} = \begin{bmatrix} x_2 \\ -x_1^3 - 0.5x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u, \tag{35}$$

the control input is bounded as $|u| \leq \bar{u} = 3$.

The desired trajectories are as below,

$$\dot{l} = Sl = \begin{bmatrix} 0 & 1 \\ -5 & 0 \end{bmatrix} l, \tag{36}$$

and the initial point is $l(0) = [0, \sqrt{5}/2]$. Reconstruct the system dynamics in (8) form, it has

$$\begin{aligned} \dot{\Lambda} &= \begin{bmatrix} \Lambda_2 \\ -(\Lambda_1 + \Lambda_3)^3 - 0.5(\Lambda_2 + \Lambda_4) + 5\Lambda_3 \\ \Lambda_4 \\ -5\Lambda_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} u \\ &= F(\Lambda) + G(\Lambda)u. \end{aligned} \tag{37}$$

The integral time sets as $T = 0.05s$. The value function (11) is set as $\lambda = 0.01$, $R = 1$, $Q_e = 1000I$. A neural network with structure 4-100-1 is adopted to approximate the iterative value function, and choose $tanh(\cdot)$ as the activation function.

The state trajectories and the desired trajectories under $u_1(\Lambda)$ are presented in Fig.1. The comparison displays that the trajectory of state x_1 diverges from the desired trajectory d_1, which implies $u_1(\Lambda)$ is not an admissible control policy.

After implementing the proposed VI algorithm for 100 iterations, a converged control policy is obtained. The state trajectories and the input trajectory under $u_{100}(\Lambda)$ are presented in Fig.2-Fig.3. The state trajectories can track the desired trajectories at around time $4s$ in Fig.2. In Fig.3, it displayes that the input trajectory satisfies the bound condition. After the state trajectories reach the desired trajectories, only the steady-state part u_d is required, and thus the input trajectory oscillates with an equal amplitude.

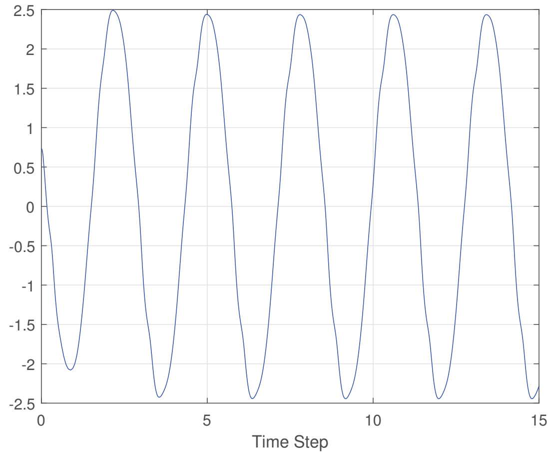

Fig. 3. Input trajectories with the obtained control.

V. Conclusion

The VI mechanism has been introduced to design the continuous-time nonlinear constrained optimal tracking controller. The original system dynamics has been augmented, and a nonquadratic performance function with discount factor has been defined. The proposed VI algorithm could be initialized by any arbitrary semi-definite positive function, which overcame the initial restriction in PI-based methods. The iterative property and the optimality of the proposed VI algorithm has been proved.

Acknowledgment

This work is supported by the National Key Research and Development Project (No.2018YFB2100401) and Key

Research and Development Program of Zhejiang Province (2020C01021).

References

[1] H. Zhang, D. Liu, Y. Luo, and D. Wang, *Adaptive Dynamic Programming for Control*. London, U.K.: Springer, 2013.

[2] H. Zhang, L. Cui, X. Zhang, and Y. Luo, "Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2226–2236, 2011.

[3] H. Modares and F. L. Lewis, "Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning," *Automatica*, vol. 50, no. 7, pp. 1780–1792, 2014.

[4] ——, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 3051–3056, 2014.

[5] G. Xiao, H. Zhang, Y. Luo, and H. Jiang, "Data-driven optimal tracking control for a class of affine non-linear continuous-time systems with completely unknown dynamics," *IET Control Theory and Applications*, vol. 10, no. 6, pp. 700–710, 2016.

[6] X. Yang, D. Liu, Q. Wei, and D. Wang, "Guaranteed cost neural tracking control for a class of uncertain nonlinear systems using adaptive dynamic programming," *Neurocomputing*, vol. 198, pp. 80–90, 2016.

[7] D. Wang, D. Liu, Y. Zhang, and H. Li, "Neural network robust tracking control with adaptive critic framework for uncertain nonlinear systems," *Neural Networks*, vol. 97, pp. 11–18, 2018.

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

[9] G. Xiao, H. Zhang, Y. Luo, and Q. Qu, "General value iteration based reinforcement learning for solving optimal tracking control problem of continuoustime affine nonlinear systems," *Neurocomputing*, vol. 245, pp. 114–123, 2017.

[10] H. Modares, F. L. Lewis, and M.-B. Naghibi-Sistani, "Integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems," *Automatica*, vol. 50, no. 1, pp. 193–202, 2014.

[11] B. Luo, H. Wu, T. Huang, and D. Liu, "Reinforcement learning solution for HJB equation arising in constrained optimal control problem," *Neural Networks*, vol. 71, pp. 150–158, 2015.

[12] X. Yang, D. Liu, B. Luo, and C. Li, "Data-based robust adaptive control for a class of unknown nonlinear constrained-input systems via integral reinforcement learning," *Information Sciences*, vol. 369, pp. 731–747, 2016.