Neurocomputing 266 (2017) 353–360

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A novel neural optimal control framework with nonlinear dynamics: Closed-loop stability and simulation verification[☆]

Ding Wang[a,b], Chaoxu Mu[c,*]

[a] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[c] Tianjin Key Laboratory of Process Measurement and Control, School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 20 February 2017
Revised 1 April 2017
Accepted 20 May 2017
Available online 25 May 2017

Communicated by Mou Chen

Keywords:
Adaptive dynamic programming
Adaptive system
Learning control
Neural network
Optimal regulator
Stability

ABSTRACT

In this paper, we focus on developing adaptive optimal regulators for a class of continuous-time nonlinear dynamical systems through an improved neural learning mechanism. The main objective lies in that establishing an additional stabilizing term to reinforce the traditional training process of the critic neural network, so that to reduce the requirement with respect to the initial stabilizing control, and therefore, bring in an obvious convenience to the adaptive-critic-based learning control implementation. It is exhibited that by employing the novel updating rule, the adaptive optimal control law can be obtained with an excellent approximation property. The closed-loop system is constructed and its stability issue is handled by considering the improved learning criterion. Experimental simulations are also conducted to verify the efficient performance of the present design method, especially the major role that the stabilizing term performed.

1. Introduction

As is known, linear optimal regulator design has been studied by control scientists and engineers for many years. For nonlinear systems, the optimal control problem always leads to cope with the nonlinear Hamilton–Jacobi–Bellman (HJB) equation, which is intractable to solve in general cases. Fortunately, a series of iterative methods have been established to tackle the optimal control problems approximately [1–3]. For adaptive/approximate dynamic programming (ADP) [3–9], the adaptive critic is taken as the basic structure and neural networks are often involved to serve as the function approximator. Generally speaking, employing the ADP method always results in approximate or adaptive optimal feedback controllers. Note that optimality and adaptivity are two important criteria of control theory and also possess grea t significance to control engineering, such as [10–16]. Hence, this kind of adaptive-critic-based optimal control design has great potentials in various control applications.

In the last decade, the methodology of ADP has been widely used for optimal control of discrete-time systems, such as [17–24] and continuous-time systems, like [25–32]. Heydari and Balakrishnan [18] investigated finite-horizon nonlinear optimal control with input constraints by adopting single network adaptive critic designs. Song et al. [19] proposed a novel ADP algorithm to solve the nearly optimal finite-horizon control problem for a class of deterministic nonaffine nonlinear time-delay systems. Mu et al. [21] studied the approximate optimal tracking control design for a class of discrete-time nonlinear systems based on the iterative globalized dual heuristic programming technique. Zhao et al. [22] gave a model-free optimal control method for optimal control of affine nonlinear systems without using the dynamics information. Qin et al. [23] studied the neural-network-based self-learning H_∞ control design for discrete-time input-affine nonlinear systems in light of ADP method. Zhong et al. [24] developed the theoretical basis of the new goal representation heuristic dynamic programming structure for general discrete-time nonlinear systems. Vamvoudakis and Lewis [25] proposed an important actor-critic algorithm to attain the continuous-time infinite horizon nonlinear optimal regulation design. Zhang et al. [26] studied the approximate optimal control for non-zero-sum differential games with continuous-time nonlinear dynamics based on single

☆ This work was supported in part by National Natural Science Foundation of China under Grants U1501251, 61533017, and 61233001, in part by Beijing Natural Science Foundation under Grant 4162065, in part by Research Fund of Tianjin Key Laboratory of Process Measurement and Control under Grant TKLPMC-201612, and in part by the Early Career Development Award of SKLMCCS.

* Corresponding author.
E-mail addresses: ding.wang@ia.ac.cn (D. Wang), cxmu@tju.edu.cn (C. Mu).

http://dx.doi.org/10.1016/j.neucom.2017.05.051
0925-2312/

network adaptive critics. Modares and Lewis [27] proposed a linear quadratic trajectory tracking control method for partially-unknown continuous-time systems based on the reinforcement learning technique. Na and Herrmann [28] proposed an online adaptive and approximate optimal trajectory tracking approach with a simplified dual approximation architecture for continuous-time unknown nonlinear controlled plants. Bian et al. [29] studied decentralized adaptive optimal control of a class of large-scale systems and its application toward the power systems. Jiang and Jiang [30] originally established the global ADP structure for continuous-time nonlinear systems. Luo et al. [31] provided the reinforcement learning solution for HJB equation with respect to the constrained optimal control problems. Gao and Jiang [32] applied ADP to design optimal output regulation of linear systems adaptively. This greatly promotes the development of the adaptive critic control designs of complex nonlinear systems. However, the traditional adaptive critic control design always depends on the choice of an initial stabilizing control, which is pretty difficult to find out in control practices. Actually, requiring an initial stabilizing control is a common property of [25,27], which weakens the application aspect of the adaptive-critic-based design to a certain extent, and correspondingly, motivates our research greatly. This paper focuses on developing nonlinear adaptive optimal regulators through an improved neural learning mechanism. The major contribution lies in that it constructs a simple reinforced structure to achieve the nonlinear optimal regulation design adaptively, without requiring the initial stabilizing controller. Moreover, the stability of the closed-loop system including the additional stabilizing term is presented with a simpler proof process. Finally, the important role that the stabilizing term plays is also verified by simulation study in detail. This can be regarded as an improvement to the traditional adaptive critic designs, like [25,27].

The rest of the current paper is organized as follows. The studied problem is described briefly in Section 2. The improved adaptive critic design technique of nonlinear adaptive optimal control is developed with closed-loop stability analysis in Section 3. The simulation studies and the concluding remarks are presented in Section 4 and Section 5, respectively. Incidentally, the main notations used in the paper are listed as follows. $\mathbb{R}$ stands for the set of all real numbers. $\mathbb{R}^n$ is the Euclidean space of all n-dimensional real vectors. $\mathbb{R}^{n\times m}$ is the space of all $n \times m$ real matrices. $\|\cdot\|$ denotes the vector norm of a vector in $\mathbb{R}^n$ or the matrix norm of a matrix in $\mathbb{R}^{n\times m}$. I_n represents the $n \times n$ identity matrix. $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ calculate the maximal and minimal eigenvalues of a matrix, respectively. Let Ω be a compact subset of $\mathbb{R}^n$ and $\mathscr{A}(\Omega)$ be the set of admissible control laws on Ω. The superscript "T" is taken for representing the transpose operation and $\nabla(\cdot)\triangleq\partial(\cdot)/\partial x$ is employed to denote the gradient operator.

2. Problem statement

In this paper, we study a class of continuous-time nonlinear systems with input-affine form given by

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

where $x(t) \in \Omega \subset \mathbb{R}^n$ is the state variable, $u(t) \in \Omega_u \subset \mathbb{R}^m$ is the control variable, and the system functions $f(\cdot) \in \mathbb{R}^n$ and $g(\cdot) \in \mathbb{R}^{n\times m}$ are known matrices and are differentiable in the arguments satisfying $f(0) = 0$. In this paper, we let the initial state at $t = 0$ be $x(0) = x_0$ and let $x = 0$ be the equilibrium point. In addition, we assume that $f(x)$ is Lipschitz continuous on a set Ω in $\mathbb{R}^n$ which contains the origin and the nonlinear plant (1) is controllable.

In order to design the optimal feedback control law $u(x)$, we let $Q(x) > 0$ when $x \neq 0$ and $Q(0) = 0$, set R as a positive definite matrix with appropriate dimension, take

$$U(x(\tau), u(\tau)) = Q(x(\tau)) + u^{\mathsf{T}}(\tau)Ru(\tau)$$

to stand for the utility function, and then define the infinite horizon cost function as

$$J(x(t), u) = \int_t^{\infty} U(x(\tau), u(\tau))\mathrm{d}\tau. \tag{2}$$

Notice here the cost $J(x(t), u)$ is often written as $J(x(t))$ or $J(x)$ for simplicity. For an admissible control law $u \in \mathscr{A}(\Omega)$, if the cost function (2) is continuously differentiable, then the related infinitesimal version is the nonlinear Lyapunov equation

$$0 = U(x, u) + (\nabla J(x))^{\mathsf{T}}[f(x) + g(x)u]$$

with $J(0) = 0$. Next, we define the Hamiltonian of system (1) as

$$H(x, u, \nabla J(x)) = U(x, u) + (\nabla J(x))^{\mathsf{T}}[f(x) + g(x)u].$$

According to Bellman's optimality principle, the optimal cost function $J^*(x)$

$$J^*(x) = \min_{u\in\mathscr{A}(\Omega)} \int_t^{\infty} U(x(\tau), u(\tau))\mathrm{d}\tau,$$

makes sure that the so-called HJB equation

$$\min_u H(x, u, \nabla J^*(x)) = 0$$

holds. Similar as [25,30], the optimal feedback control law is computed by

$$u^*(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\nabla J^*(x). \tag{3}$$

Noticing the optimal control expression (3), the HJB equation is in fact

$$\begin{aligned} 0 &= U(x, u^*) + (\nabla J^*(x))^{\mathsf{T}}[f(x) + g(x)u^*] \\ &= Q(x) + (\nabla J^*(x))^{\mathsf{T}} f(x) \\ &\quad - \frac{1}{4}(\nabla J^*(x))^{\mathsf{T}}g(x)R^{-1}g^{\mathsf{T}}(x)\nabla J^*(x), J^*(0) = 0. \end{aligned} \tag{4}$$

Eq. (4) is actually $H(x, u^*, \nabla J^*(x)) = 0$, which is difficult to get the solution theoretically. In other words, it is clearly not easy to obtain the optimal control law (3) for general nonlinear systems, which inspires us to effectively design a class of approximate optimal control schemes.

3. Approximate optimal control design and its stability

During the approximate control algorithm implementation, the idea of adaptive critic is adopted with neural network approximation. Using the universal approximation property, the optimal cost function $J^*(x)$ can be expressed by a neural network with a single hidden layer on a compact set Ω as

$$J^*(x) = \omega_c^{\mathsf{T}}\sigma_c(x) + \varepsilon_c(x), \tag{5}$$

where $\omega_c \in \mathbb{R}^{l_c}$ is the ideal weight vector that is upper bounded, l_c is the number of hidden neurons, $\sigma_c(x) \in \mathbb{R}^{l_c}$ is the activation function, and $\varepsilon_c(x) \in \mathbb{R}$ is the reconstruction error. Then, the gradient vector is

$$\nabla J^*(x) = (\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x).$$

Noticing the ideal weight is unknown in advance, a critic network is developed to approximate the optimal cost function as

$$\hat{J}^*(x) = \hat{\omega}_c^{\mathsf{T}}\sigma_c(x), \tag{6}$$

where $\hat{\omega}_c \in \mathbb{R}^{l_c}$ denotes the estimated weight vector. Similarly, we derive the gradient vector as

$$\nabla\hat{J}^*(x) = (\nabla\sigma_c(x))^{\mathsf{T}}\hat{\omega}_c.$$

Considering the feedback formulation (3) and the neural network expression (5), the optimal control law can be rewritten as

$$u^*(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\left[(\nabla\sigma_c(x))^{\mathsf{T}}\omega_c + \nabla\varepsilon_c(x)\right]. \tag{7}$$