Neural Networks 99 (2018) 19–30

Contents lists available at ScienceDirect

Neural Networks

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

Self-learning robust optimal control for continuous-time nonlinear systems with mismatched disturbances ☆

Xiong Yang [a], Haibo He [b,*]

[a] School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China
[b] Department of Electrical, Computer and Biomedical Engineering, University of Rhode Island, Kingston, RI 02881, USA

ARTICLE INFO

Article history:
Received 7 August 2017
Received in revised form 16 October 2017
Accepted 28 November 2017
Available online 13 December 2017

Keywords:
Adaptive dynamic programming
Neural network
Reinforcement learning
Robust optimal control
Mismatched disturbance

ABSTRACT

This paper presents a novel adaptive dynamic programming(ADP)-based self-learning robust optimal control scheme for input-affine continuous-time nonlinear systems with mismatched disturbances. First, the stabilizing feedback controller for original nonlinear systems is designed by modifying the optimal control law of the auxiliary system. It is also demonstrated that this feedback controller can optimize a specified value function. Then, within the framework of ADP, a single critic network is constructed to solve the Hamilton–Jacobi–Bellman equation associated with the auxiliary system optimal control law. To update the critic network weights, an indicator function and a concurrent learning technique are employed. By using the proposed update law for the critic network, the restrictive conditions including the initial admissible control and the persistence of excitation condition are relaxed. Moreover, the stability of the closed-loop auxiliary system is guaranteed in the sense that all the signals are uniformly ultimately bounded. Finally, the applicability of the developed control strategy is illustrated through simulations for an unstable nonlinear plant and a power system.

1. Introduction

Owing to powerful abilities of self-learning and adaptivity, adaptive dynamic programming (ADP) and reinforcement learning (RL) have become significant tools for designing optimal controllers for nonlinear systems in the past few decades (Bertsekas & Tsitsiklis, 1996; He, 2011; Jiang & Jiang, 2017). ADP and RL are almost in the same spirits when solving optimal control problems. Therefore, the two names are often interchangeable in the literature (Lewis & Liu, 2013). A common structure utilized in ADP and RL is the actor-critic architecture, which employs two neural networks (NNs). To be specific, the actor NN generates a control policy to the controlled system, and the critic NN evaluates the value of that control policy and gives feedback information to the actor NN (Liu, Wei, Wang, Yang, & Li, 2017). Based on this architecture, many ADP and RL approaches have been proposed.

For discrete-time nonlinear systems, Zhong, Ni, and He (2016) introduced a goal representation ADP to deal with nonlinear optimal control problems. After that, Wei, Liu, Lewis, Liu, and Zhang (2017) developed a mixed iterative ADP to study the optimal battery energy control problem arising in smart residential micro-grids. Differing from Wei et al. (2017) and Zhong et al. (2016) considering optimal control problems in time-triggering mechanisms, Sahoo, Xu, and Jagannathan (2016) presented an event-triggered near optimal control scheme for unknown nonlinear systems via ADP. For continuous-time (CT) nonlinear systems, Vrabie and Lewis (2013) proposed an integral RL to calculate the Nash strategies of partially unknown nonzero-sum games. After that, Lee, Park, and Choi (2015) introduced a novel integral RL to solve the optimal control problem of completely unknown nonlinear systems. Meanwhile, Luo, Wu, Huang, and Liu (2015) developed an off-policy RL to study the optimal control problem of unknown nonlinear systems with input constraints. Later, instead of addressing optimal regulation of nonlinear systems, Kamalapurkar, Andrews, Walters, and Dixon (2017) proposed a model-based RL to design the optimal tracking controller for input-affine nonlinear systems. More recently, Vamvoudakis, Mojoodi, and Ferraz (2017) presented an event-triggered optimal tracking control of input-affine nonlinear systems. In the aforementioned literature, disturbances/perturbations were generally not taken into consideration while designing optimal controllers for nonlinear systems. Nevertheless, in real industry applications, many plants often suffer from external disturbances/perturbations. Therefore, it is necessary to develop robust optimal control strategies for nonlinear systems with disturbances/perturbations.

☆ This work was supported in part by the National Natural Science Foundation of China under Grant 61503379, and in part by China Scholarship Council under the State Scholarship Fund, and in part by the U.S. National Science Foundation under Grants CCF-1439011 and CMMI 1526835.
* Corresponding author.
E-mail addresses: xiong.yang@tju.edu.cn (X. Yang), he@ele.uri.edu (H. He).

https://doi.org/10.1016/j.neunet.2017.11.022
0893-6080/

Recently, a robust ADP was introduced by Jiang and Jiang (2014) to design the robust optimal controller for CT nonlinear systems with matched uncertainties. After that, Wang, Li, Liu, and Mu (2016) extended the robust ADP to develop a data-based robust optimal control scheme for CT nonlinear plants subject to matched disturbances. In the above mentioned literature, the controlled systems satisfied matched conditions (Corless & Leitmann, 1981). In general, robust optimal control approaches for nonlinear systems with matched uncertainties do not always hold for those systems with mismatched uncertainties/disturbances. Though there already exist a few studies on robust optimal control of general nonlinear systems (including systems with matched/mismatched disturbances) (Fu & Chai, 2016; Luo, Wu, & Huang, 2015; Modares, Lewis, & Jiang, 2015; Song, Wei, & Song, 2017; Wang, He, & Liu, 2017; Wei, Song, & Yan, 2016), most of them obtain the robust optimal control policies via solving the zero-sum games. The main difficulty in solving zero-sum games lies in that one needs to judge whether the saddle point exists or not beforehand. According to Zhang, Wei, and Liu (2011), it is generally hard to judge the existence of such saddle points of nonlinear zero-sum games. Even worse, some nonlinear zero-sum games might not exist the saddle points. This difficulty motivates our research.

On the other hand, to derive the robust optimal control schemes, most of the above mentioned methods require to meet certain conditions including the initial admissible control and the persistence of excitation condition. The two conditions are generally difficult to be satisfied (note: a detailed expression has been provided in Section 4). Recently, in order to remove the initial admissible control condition, Dierks and Jagannathan (2010) proposed a single approximator-based control scheme to solve the Hamilton–Jacobi–Isaacs equation. After that, by using a similar architecture as in Dierks and Jagannathan (2010), Nodland, Zargarzadeh, and Jagannathan (2013) designed an NN-based optimal output feedback controller for the helicopter unmanned aerial vehicles. At the same time, Chowdhary, Yucelen, Mühlegg, and Johnson (2013) introduced a concurrent learning technique to relax the persistence of excitation condition. Later, Modares, Lewis, and Naghibi Sistani (2014) applied this technique to design the optimal controller for input-constrained partially unknown CT nonlinear systems. More recently, Zhang, Zhao, and Wang (in press) extended this technique to develop an event-triggered robust control policy for uncertain CT nonlinear systems. However, to the best of our knowledge, there are few studies on developing robust optimal control strategies for nonlinear systems subject to mismatched perturbations with requiring neither the initial admissible control nor the persistence of excitation condition. This also motivates our study.

In this paper, an ADP-based self-learning robust optimal control scheme is developed for a class of CT nonlinear systems with mismatched disturbances for the first time. To begin with, a stabilizing feedback controller for original nonlinear systems is designed by modifying the optimal control law of the auxiliary system. Meanwhile, the stabilizing feedback controller is proved to be able to make a specified value function achieve optimality. To remove the requirements of the initial admissible control and the persistence of excitation condition simultaneously, an indicator function and a concurrent learning technique are employed. In addition, by using Lyapunov's method, all the signals in the closed-loop auxiliary system are demonstrated to be stable in the sense of uniform ultimate boundedness (UUB).

The outline of this paper is illustrated as follows. After briefly presenting the problem description in Section 2, we develop the robust control scheme in Section 3. Section 4 shows that the approximation solution of the Hamilton–Jacobi–Bellman (HJB) equation can be derived via ADP, and Section 5 presents the stability analysis of the closed-loop auxiliary system. Section 6 provides two examples to validate the developed theoretical results. Finally, Section 7 gives concluding remarks of this paper.

Notation: $\mathbb{R}$ denotes the set of all real numbers. $\mathbb{R}^m$ and $\mathbb{R}^{n\times m}$ denote the Euclidean space of all real m-vectors and the space of all $n \times m$ real matrices, respectively. I_n is the identity matrix of dimension $n \times n$. T denotes the transpose. $\triangleq$ means 'equal by definition'. C^1 denotes the class of functions with the continuous derivative. When the vector $x = [x_1, \ldots, x_n]^{\mathsf{T}} \in \mathbb{R}^n$, $\|x\| = \sqrt{\sum_{i=1}^{n} |x_i|^2}$ represents the Euclidean norm of x. When the matrix $A \in \mathbb{R}^{n\times m}$, $\|A\|$ represents the Frobenius-norm of A. V_x denotes the partial derivative of the value function $V(x)$ with respect to the state x, i.e., $V_x = \partial V(x)/\partial x$.

2. Problem description

Consider the robust optimal control problem of an uncertain nonlinear system formulated as follows:

$$\min J(x(t), u(t)) = \int_t^{\infty} r(x(s), u(s))\mathrm{d}s \tag{1}$$

$$s.t.$$

$$\dot{x}(t) = f(x(t)) + g(x(t))u(t) + \Delta f(x(t)) \tag{2}$$

where $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$ are the state and the control vectors, respectively, $r(x, u)$ is a nonnegative function with respect to x and u, $f(x) \in \mathbb{R}^n$ and $g(x) \in \mathbb{R}^{n\times m}$ are known functions, and $\Delta f(x) \in \mathbb{R}^n$ is an uncertain nonlinear function.

To facilitate later analyses, we impose two assumptions which have been utilized in Tripathy, Kar, and Paul (in press), Wang, Mu, Yang, and Liu (2017) and Yang, He, Liu, and Zhu (2017).

Assumption 1. $f(x)$ and $g(x)$ are locally Lipschitz continuous in their arguments. Meanwhile, $f(0) = 0$, i.e., $x = 0$ is an equilibrium point of system (2) when $u(t) = 0$ and $\Delta f(x(t)) = 0$ for all $t \geq 0$. In addition, $x_0 = x(0)$ is the initial system state.

Assumption 2. The uncertain term $\Delta f(x)$ satisfies the mismatched condition, that is,

$$\Delta f(x) = k(x)\omega(x) \ \ (k(x) \neq g(x) \text{ if } m = p)$$

where $k(x) \in \mathbb{R}^{n\times p}$ is a known function and $\omega(x) \in \mathbb{R}^p$ is an uncertain disturbance bounded by a known nonnegative function $\omega_M(x)$, i.e., $\|\omega(x)\| \leq \omega_M(x)$ with $\omega(0) = 0$ and $\omega_M(0) = 0$. Moreover, there exists a nonnegative function $\psi_M(x)$ such that, for every $x \in \mathbb{R}^n$,

$$\left\|g^{+}(x)\Delta f(x)\right\| \leq \psi_M(x)$$

with $\psi_M(0) = 0$ and $g^{+}(x)$ the Moore–Penrose pseudo-inverse of $g(x)$.

The goals of this paper include two aspects: (i) Find a feedback control $u(x)$ such that system (2) is asymptotically stable. (ii) Find a nonnegative function $r(x, u)$ given as in (1) such that the value function $J(x, u)$ achieves optimality under the feedback control $u(x)$.

Owing to the existence of the uncertain disturbance $\omega(x)$, it is often difficult to achieve the objectives (i) and (ii) by using direct methods. To overcome the difficulty, we present an indirect method based on the works of Lin (2007) and Lin and Brandt (1998). Specifically, we will transform the robust optimal control problem (1) and (2) into an optimal control problem of the auxiliary system.

3. Robust optimal control scheme

This section contains two parts. First, we present the auxiliary system and the associated HJB equation. Then, we demonstrate that the robust optimal control of system (2) can be obtained by solving the HJB equation.