Automatica 50 (2014) 1780–1792

Contents lists available at ScienceDirect

Automatica

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

Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning[☆]

Hamidreza Modares[1], Frank L. Lewis

University of Texas at Arlington Research Institute, 7300 Jack Newell Blvd. S., Ft. Worth, TX 76118, USA

ARTICLE INFO

Article history:
Received 20 June 2013
Received in revised form
25 November 2013
Accepted 21 March 2014
Available online 27 May 2014

Keywords:
Optimal tracking control
Integral reinforcement learning
Input constrainers
Neural networks

ABSTRACT

In this paper, a new formulation for the optimal tracking control problem (OTCP) of continuous-time nonlinear systems is presented. This formulation extends the integral reinforcement learning (IRL) technique, a method for solving optimal regulation problems, to learn the solution to the OTCP. Unlike existing solutions to the OTCP, the proposed method does not need to have or to identify knowledge of the system drift dynamics, and it also takes into account the input constraints a priori. An augmented system composed of the error system dynamics and the command generator dynamics is used to introduce a new nonquadratic discounted performance function for the OTCP. This encodes the input constrains into the optimization problem. A tracking Hamilton–Jacobi–Bellman (HJB) equation associated with this nonquadratic performance function is derived which gives the optimal control solution. An online IRL algorithm is presented to learn the solution to the tracking HJB equation without knowing the system drift dynamics. Convergence to a near-optimal control solution and stability of the whole system are shown under a persistence of excitation condition. Simulation examples are provided to show the effectiveness of the proposed method.

1. Introduction

Reinforcement learning (RL) (Bertsekas & Tsitsiklis, 1996; Powell, 2007; Sutton & Barto, 1998), inspired by learning mechanisms observed in mammals, is concerned with how an agent or actor ought to take actions so as to optimize a cost of its long-term interactions with the environment. The agent or actor learns an optimal policy by modifying its actions based on stimuli received in response to its interaction with its environment. Similar to RL, optimal control involves finding an optimal policy based on optimizing a long-term performance criterion. Strong connections between RL and optimal control have prompted a major effort towards introducing and developing online and model-free RL algorithms to learn the solution to optimal control problems (Lewis & Liu, 2012; Vrabie, Vamvoudakis, & Lewis, 2013; Zhang, Liu, Luo, & Wang, 2012).

During the last few years, RL methods have been successfully used to solve the optimal regulation problems by learning the solution to the so-called Hamilton–Jacobi–Bellman (HJB) equation (Lewis, Vrabie, & Syrmos, 2012; Liu & Wei, 2013). For continuous-time systems, Vrabie and Lewis (2009) and Vrabie, Pastravanu, Abou-Khalaf, and Lewis (2009) proposed a promising RL algorithm, called integral reinforcement learning (IRL), to learn the solution to the HJB equation using only partial knowledge about the system dynamics. They used an iterative online policy iteration (PI) (Howard, 1960) procedure to implement their IRL algorithm. Later, inspired by Vrabie and Lewis (2009) and Vrabie et al. (2009), some online PI algorithms were presented to solve the optimal regulation problem for completely unknown linear systems (Jiang & Jiang, 2012; Lee, Park, & Choi, 2012). Also, in Liu, Yang, and Li (2013) the authors presented an IRL algorithm to find the solution to the HJB equation related to a discounted cost function. Other than the IRL-based PI algorithms, efficient synchronous PI algorithms with guaranteed closed-loop stability were proposed in Bhasin et al. (2012), Modares, Naghibi-Sistani, and Lewis (2013), Vamvoudakis and Lewis (2010), to learn the solution to the HJB equation. Synchronous IRL algorithms were also presented for solving the HJB equation in Modares, Naghibi-Sistani, and Lewis (2014) and Vamvoudakis, Vrabie, and Lewis (in press).

[☆] The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Giancarlo Ferrari-Trecate under the direction of Editor Ian R. Petersen.

E-mail addresses: reza_modares@yahoo.com (H. Modares), lewis@uta.edu (F.L. Lewis).

[1] Tel.: +1 8173818059; fax: +1 985118172233.

http://dx.doi.org/10.1016/j.automatica.2014.05.011
0005-1098/© 2014 Elsevier Ltd. All rights reserved.

Although RL algorithms have been widely used to solve the optimal regulation problems, few results considered solving the optimal tracking control problem (OTCP) for both discrete-time (Dierks & Jagannathan, 2009; Kiumarsi, Lewis, Modares, Karimpour, & Naghibi-Sistani, 2014; Wang, Liu, & Wei, 2012; Zhang, Wei, & Luo, 2008) and continuous-time systems (Dierks & Jagannathan, 2010; Zhang, Cui, Zhang, & Luo, 2011). Moreover, existing methods for continuous-time systems require the exact knowledge of the system dynamics a priori while finding the feedforward part of the control input using the dynamic inversion concept. In order to attain the required knowledge of the system dynamics, in Zhang et al. (2011), a plant model was first identified and then an RL-based optimal tracking controller was synthesized using the identified model. To our knowledge, there has been no attempt to develop RL-based techniques to solve the OTCP for continuous-time systems with unknown or partially-unknown dynamics using only measured data in real time. While the importance of the IRL algorithm is well understood for solving optimal regulation problems using only partial knowledge of the system dynamics, the requirement of the exact knowledge of the system dynamics for finding the steady-state part of the control input in the existing OTCP formulation does not allow extending the IRL algorithm for solving the OTCP.

Another important issue which is ignored in the existing RL-based solutions to the OTCP is the amplitude limitation on the control inputs. In fact, in the existing formulation for the OTCP, it is not possible to encode the input constraints into the optimization problem a priori, as only the cost of the feedback part of the control input is considered in the performance function. Therefore, the existing RL-based solutions to the OTCP offer no guarantee on the remaining control inputs on their permitted bounds during and after learning. This may result in performance degradation or even system instability. In the context of the constrained optimal regulation problem, however, an offline PI algorithm (Abou-Khalaf & Lewis, 2005) and online PI algorithms (Modares et al., 2013, 2014) were presented to find the solution to the constrained HJB equation.

In this paper, we develop an online adaptive controller based on the IRL technique to learn the OTCP solution for nonlinear continuous-time systems without knowing the system drift dynamics or the command generator dynamics. The contributions of this paper are as follows. First, a new formulation for the OTCP is presented. In fact, an augmented system is constructed from the tracking error dynamics and the command generator dynamics to introduce a new discounted performance function for the OTCP. Second, the input constraints are encoded into the optimization problem a priori by employing a suitable nonquadratic performance function. Third, a tracking HJB equation related to this nonquadratic performance function is derived which gives both feedforward and feedback parts of the control input simultaneously. Fourth, the IRL algorithm is extended for solving the OTCP. An IRL algorithm, implemented on an actor–critic structure, is used to find the solution to the tracking HJB equation online using only partial knowledge about the system dynamics. In contrast to the existing work, a preceding identification procedure is not needed and the optimal policy is learned using only measured data from the system. Convergence of the proposed learning algorithm to a near-optimal control solution and the boundness of the tracking error and the actor and critic NNs weights during learning are also shown.

2. Optimal tracking control problem (OTCP)

In this section, a review of the OTCP for continuous-time nonlinear systems is given. It is pointed out that the standard solution to the given problem requires complete knowledge of the system dynamics. It is also pointed out that the input constraints caused by the actuator saturation cannot be encoded into the standard performance function a priori. A new formulation of the OTCP problem is given in the next section to overcome these shortcomings.

2.1. Problem formulation

Consider the affine CT dynamical system described by

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

where $x \in \mathbb{R}^n$ is the measurable system state vector, $f(x) \in \mathbb{R}^n$ is the drift dynamics of the system, $g(x) \in \mathbb{R}^{n\times m}$ is the input dynamics of the system, and $u(t) \in \mathbb{R}^m$ is the control input. The elements of $u(t)$ are defined by $u_i(t),\ i = 1, \ldots, m$.

Assumption 1. It is assumed that $f(0) = 0$ and $f(x)$ and $g(x)$ are Lipschitz, and that the system (1) is controllable in the sense that there exists a continuous control on a set $\Omega \subseteq \mathbb{R}^n$ which stabilizes the system.

Assumption 2 (*Bhasin et al., 2012; Vamvoudakis & Lewis, 2010*)**.** The following assumptions are considered on the system dynamics:

(a) $\|f(x)\| \le b_f \|x\|$ for some constant b_f.
(b) $g(x)$ is bounded by a constant b_g, i.e. $\|g(x)\| \le b_g$.

Note that Assumption 2(a) requires $f(x)$ be Lipschitz and $f(0) = 0$ (see Assumption 1) which is a standard assumption to make sure the solution $x(t)$ of the system (1) is unique for any finite initial condition. On the other hand, although Assumption 2(b) restricts the considered class of nonlinear systems, many physical systems, such as robotic systems (Slotine & Li, 1991) and aircraft systems (Sastry, 1991) fulfill such a property.

The goal of the optimal tracking problem is to find the optimal control policy $u^*(t)$ so as to make the system (1) track a desired (reference) trajectory $x_d(t) \in \mathbb{R}^n$ in an optimal manner by minimizing a predefined performance function. Moreover, the input must be constrained to remain within predefined limits $|u_i(t)| \le \lambda,\ i = 1, \ldots, m$.

Define the tracking error as

$$e_d(t) \triangleq x(t) - x_d(t). \tag{2}$$

A general performance function leading to the optimal tracking controller can be expressed as

$$V(e_d(t), x_d(t)) = \int_t^\infty e^{-\gamma(\tau - t)}[E(e_d(\tau)) + U(u(\tau))]\,d\tau \tag{3}$$

where $E(e_d)$ is a positive-definite function, $U(u)$ is a positive-definite integrand function, and γ is the discount factor.

Note that the performance function (3) contains both the tracking error cost and the whole control input energy cost. The following assumption is made in accordance to other work in the literature.

Assumption 3. The desired reference trajectory $x_d(t)$ is bounded and there exists a Lipschitz continuous command generator function $h_d(x_d(t)) \in \mathbb{R}^n$ such that

$$\dot{x}_d(t) = h_d(x_d(t)) \tag{4}$$

and $h_d(0) = 0$.

Note that the reference dynamics needs only to be stable in the sense of Lyapunov, not necessarily asymptotically stable.