Automatica 51 (2015) 40–48

Contents lists available at ScienceDirect

Automatica

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

Brief paper

Approximate optimal trajectory tracking for continuous-time nonlinear systems[☆]

Rushikesh Kamalapurkar[a], Huyen Dinh[b], Shubhendu Bhasin[c], Warren E. Dixon[a]

[a] Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, USA
[b] Department of Mechanical Engineering, University of Transport and Communications, Hanoi, Viet Nam
[c] Department of Electrical Engineering, Indian Institute of Technology, Delhi, India

ARTICLE INFO

Article history:
Received 21 September 2012
Received in revised form
23 May 2014
Accepted 29 September 2014

Keywords:
Time-varying systems
Optimal control
Adaptive control
Tracking applications
Actor–critic

ABSTRACT

Adaptive dynamic programming has been investigated and used as a method to approximately solve optimal regulation problems. However, the extension of this technique to optimal tracking problems for continuous-time nonlinear systems has remained a non-trivial open problem. The control development in this paper guarantees ultimately bounded tracking of a desired trajectory, while also ensuring that the enacted controller approximates the optimal controller.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Reinforcement learning (RL) is a concept that can be used to enable an agent to learn optimal policies from interaction with the environment. The objective of the agent is to learn the policy that maximizes or minimizes a cumulative long term reward. Almost all RL algorithms use some form of generalized policy iteration (GPI). GPI is a set of two simultaneous interacting processes, policy evaluation and policy improvement. Starting with an estimate of the state value function and an admissible policy, policy evaluation makes the estimate consistent with the policy and policy improvement makes the policy greedy with respect to the value function. These algorithms exploit the fact that the optimal value function satisfies Bellman's principle of optimality (Kirk, 2004; Sutton & Barto, 1998).

When applied to continuous-time systems the principle of optimality leads to the Hamilton–Jacobi–Bellman (HJB) equation which is the continuous-time counterpart of the Bellman equation (Doya, 2000). Similar to discrete-time adaptive dynamic programming (ADP), continuous-time ADP approaches aim at finding approximate solutions to the HJB equation. Various methods to solve this problem are proposed in Abu-Khalaf and Lewis (2002), Beard, Saridis, and Wen (1997), Bhasin et al. (2013), Jiang and Jiang (2012), Vamvoudakis and Lewis (2010), Vrabie and Lewis (2009) and Zhang, Luo, and Liu (2009) and the references therein. An infinite horizon regulation problem with a quadratic cost function is the most common problem considered in ADP literature. For these problems, function approximation techniques can be used to approximate the value function because it is time-invariant.

Approximation techniques like neural networks (NNs) are commonly used in ADP literature for value function approximation. ADP-based approaches are presented in results such as (Dierks & Jagannathan, 2010; Zhang, Cui, Zhang, & Luo, 2011) to address the tracking problem for continuous-time systems, where the value function, and the controller presented are time-varying functions of the tracking error. However, for the infinite horizon optimal control problem, time does not lie on a compact set, and NNs can only approximate functions on a compact domain. Thus, it is unclear how a NN with the tracking error as an input can approximate the time-varying value function and controller.

For discrete-time systems, several approaches have been developed to address the tracking problem. Park, Choi, and Lee

[☆] This research is supported in part by NSF award numbers 1161260, 1217908, ONR grant number N00014-13-1-0151, and a contract with the AFRL Mathematical Modeling and Optimization Institute. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsoring agency. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Antonio Loria under the direction of Editor Andrew R. Teel.

E-mail addresses: rkamalapurkar@ufl.edu (R. Kamalapurkar), huyendtt214@gmail.com (H. Dinh), sbhasin@ee.iitd.ac.in (S. Bhasin), wdixon@ufl.edu (W.E. Dixon).

http://dx.doi.org/10.1016/j.automatica.2014.10.103
0005-1098/© 2014 Elsevier Ltd. All rights reserved.

(1996) use generalized back-propagation through time to solve a finite horizon tracking problem that involves offline training of NNs. An ADP-based approach is presented in Dierks and Jagannathan (2009) to solve an infinite horizon optimal tracking problem where the desired trajectory is assumed to depend on the system states. Greedy heuristic dynamic programming based algorithms are presented in results such as (Luo & Liang, 2011; Wang, Liu, & Wei, 2012; Zhang, Wei, & Luo, 2008) which transform the nonautonomous system into an autonomous system, and approximate convergence of the sequence of value functions to the optimal value function is established. However, these results lack an accompanying stability analysis.

In this result, the tracking error and the desired trajectory both serve as inputs to the NN. This makes the developed controller fundamentally different from previous results, in the sense that a different HJB equation must be solved and its solution, i.e. the feedback component of the controller, is a time-varying function of the tracking error. In particular, this paper addresses the technical obstacles that result from the time-varying nature of the optimal control problem by including the partial derivative of the value function with respect to the desired trajectory in the HJB equation, and by using a system transformation to convert the problem into a time-invariant optimal control problem in such a way that the resulting value function is a time-invariant function of the transformed states, and hence, lends itself to approximation using a NN. A Lyapunov-based analysis is used to prove ultimately bounded tracking and that the enacted controller approximates the optimal controller. Simulation results are presented to demonstrate the applicability of the presented technique. To gauge the performance of the proposed method, a comparison with a numerical optimal solution is presented.

For notational brevity, unless otherwise specified, the domain of all the functions is assumed to be $\mathbb{R}_{\geq 0}$. Furthermore, time-dependence is suppressed while denoting trajectories of dynamical systems. For example, the trajectory $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is defined by abuse of notation as $x \in \mathbb{R}^n$, and referred to as x instead of $x(t)$, and unless otherwise specified, an equation of the form $f + h(y, t) = g(x)$ is interpreted as $f(t) + h(y(t), t) = g(x(t))$ for all $t \in \mathbb{R}_{\geq 0}$.

2. Formulation of time-invariant optimal control problem

Consider a class of nonlinear control affine systems

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

where $x \in \mathbb{R}^n$ is the state, and $u \in \mathbb{R}^m$ is the control input. The functions $f : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are locally Lipschitz and $f(0) = 0$. The control objective is to track a bounded continuously differentiable signal $x_d \in \mathbb{R}^n$. To quantify this objective, a tracking error is defined as $e \triangleq x - x_d$. The open-loop tracking error dynamics can then be expressed as

$$\dot{e} = f(x) + g(x) u - \dot{x}_d. \tag{1}$$

The following assumptions are made to facilitate the formulation of an approximate optimal tracking controller.

Assumption 1. The function g is bounded, the matrix $g(x)$ has full column rank for all $x \in \mathbb{R}^n$, and the function $g^+ : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ defined as $g^+ \triangleq \left(g^T g\right)^{-1} g^T$ is bounded and locally Lipschitz.

Assumption 2. The desired trajectory is bounded such that $\|x_d\| \leq d \in \mathbb{R}$, and there exists a locally Lipschitz function $h_d : \mathbb{R}^n \to \mathbb{R}^n$ such that $\dot{x}_d = h_d(x_d)$ and $g(x_d) g^+(x_d) (h_d(x_d) - f(x_d)) = h_d(x_d) - f(x_d)$, $\forall t \in \mathbb{R}_{\geq t_0}$.

The steady-state control policy $u_d : \mathbb{R}^n \to \mathbb{R}^m$ corresponding to the desired trajectory x_d is

$$u_d(x_d) = g_d^+ (h_d(x_d) - f_d), \tag{2}$$

where $g_d^+ \triangleq g^+(x_d)$ and $f_d \triangleq f(x_d)$. To transform the time-varying optimal control problem into a time-invariant optimal control problem, a new concatenated state $\zeta \in \mathbb{R}^{2n}$ is defined as (Zhang et al., 2008)

$$\zeta \triangleq \left[e^T, x_d^T\right]^T. \tag{3}$$

Based on (1) and Assumption 2, the time derivative of (3) can be expressed as

$$\dot{\zeta} = F(\zeta) + G(\zeta)\mu, \tag{4}$$

where the functions $F : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $G : \mathbb{R}^{2n} \to \mathbb{R}^{2n\times m}$, and the control $\mu \in \mathbb{R}^m$ are defined as

$$F(\zeta) \triangleq \begin{bmatrix} f(e + x_d) - h_d(x_d) + g(e + x_d) u_d(x_d) \\ h_d(x_d) \end{bmatrix},$$

$$G(\zeta) \triangleq \begin{bmatrix} g(e + x_d) \\ 0 \end{bmatrix}, \qquad \mu \triangleq u - u_d. \tag{5}$$

Local Lipschitz continuity of f and g, the fact that $f(0) = 0$, and Assumption 2 imply that $F(0) = 0$ and F is locally Lipschitz.

The objective of the optimal control problem is to design a policy $\mu^* : \mathbb{R}^{2n} \to \mathbb{R}^m \in \Psi$ such that the control law $\mu = \mu^*(\zeta)$ minimizes the cost functional

$$J(\zeta, \mu) \triangleq \int_0^\infty r(\zeta(\rho), \mu(\rho))\, d\rho,$$

subject to the dynamic constraints in (4), where Ψ is the set of admissible policies (Beard et al., 1997), and $r : \mathbb{R}^{2n} \times \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ is the local cost defined as

$$r(\zeta, \mu) \triangleq \zeta^T \overline{Q} \zeta + \mu^T R \mu. \tag{6}$$

In (6), $R \in \mathbb{R}^{m\times m}$ is a positive definite symmetric matrix of constants, and $\overline{Q} \in \mathbb{R}^{2n\times 2n}$ is defined as

$$\overline{Q} \triangleq \begin{bmatrix} Q & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} \end{bmatrix}, \tag{7}$$

where $Q \in \mathbb{R}^{n\times n}$ is a positive definite symmetric matrix of constants with the minimum eigenvalue $\underline{q} \in \mathbb{R}_{>0}$, and $0_{n\times n} \in \mathbb{R}^{n\times n}$ is a matrix of zeros. For brevity of notation, let $(\cdot)'$ denote $\partial(\cdot)/\partial\zeta$.

3. Approximate optimal solution

Assuming that a minimizing policy exists and that the optimal value function $V^* : \mathbb{R}^{2n} \to \mathbb{R}_{\geq 0}$ defined as

$$V^*(\zeta) \triangleq \min_{\mu(\tau)|\tau\in\mathbb{R}_{\geq t}} \int_t^\infty r\left(\phi^\mu(\tau; t, \zeta), \mu(\tau)\right) d\tau \tag{8}$$

is continuously differentiable, the HJB equation for the optimal control problem can be written as

$$H^* = V^{*\prime}(\zeta)\left(F(\zeta) + G(\zeta)\mu^*(\zeta)\right) + r\left(\zeta, \mu^*(\zeta)\right) = 0, \tag{9}$$

for all ζ, with the boundary condition $V^*(0) = 0$, where H^* denotes the Hamiltonian, and $\mu^* : \mathbb{R}^{2n} \to \mathbb{R}^m$ denotes the optimal policy. In (8) $\phi^\mu(\tau; t, \zeta)$ denotes the trajectory of (4) under the controller μ starting at initial time t and initial state ζ. For the local cost in (6) and the dynamics in (4), the optimal policy can be obtained in closed-form as (Kirk, 2004)

$$\mu^*(\zeta) = -\frac{1}{2} R^{-1} G^T(\zeta) \left(V^{*\prime}(\zeta)\right)^T. \tag{10}$$

Download English Version:

https://daneshyari.com/en/article/7110043

Download Persian Version:

https://daneshyari.com/article/7110043

Daneshyari.com