Automatica 83 (2017) 108–115

Contents lists available at ScienceDirect

Automatica

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

Brief Paper

Drift counteraction optimal control for deterministic systems and enhancing convergence of value iteration[☆]

Robert A.E. Zidek, Ilya V. Kolmanovsky

Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 26 July 2016
Received in revised form 19 April 2017
Accepted 1 May 2017

Keywords:
Optimal control
Constrained nonlinear systems
Approximate dynamic programming
Aerospace applications

ABSTRACT

The paper treats a class of optimal control problems for deterministic nonlinear discrete-time systems with the objective of maximizing the time or total yield until prescribed constraints are violated. Such problems are referred to as drift counteraction optimal control (DCOC) problems as the corresponding control policy may be viewed as optimally counteracting drift imposed by disturbances or system dynamics. We derive conditions for the existence of an optimal solution. The optimal control policy is characterized by the value function and a new algorithm based on proportional feedback is presented that obtains the value function faster than conventional dynamic programming algorithms. In addition, an approximate dynamic programming (ADP) approach using Gaussian process regression is formulated based on the new algorithm. Two numerical examples are reported, a time maximization problem for a van der Pol oscillator and a satellite life extension problem.

1. Introduction

Let $x_t \in \mathbb{R}^n$ be the state vector of a dynamic system at a discrete time instant $t \in \mathbb{Z}_+$ and $\pi : \mathbb{R}^n \to U$ an admissible control policy. We consider a class of exit-time problems for deterministic nonlinear discrete-time systems, where the first exit-time of x from a prescribed set $G \subset \mathbb{R}^n$, given the initial state vector x_0 and control policy π, is defined as follows

$$\tau(x_0, \pi) = \inf\left\{t \in \mathbb{Z}_{\geq 0} : x_t \notin G | x_0 \in G, \pi \in \Pi\right\}, \tag{1}$$

with Π as the set of admissible control policies. The optimal control problem is given by

$$J(x_0, \pi) = \sum_{t=0}^{\tau(x_0,\pi)-1} g(x_t, u_t) \to \max_{\pi \in \Pi} \qquad \text{subject to } x_{t+1} = f(x_t, u_t),\ x_0 \in G, \tag{2}$$

where $g : G \times U \to \mathbb{R}_+$ is the instantaneous yield and $u_t = \pi(x_t) \in U \subset \mathbb{R}^p$ is the control input vector at a time instant t. We refer to such problems as *drift counteraction optimal control* (DCOC) problems since the optimal control policy is counteracting drift imposed by system dynamics or disturbances in order to maximize J. Note that if $g \equiv 1$ in (2), the objective is to maximize the first exit-time from G. DCOC problems can be found in many engineering applications, in particular, those where resources (fuel, energy, component life, etc.) are finite. For example, Kolmanovsky and Filev (2009) applied DCOC to adaptive cruise control and hybrid electric vehicle energy management and Zidek and Kolmanovsky (2015) used DCOC to maximize the lifetime of a satellite.

The approach in this paper to solve (2) is based on dynamic programing (DP), where the optimal control policy is characterized by the value function V, which is defined by

$$V(x) = \sup_{\pi \in \Pi} J(x, \pi). \tag{3}$$

The DCOC problem for stochastic systems was considered by Kolmanovsky, Lezhnev, and Maizenberg (2008). They showed that the value iteration (VI) algorithm converges to the value function if an optimal solution exists. Zidek and Kolmanovsky (2015) applied DCOC to deterministic systems and introduced proportional feedback VI to obtain the value function. Numerical examples showed that proportional feedback VI converges to the value function faster than conventional VI. The present paper extends significantly our previous conference paper (Zidek & Kolmanovsky, 2015). In particular, we present additional results and applications, details of the proofs, and discussions. The main contributions are the derivation of conditions for the existence of a solution to (2) and the analysis of convergence of proportional feedback VI, theoretically proving convergence for any proportional gain between 0 and 2. Furthermore, we provide an explanation for gains different

[☆] This research is supported by the National Science Foundation Award Number EECS 1404814. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Andrey V. Savkin under the direction of Editor Ian R. Petersen.

E-mail addresses: robzidek@umich.edu (R.A.E. Zidek), ilya@umich.edu (I.V. Kolmanovsky).

http://dx.doi.org/10.1016/j.automatica.2017.06.015

than 1 providing faster convergence when VI is implemented approximately. We consider a discrete-mesh-based approximation and a Kriging-based approximate dynamic programming (ADP) implementation and we present numerical examples which confirm improved convergence properties with the proper selection of the gain differently from the ideally optimal value of 1.

Problems similar to DCOC, subject to stochastic continuous-time systems, were treated by Bayraktar, Song, and Yang (2010), Buckdahn and Nie (2016), Fleming and Soner (2006), Gorodetsky, Karaman, and Marzouk (2015), Kolmanovsky and Maizenberg (2002) and Lions (1983b). It was shown that, under suitable assumptions, the optimal control and its corresponding value function satisfy the Hamilton–Jacobi–Bellman (HJB) equation in the viscosity sense, where the HJB equation is a second-order partial differential equation (PDE) for the value function (Lions, 1983a). Related problems for deterministic continuous-time systems were treated by Barles and Perthame (1988), Blanc (1997), Cannarsa, Pignotti, and Sinestrari (2000), Malisoff (2002) and Rungger and Stursberg (2011). As for the stochastic case, the value function was shown to be a weak solution of the HJB equation, which is a first order PDE in the deterministic case.

Explicit solutions to the HJB equation only exist for some special problems. Otherwise, a solution can only be obtained approximately using numerical methods. Therefore, as also noted in Kolmanovsky et al. (2008), in contrast to the continuous-time treatment of the problem and solving a PDE numerically, the formulation of the problem in discrete-time appears to be computationally more tractable for determining the value function. In fact, numerical schemes for solving the HJB equation require both a time and state space discretization, where the VI algorithm may be used to solve the discretized problem (Barles & Souganidis, 1991; Kushner & Dupuis, 2013; Rungger & Stursberg, 2011).

The structure of the paper is as follows. A characterization of the optimal solution and existence conditions are given in Section 2. The computation of the optimal control policy, including a Kriging-based ADP approach, is discussed in Section 3. Section 4 presents two numerical examples of maximizing the time until a van der Pol oscillator violates constraints and of maximizing the lifetime of a satellite in low Earth orbit (LEO). A conclusion is given in Section 5.

2. Characterization of optimal solution

We make the following assumption about g.

Assumption 1. There exists a real-valued $\bar{g} > 0$ such that $g(x, u) \leq \bar{g}$ for all $(x, u) \in G \times U$.

Theorem 1 provides conditions under which the total yield and the value function are bounded. It is based on the following assumption about $\tau(x, \pi)$.

Assumption 2. There exists an integer $\bar{T} > 0$ such that $\tau(x, \pi) \leq \bar{T}$ for all $x \in G$ and $\pi \in \Pi$.

This assumption is reasonable in DCOC problems in which every trajectory will eventually violate the constraints and the objective is either to delay this event or to maximize yield before it happens. This is the case, for example, in applications where resources such as fuel are limited, see Section 1, or where insufficient control authority is available.

Theorem 1. *Suppose Assumptions 1 and 2 hold. Then there exists $\bar{V} > 0$ such that $J(x, \pi) \leq V(x) \leq \bar{V}$ for all $x \in G$ and $\pi \in \Pi$.*

Proof. Let $x = x_0 \in G$ be any given state and $\pi \in \Pi$. Using Assumptions 1 and 2, we get

$$J(x, \pi) = \sum_{t=0}^{\tau(x,\pi)-1} g(x_t, u_t) \leq \sum_{t=0}^{\tau(x,\pi)-1} \bar{g} \leq \bar{T}\bar{g}. \tag{4}$$

This and (3) imply that $V(x) \leq \bar{V} = \bar{T}\bar{g}$. □

The next theorem provides sufficient conditions for a control policy to be optimal.

Theorem 2. *Suppose Assumptions 1 and 2 hold and let $L^{\pi}V(x) = V(x) - V(f(x, \pi(x)))$. Then $\pi^* \in \Pi$ satisfies*

$$\begin{aligned} &L^{\pi^*}V(x) = g(x, \pi^*(x)), \quad \text{if } x \in G, \\ &L^{\pi}V(x) \geq g(x, \pi(x)), \quad \text{if } x \in G, \quad \pi \neq \pi^*, \\ &V(x) = 0, \quad \text{if } x \notin G, \end{aligned} \tag{5}$$

for all $x \in \mathbb{R}^n$ and $\pi \in \Pi$ if and only if π^* maximizes $J(x, \pi)$ for all $x \in G$. Furthermore, $V(x) = J(x, \pi^*)$ and

$$\pi^*(x) \in \Pi^*(x) = \arg\max_{u \in U} \{g(x, u) + V(f(x, u))\}. \tag{6}$$

Proof. Since $J(x, \pi) = 0$ for all $x \notin G$, $V(x) = 0$ for all $x \notin G$. Now let $x = x_0 \in G$ be any given state and $\pi \in \Pi$. For the first part of the proof, assume π^* satisfies (5). Thus, we have

$$\begin{aligned} J(x, \pi) &= \sum_{t=0}^{\tau(x,\pi)-1} g(x_t, \pi(x_t)) \\ &\leq \sum_{t=0}^{\tau(x,\pi)-1} L^{\pi}V(x_t) = V(x), \end{aligned} \tag{7}$$

since $V(x_{\tau(x,\pi)}) = 0$ due to $x_{\tau(x,\pi)} \notin G$. Similarly,

$$\begin{aligned} J(x, \pi^*) &= \sum_{t=0}^{\tau(x,\pi^*)-1} g(x_t, \pi^*(x_t)) \\ &= \sum_{t=0}^{\tau(x,\pi^*)-1} L^{\pi^*}V(x_t) = V(x). \end{aligned} \tag{8}$$

We can compare (7) and (8) because V is bounded by Theorem 1, which shows that $J(x, \pi^*) \geq J(x, \pi)$. It immediately follows from (5) that $\pi^*(x) \in \Pi^*(x)$ according to (6). For the second part of the proof, assume that π^* maximizes $J(x, \pi)$ for all $x \in G$. Then, by (3), $V(x) = J(x, \pi^*)$ for all $x \in G$. This implies

$$\begin{aligned} V(x) &= g(x, \pi^*(x)) + J(f(x, \pi^*(x)), \pi^*) \\ &= g(x, \pi^*(x)) + V(f(x, \pi^*(x))). \end{aligned} \tag{9}$$

Since $V(x)$ is the optimal value, it follows that, for any admissible policy $\pi \neq \pi^*$,

$$\begin{aligned} V(x) &\geq g(x, \pi(x)) + J(f(x, \pi(x)), \pi^*) \\ &= g(x, \pi(x)) + V(f(x, \pi(x))). \quad \square \end{aligned} \tag{10}$$

Remark 1. The optimal solution π^* to (2), if exists, may not be unique. In case of non-uniqueness, additional criteria, for instance, minimizing 2-norm, may be used for selecting the control from the set of maximizers in (6).

Theorem 3. *If an optimal solution to (2) exists, V is the unique solution to (5).*

Proof. Suppose $\pi^* \in \Pi$ is an optimal solution to (2). Furthermore, suppose that, in addition to V, another function $\hat{V}$ satisfies (5). It follows from the proof of Eqs. (2) and (8) that, for all $x \in G$, $V(x) = J(x, \pi^*)$ and $\hat{V}(x) = J(x, \pi^*)$, which implies $\hat{V} = V$. □

The existence of an optimal solution to (2) can be studied using the set $\Pi^*(x)$.

Theorem 4. *An optimal solution $\pi^* \in \Pi$ to (2) exists for all $x \in G$ if and only if the set $\Pi^*(x)$ defined in (6) is nonempty for all $x \in G$.*