Automatica 87 (2018) 159–165

Contents lists available at ScienceDirect

Automatica

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

Brief Paper

Stochastic optimal control via forward and backward stochastic differential equations and importance sampling ☆

Ioannis Exarchos, Evangelos A. Theodorou

School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0150, United States

ARTICLE INFO

Article history:
Received 1 December 2015
Received in revised form 2 May 2017
Accepted 28 August 2017
Available online 7 November 2017

Keywords:
Stochastic control
Forward and backward stochastic differential equations
Importance sampling

ABSTRACT

The aim of this work is to present a novel sampling-based numerical scheme designed to solve a certain class of stochastic optimal control problems, utilizing forward and backward stochastic differential equations (FBSDEs). By means of a nonlinear version of the Feynman–Kac lemma, we obtain a probabilistic representation of the solution to the nonlinear Hamilton–Jacobi–Bellman equation, expressed in the form of a system of decoupled FBSDEs. This system of FBSDEs can be solved by employing linear regression techniques. The proposed framework relaxes some of the restrictive conditions present in recent sampling based methods within the Linearly Solvable Optimal Control framework, and furthermore addresses problems in which the time horizon is not prespecified. To enhance the efficiency of the proposed scheme when treating more complex nonlinear systems, we then derive an iterative algorithm based on Girsanov's theorem on the change of measure, which features importance sampling. This scheme is shown to be capable of learning the optimal control without requiring an initial guess.

1. Introduction

By and large, the problem of obtaining an optimal control is associated with the solution of a generally nonlinear, second-order partial differential equation (PDE), known as the Hamilton–Jacobi–Bellman (HJB) equation. A classification of different available methods can be achieved based on whether the solution of this PDE is sought for over the entire domain, or locally around a nominal system trajectory. In the first case, several attempts have been made to address the difficulty inherent in solving such nonlinear PDEs, as well as the curse of dimensionality, with various different methods and approaches (Beard, Saridis, & Wen, 1997; Lasserre, Henrion, Prieur, & Trelat, 2008; McEneaney, 2007) for deterministic control problems, while a stochastic setting is considered in Gorodetsky, Karaman, and Marzouk (2015), Horowitz and Burdick (2014) and Horowitz, Damle, and Burdick (2014). With only but a few exceptions, most of these methods suffer from the curse of dimensionality. On the other hand, the latter category of local methods includes traditional approaches such as Stochastic Differential Dynamic Programing (S-DDP) (Theodorou, Tassa, & Todorov, 2010; Todorov & Li, 2005), which is based on linearization of the dynamics and a quadratic approximation of the value function around nominal trajectories, as well as sampling-based methods.

Sampling-based methods, within stochastic control, rely on a probabilistic representation of the solution to linear backward PDEs. This probabilistic representation is addressed by forward sampling of state trajectories via Stochastic Differential Equations (SDEs), and the numerical evaluation of expectations. Several results based on this framework appear in the literature under the names of Path Integral (PI) Control (Kappen, 2005; Theodorou et al., 2010), Kullback–Leibler (KL) Control, or Linearly Solvable Optimal Control (LSOC) (Dvijotham & Todorov, 2012; Todorov, 2009). These methods have become an exceedingly popular approach to solve nonlinear stochastic optimal control problems due to their ability to accommodate scalable iterative schemes. Their fundamental characteristic is that they rely on the exponential transformation of the value function; under the exponential transformation, and by introducing certain restrictions between control authority, cost and stochasticity, there exists a direct relationship between the HJB PDE and the backward Chapman–Kolmogorov PDE. The latter PDE, being linear, permits then the use of the linear Feynman–Kac lemma (Karatzas & Shreve, 1991), which relates backward linear PDEs to forward SDEs. Thus, the corresponding optimal control problem can be solved using forward sampling. While forward sampling-based methods exhibit several advantages against traditional methods of stochastic control, such as the mild conditions on the differentiability of the cost and the stochastic dynamics,

☆ This research was supported by the Army Research Office grant ARO W911NF-16-1-0390. The material in this paper was partially presented at the 2016 American Control Conference, July 6–8, 2016, Boston, MA, USA. This paper was recommended for publication in revised form by Associate Editor Valery Ugrinovskii under the direction of Editor Ian R. Petersen.

E-mail addresses: exarchos@gatech.edu (I. Exarchos), evangelos.theodorou@ae.gatech.edu (E.A. Theodorou).

https://doi.org/10.1016/j.automatica.2017.09.004
0005-1098/

there are also some key disadvantages which pertain to the nature of the exponential transformation. In particular, the effect of the exponential transformation can be identified as the mapping of the value function $v(t, x)$, which has range $[0, \infty)$, to the desirability function $\psi(t, x)$, whose range is $(0, 1]$. This mapping leads to a drastic reduction in the ability to distinguish states with high cost (low desirability) from states with low cost (high desirability). This issue has been partially addressed with renormalization of the trajectory cost (Theodorou et al., 2010). Finally, while the necessary constraint introduced between control authority and stochasticity can lead to symmetry breaking phenomena and delayed decision (Kappen, 2005), it is a rather restrictive assumption whenever applications to engineered systems are considered.

In this work, we present a learning control algorithm which capitalizes on the innate relationship between certain nonlinear PDEs and Forward and Backward SDEs, demonstrated by a *nonlinear* Feynman–Kac lemma. By means of this lemma, we obtain a probabilistic representation of the solution to the nonlinear HJB equation, expressed in the form of a system of decoupled FBSDEs. This system of FBSDEs can be solved by employing linear regression techniques. To enhance the efficiency of the proposed scheme when treating more complex nonlinear systems, we then derive an iterative algorithm based on Girsanov's theorem on the change of measure, which features importance sampling for the case of FBSDEs. A brief summary of some of the contents of this work has been published by the authors in Exarchos and Theodorou (2016). The herein proposed framework has also been extended to differential games and risk-sensitive control in Exarchos, Theodorou, and Tsiotras (2016).

2. Problem statement

Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete, filtered probability space on which a p-dimensional standard Brownian motion W_t is defined, such that $\{\mathscr{F}_t\}_{t\geq 0}$ is the natural filtration of W_t augmented by all $\mathbb{P}$-null sets. Consider the problem of minimizing the expected cost defined by the cost functional

$$J(\tau, x_\tau; u(\cdot)) = \mathbb{E}\Big[g(x(T)) + \int_\tau^T q(t, x(t)) + \frac{1}{2}u^\top(t)Ru(t)\mathrm{d}t\Big], \tag{1}$$

associated with the stochastic controlled system, which is represented by the Itô stochastic differential equation (SDE)

$$\begin{cases} \mathrm{d}x(t) = f(t, x(t))\mathrm{d}t + G(t, x(t))u(t)\mathrm{d}t + \Sigma(t, x(t))\mathrm{d}W_t, \\ t \in [\tau, T], \qquad x(\tau) = x_\tau, \end{cases} \tag{2}$$

with $T > \tau \geq 0$, wherein T is a fixed time of termination (this requirement will be relaxed in Section 4), $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^\nu$ is the control vector, and R is a $\nu \times \nu$ positive definite matrix. The functions $g : \mathbb{R}^n \to \mathbb{R}$, $q : [0, T] \times \mathbb{R}^n \to \mathbb{R}$, $f : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $G : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times\nu}$, and $\Sigma : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n\times p}$ are deterministic, that is, they do not depend explicitly on $\omega \in \Omega$, continuous w.r.t. time t (in case there is explicit dependence), Lipschitz (uniformly in t) with respect to the state variables, and uniformly bounded w.r.t. time t. These standard assumptions (Yong & Zhou, 1999) guarantee that the SDE solution is unique and does not have a finite escape time, similar to the case of ordinary differential equations, in addition to a well-defined cost functional (1). Furthermore, the square-integrable process $u : [0, T] \times \Omega \to U \subseteq \mathbb{R}^\nu$ is $\{\mathscr{F}_t\}_{t\geq 0}$-adapted (also called *progressively measurable*), which essentially translates into the control input being non-anticipating, i.e., relying only on past and present information. For any given initial condition (τ, x_τ), we wish to minimize (1) under all admissible functions. The solution is obtained by solving the associated Hamilton–Jacobi–Bellman (HJB) equation for the *Value function*. Specifically, by applying the stochastic version of Bellman's principle of optimality, it is shown (Fleming & Soner, 2006; Yong & Zhou, 1999) that if the Value function is in $C^{1,2}([0, T] \times \mathbb{R}^n)$, then it is a solution to the following terminal value problem of a second-order partial differential equation, known as the HJB equation, which, for the problem at hand, and suppressing function arguments for notational compactness, takes the form

$$\begin{cases} v_t + \inf\limits_{u\in U}\Big\{\frac{1}{2}\mathrm{tr}(v_{xx}\Sigma\Sigma^\top) + v_x^\top(f + Gu) + q + \frac{1}{2}u^\top Ru\Big\} \\ = 0,\ (t, x) \in [0, T) \times \mathbb{R}^n, \quad v(T, x) = g(x), \quad x \in \mathbb{R}^n \end{cases} \tag{3}$$

where v_x and v_{xx} denote the gradient and the Hessian of v, respectively. The term inside the brackets is the Hamiltonian. Note that this result can be extended to include cases where the Value function does not satisfy the smoothness condition. Then, if one also considers viscosity solutions of (3), the Value function is proven to be a viscosity solution of (3). Furthermore, the viscosity solution is equal to the classical solution, if a classical solution exists. For the chosen form of the cost integrand, and assuming that the optimal control lies in the interior of U, we may carry out the infimum operation by taking the gradient of the Hamiltonian with respect to u and setting it equal to zero, thus obtaining

$$u^*(t, x) = -R^{-1}G^\top(t, x)v_x(t, x), \qquad (t, x) \in [0, T] \times \mathbb{R}^n. \tag{4}$$

Inserting the above expression back into the original HJB equation and suppressing function arguments for notational brevity, we obtain the equivalent characterization

$$\begin{cases} v_t + \frac{1}{2}\mathrm{tr}(v_{xx}\Sigma\Sigma^\top) + v_x^\top f + q - \frac{1}{2}v_x^\top GR^{-1}G^\top v_x = 0, \\ v(T, x) = g(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{5}$$

3. A Feynman–Kac type representation through FBSDEs

There is an innate relation between stochastic differential equations and second-order partial differential equations (PDEs) of parabolic or elliptic type. Specifically, solutions to a certain class of nonlinear PDEs can be represented by solutions to forward–backward stochastic differential equations (FBSDEs), in the same spirit as demonstrated by the well-known Feynman–Kac formulas (Karatzas & Shreve, 1991) for linear PDEs. In what follows, we shall briefly state the definitions of forward and backward processes, and then proceed to link their solution with the solution of PDEs, in light of a nonlinear Feynman–Kac formula. As a forward process we shall define the square-integrable, $\{\mathscr{F}_s\}_{s\geq 0}$-adapted process $X(\cdot)$[1], which, for any given initial condition $(t, x) \in [0, T] \times \mathbb{R}^n$, satisfies the Itô FSDE

$$\begin{cases} \mathrm{d}X_s = b(s, X_s)\mathrm{d}s + \Sigma(s, X_s)\mathrm{d}W_s, \qquad s \in [t, T], \\ X_t = x. \end{cases} \tag{6}$$

The forward process (6) is also called the *state process* in the literature. We shall denote the solution to the forward SDE (6) as $X_s^{t,x}$, wherein (t, x) are the initial condition parameters. In contrast to the forward process, the associated backward process is the square-integrable, $\{\mathscr{F}_s\}_{s\geq 0}$-adapted pair $(Y(\cdot), Z(\cdot))$ defined via a BSDE satisfying a terminal condition

$$\begin{cases} \mathrm{d}Y_s = -h(s, X_s^{t,x}, Y_s, Z_s)\mathrm{d}s + Z_s^\top \mathrm{d}W_s \qquad s \in [t, T], \\ Y_T = g(X_T). \end{cases} \tag{7}$$

The function $h(\cdot)$ is called *generator* or *driver*. The solution is implicitly defined by the initial condition parameters (t, x) of the FSDE since it obeys the terminal condition $g(X_T^{t,x})$, and thus we will similarly use the notation $Y_s^{t,x}$ and $Z_s^{t,x}$ to denote the solution

[1] While X is a function of s and ω, we shall use X_s for notational brevity.