View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Radboud Repository

Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/144515

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

Path integral control and state-dependent feedback

Sep Thijssen* and H. J. Kappen†
Department of Neurophysics, Donders Institute for Neuroscience, Radboud University, Nijmegen, The Netherlands
(Received 13 June 2014; revised manuscript received 19 September 2014; published 2 March 2015)

In this paper we address the problem of computing state-dependent feedback controls for path integral control problems. To this end we generalize the path integral control formula and utilize this to construct parametrized state-dependent feedback controllers. In addition, we show a relation between control and importance sampling: Better control, in terms of control cost, yields more efficient importance sampling, in terms of effective sample size. The optimal control provides a zero-variance estimate.

DOI: 10.1103/PhysRevE.91.032104 PACS number(s): 02.50.Ey, 02.30.Yy, 05.10.Ln, 05.10.Gg

I. INTRODUCTION

Control methods are used widely in many engineering applications, such as mechanical systems, chemical plants, finance, and robotics. Often, these methods are used to stabilize the system around a particular set point or trajectory using state feedback. In robotics, the problem may be to plan a sequence of actions that yield a motor behavior such as walking or grasping an object [1,2]. In finance, the problem may be to devise a sequence of buy and sell actions to optimize a portfolio of assets, or to determine the optimal option price [3].

Optimal control theory provides an elegant mathematical framework for computing an optimal controller using the Hamilton-Jacobi-Bellman (HJB) equation. In general the HJB equation is impossible to solve analytically, and numerical solutions are intractable due to the problem of dimensionality. As a result, often a suboptimal linear feedback controller such as a proportional-integral-derivative (PID) controller [4] or another heuristic approach is used instead. The use of suboptimal controllers may be particularly problematic for nonlinear stochastic problems, where noise affects the optimality of the controller.

One way to proceed is to consider the class of control problems in which the HJB equation can be linearized. Such problems can be divided into two closely related cases [5]. The first considers infinite-time-average cost problems, while the second considers finite-time problems. Approaches of the first kind [2,6] solve the control problem as an eigenvalue problem. This class has the advantage that the solution also computes a feedback signal, but the disadvantage that a discrete representation of the state space is required. In the second case the optimal control solution is given as a path integral [7]. This case will be the subject of this work. Path integral approaches have led to efficient computational methods that have been successfully applied to multiagent systems and robot movement [1,8–11].

Despite the success of the method, two key aspects have apparently not yet been addressed.

(1) The issue of state feedback has been largely ignored in path integral approaches and the resulting "open-loop" controllers are independent of the state; they are possibly augmented with an additional PID controller to ensure stability [1].

(2) The path integral is computed using Monte Carlo sampling. The use of an exploring control as a type of importance sampling has been suggested to improve the efficiency of the sampling [3,12] but there appear to be no theoretical results to back this up.

These two aspects are related because the exploring controls are most effective if they are state feedback controls. In this paper we propose solutions to these two issues. We generalize the path integral control formula and utilize this to construct parametrized state-dependent feedback controllers. In Corollary 4 we show how a feedback controller might be obtained using path integral control computations that can be approximated to arbitrary precision in this way if the parametrization is correct. The parameters for all future times can be computed using a single set of Monte Carlo samples.

We derive the key property that the path integral is independent of the importance sampling when using infinite samples. However, importance sampling strongly affects the efficiency of the sampler. In Theorem 2 we derive a bound which implies that, when the importance control approaches the optimal control, the variance in the estimates reduces to zero and the effective sample size becomes maximal. This allows us to improve the estimates iteratively by using better and better importance sampling with increasing effective sample size.

This work is structured as follows. In Sec. II we review path integral control and we extend the existing theory in Sec. III. Using this we prove additional variance bounds in Sec. IV and generalized path integral control formulas in Sec. V. In Sec. VI we construct a feedback controller and describe how to compute it efficiently. In Sec. VII we show in an example how to compute several nonlinear feedback controllers for a nonlinear control problem.

II. THE PATH INTEGRAL CONTROL PROBLEM

Consider the dynamical system

$$dX^u(t) = b(t,X^u(t))dt + \sigma(t,X^u(t))[u(t,X^u(t))dt + dW(t)], \quad (1)$$

for $t_0 \leqslant t \leqslant t_1$ and with $X^u(t_0) = x_0$. Here $W(t)$ is m-dimensional standard Brownian motion, and we take $b : [t_0,t_1] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma : [t_0,t_1] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$, and $u : [t_0,t_1] \times \mathbb{R}^n \to \mathbb{R}^m$ such that a solution of Eq. (1) exists. Formulating exact conditions that guarantee existence is not

*s.thijssen@donders.ru.nl
†b.kappen@science.ru.nl; http://www.snn.ru.nl/~bertk

 ©2015 American Physical Society

the aim of this work. (See [13,14] for details of the theory, or [15] for a mathematical approach to path integral control.)

Given a function $u(t,x)$ that defines the control for each state x and each time $t_0 \leqslant t \leqslant t_1$, we define the cost

$$S^u(t) = \int_t^{t_1} V(s,X^u(s)) + \frac{1}{2}u(s,X^u(s))'u(s,X^u(s))ds + \int_t^{t_1} u(s,X^u(s))'dW(s), \quad (2)$$

where the prime denotes the transpose. Note that S depends on future values of X and is therefore not adaptive [13,14] with respect to the Brownian motion.

It is unusual to include a stochastic integral with respect to Brownian motion in the cost because it vanishes when taking the expectation value. However, when performing importance sampling with u, such a term appears naturally (see Sec IV).

The goal in stochastic optimal control is to minimize the expected cost with respect to the control:

$$J(t,x) = \min_u \mathbb{E}[S^u(t) \mid X^u(t) = x],$$
$$u^*(\cdot,\cdot) = \arg\min_u \mathbb{E}[S^u(t_0)].$$

Here $\mathbb{E}$ denotes the expected value with respect to the stochastic process from Eq. (1). The following, previously established, result [7,11] gives a solution of the control problem in terms of path integrals.

Theorem 1. The solution of the control problem is given by

$$J(t_0,x_0) = -\ln \mathbb{E}\, e^{-S^u(t_0)}, \quad (3)$$

$$u^*(t_0,x_0) - u(t_0,x_0) = \lim_{t\to t_0} \frac{\mathbb{E}\left[e^{-S^u(t_0)}\int_{t_0}^t dW(s)\right]}{(t-t_0)\mathbb{E}[e^{-S^u(t_0)}]}. \quad (4)$$

Proof. Equation (3) will be proven in Remark 2 and Eq. (4) follows from the generalized main theorem in Sec. V. ■

Because the solution of the control problem is given in terms of a path integral Eqs. (3) and (4), the control problem Eqs. (1) and (2) is referred to as a path integral control problem. The formulas from Theorem 1 provide a solution at t_0. Of course, since t_0 is arbitrary, this can be utilized at any time t. However, for $t > t_0$, the state $X^u(t)$ is probabilistic, and consequently, the optimal control must be recomputed for each t,x separately. This issue will be partly resolved in the main theorem, where we show that all expected optimal future controls can be expressed using a single path integral.

The optimal control solution holds for any function u. In particular, it holds for $u = 0$ in which case we refer to Eq. (1) as the uncontrolled dynamics. Computing the optimal control in Eq. (4) with $u \neq 0$ implements a type of importance sampling, which is further discussed in Sec. IV.

Remark 1. It is straightforward, but notationally tedious, to generalize the control problem to the following slightly more general form

$$dX = bdt + \sigma(udt + \rho dW),$$
$$S = \Phi(x(T)) + \int_{t_0}^{t_1} V + \frac{1}{2}u'Rudt + \int_{t_0}^{t_1} u'R\rho dW,$$

with $\Phi \in \mathbb{R}$, and $R,\sigma \in \mathbb{R}^{m\times m}$ with $\lambda I = R\rho\rho'$ and $\lambda \in \mathbb{R}_{>0}$. Note that we dropped the dependence on $t, X^u(t)$ for brevity.

III. LINEARIZABLE HJB EQUATION AND STOCHASTIC PROCESSES

In this work we use the HJB equation as a means of solving the control problem. The path integral control problem is characterized by the fact that the HJB equation can be linearized. This will be utilized in this section to obtain the main lemma.

Definition 1. Throughout the rest of this work we define

$$\psi(t,x) = e^{-J(t,x)},$$
$$\psi(t) = \psi(t,X^u(t)),$$
$$\phi(t) = e^{-S^u(t_0)+S^u(t)}.$$

Note that $\psi(\cdot,\cdot)$ denotes a *function* of time and state, while $\phi(\cdot)$ and $\psi(\cdot)$ denote *stochastic processes*, the latter being equal to the function $\psi(\cdot,\cdot)$ of the stochastic process Eq. (1). This convention will also be used for other functions, e.g., $u(t) = u(t,X^u(t))$. We remark that, in contrast to $S^u(t)$, the processes $\psi(t)$ and $\phi(t)$ are adapted: They do not depend on future values of X.

Lemma 1. (main lemma).

$$e^{-S^u(t)} - \psi(t) = \frac{1}{\phi(t)}\int_t^{t_1} \phi(s)\psi(s)[u^*(s) - u(s)]'dW(s). \quad (5)$$

Proof. The HJB equation [14] for the control problem is

$$-J_t = \min_u \left(V + \tfrac{1}{2}u'u + (b+\sigma u)'J_x + \tfrac{1}{2}\mathrm{Tr}(\sigma\sigma' J_{xx})\right),$$

with boundary condition $J(t_1,x) = 0$. We can solve for u which gives

$$u^* = -\sigma' J_x,$$
$$-J_t = V - \tfrac{1}{2}J_x'\sigma\sigma' J_x + b'J_x + \tfrac{1}{2}\mathrm{Tr}(\sigma\sigma' J_{xx}). \quad (6)$$

This partial differential equation (PDE) becomes linear in terms of ψ. We have

$$\psi_t + b'\psi_x + \frac{1}{2}\mathrm{Tr}\sigma\sigma'\psi_{xx} = V\psi,$$
$$u^* = \frac{1}{\psi}\sigma'\psi_x, \quad (7)$$

with boundary condition $\psi(t_1,x) = e^{-J(t_1,x)} = 1$.

Using Itô's lemma [13,14] we obtain a stochastic differential equation (SDE) for the process $\psi(t)$ (dropping the dependence on time for brevity)

$$d\psi = \left(\psi_t + \psi_x'(b+\sigma u) + \tfrac{1}{2}\mathrm{Tr}\sigma\sigma'\psi_{xx}\right)dt + \psi_x'\sigma dW$$
$$= V\psi dt + \psi_x'\sigma(udt + dW),$$

where the last equation follows because $\psi(\cdot,\cdot)$ satisfies Eq. (7). From the definition of ϕ one readily verifies that it satisfies the SDE $d\phi(t) = -\phi(t)[V(t)dt + u(t)'dW(t)]$ with initial condition $\phi(t_0) = 1$. Using the product rule from stochastic

calculus [13] we obtain

$$\begin{aligned} d(\phi\psi) &= \psi d\phi + \phi d\psi + d[\phi,\psi] \\ &= -\phi\psi u' dW + \phi\psi'_x \sigma dW \\ &= \phi\psi(u^* - u)' dW. \end{aligned} \tag{8}$$

Integrating the above from t to t_1 gives

$$\begin{aligned} &\phi(t_1)\psi(t_1) - \phi(t)\psi(t) \\ &\quad = \int_t^{t_1} \phi(s)\psi(s)[u^*(s) - u(s)]' dW(s). \end{aligned}$$

Note that $\psi(t_1) = 1$ and that $\phi(t_1) = \phi(t)e^{-S^u(t)}$. Dividing by $\phi(t)$ we obtain the statement of the lemma. ■

IV. OPTIMAL IMPORTANCE SAMPLING

A Monte Carlo approximation of the optimal control solution Eq. (4) is a weighted average, where the weight depends on the path cost. If the variance of the weights is high, then a lot of samples are required to obtain a good estimate. Critically, Eq. (4) holds for all u, so that it can be chosen to reduce the variance of the path weights. This induces a change of measure and an importance sampling scheme. By the Girsanov theorem [13,14], the change in measure does not affect the weighted average (for a more detailed description in the context of path integral control, see [5]). The Radon-Nikodym derivative $e^{-\int[(1/2)u'udt+u'dW]}$ is the correction term for importance sampling with u, which explains why we included $\int u'dW$ in the definition of S.

In this section we will show that the optimal u for sampling purposes turns out to be u^*. More generally, the variance will decrease as u gets closer to u^*. This motivates policy iteration, in which increasingly better estimates u of u^* improve sampling so that even better approximations of u^* might be obtained.

Definition 2. Given the process $X^u(t)$ for $t_0 < t < t_1$:

(1) The weight of a path is defined as $\alpha^u = \frac{e^{-S^u(t_0)}}{\mathbb{E}[e^{-S^u(t_0)}]}$.

(2) The fraction λ^u of effective samples is $\lambda^u = \frac{1}{\mathbb{E}[(\alpha^u)^2]}$.

Theorem 2. We have the following upper and lower bounds for the variance of the weight:

$$\mathrm{Var}(\alpha^u) \leqslant \int_{t_0}^{t_1} \mathbb{E}[(u^* - u)'(u^* - u)(\alpha^u)^2]dt, \tag{9}$$

$$\mathrm{Var}(\alpha^u) \geqslant \int_{t_0}^{t_1} \mathbb{E}[(u^* - u)\alpha^u]'\mathbb{E}[(u^* - u)\alpha^u]dt. \tag{10}$$

Because $\mathrm{Var}(\alpha^u) + 1 = \mathbb{E}[(\alpha^u)^2]$, the fraction of effective samples as defined in Definition 2.2 satisfies $0 < \lambda^u \leqslant 1$. It has been suggested [16] that this fraction can be used to determine how well one can compute a sample estimate of a weighted average. This can be connected with Theorem 2 as follows.

Corollary 1. If $||u^* - u||^2 \leqslant \epsilon/(t_1 - t_0)$, then

$$\lambda^u \geqslant 1 - \epsilon.$$

Proof. This follows readily from Eq. (9). ■

A numerical illustration of Theorem 2 can be found in Fig. 1. Before we prove Theorem 2, we deduce a few useful facts that follow from the main lemma.

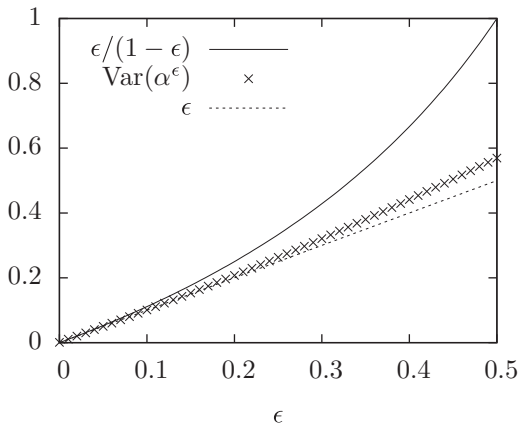

FIG. 1. Estimate of $\mathrm{Var}(\alpha^\epsilon)$, where $\alpha^\epsilon := e^{-S^{u^\epsilon}(t_0)}/\psi(t_0,x_0)$ with upper and lower bounds from Theorem 2 with respect to the control problem in Example 1. Here we considered a range of suboptimal importance controls $u^\epsilon(t,x) = u^*(t,x) + \sqrt{\epsilon}$. The estimate of the variance is based on 10^4 paths that were generated with $dt = 0.001$.

Corollary 2. An optimally controlled random path is an instance of Eq. (1) with $u = u^*$. Although such a path is random, its attributed cost has zero variance and is equal to the expected optimal cost to go:

$$S^{u^*}(t_0) = -\ln\psi(t_0,x_0) = J(t_0,x_0).$$

Furthermore we have $\alpha^{u^*} = 1$, such that the weighted average, which is independent of u, equals the expectation under the optimal process.

Proof. Take $u = u^*$ and $t = t_0$ in Eq. (5). ■

Corollary 3. The following Feynman-Kac formula [13,14] expresses ψ as a path integral:

$$\psi(t) = \mathbb{E}[e^{-S^u(t)} \mid \mathcal{F}_t]. \tag{11}$$

Here the filtration $\mathcal{F}_t$ denotes that we are taking the expected value conditioned on events up to time t.

Proof. Take the expected value on both sides of Eq. (5). ■

Remark 2. When we consider Eq. (11) with $t = t_0$, and take minus the logarithm on both sides, we obtain Eq. (3): a path integral formula for the optimal cost to go function.

Proof of Theorem 2. Consider Eq. (5) with $t = t_0$, and divide by $\psi(t_0,x_0)$ such that

$$\begin{aligned} &\mathrm{Var}(\alpha^u) \\ &= \mathbb{E}\left[\left(\int_{t_0}^{t_1} \frac{\phi(t)\psi(t)}{\psi(t_0)}[u^*(t) - u(t)]' dW(t)\right)^2\right] \\ &= \mathbb{E}\int_{t_0}^{t_1} \frac{\phi(t)^2\psi(t)^2}{\psi(t_0)^2}[u^*(t) - u(t)]'[u^*(t) - u(t)]dt \\ &= \mathbb{E}\int_{t_0}^{t_1} \left[\alpha^u\psi(t)e^{S^u(t)}\right]^2[u^*(t) - u(t)]'[u^*(t) - u(t)]dt. \end{aligned} \tag{12}$$

In the first line we used that $\phi(t_0) = 1$, and in the second line we applied the Itô isometry [13]. In the third line we used $\alpha^u = e^{-S^u(t_0)}/\psi(t_0)$, which follows from Eq. (11) with $t = t_0$.

For the upper bound we consider Eq. (11) and apply Jensen's inequality

$$\psi(t)^2 = \mathbb{E}\left[e^{-S^u(t)}\middle|\mathcal{F}_t\right]^2 \leqslant \mathbb{E}\left[e^{-2S^u(t)}\middle|\mathcal{F}_t\right].$$

Substituting in Eq. (12) and using the law of total expectation we obtain the inequality (9).

For the lower bound we use Jensen's inequality on the whole integrand of Eq. (12) to obtain

$$\mathrm{Var}(\alpha^u) \geqslant \int_{t_0}^{t_1} \mathbb{E}\{\alpha^u \psi(t) e^{S^u(t)}[u^*(t) - u(t)]'\} \mathbb{E}\left\{\alpha^u \psi(t) e^{S^u(t)}[u^*(t) - u(t)]\right\} dt.$$

Using Eq. (11) and the law of total expectation we obtain the inequality (10). ■

We conclude that the optimal control problem is equivalent to the optimal sampling problem. An important consequence, which is given in Corollary 1, is that if the importance control is close to optimal, then so is the sampling efficiency.

V. THE MAIN PATH INTEGRAL THEOREM

The main theorem is a generalization of Theorem 1 that gives a solution of the control problem in terms of path integrals. The disadvantage of Theorem 1 is that it requires us to recompute the optimal control for each t,x separately. Here, we show that we can also compute the *expected* optimal future controls using a single set of trajectories with initialization $X(t_0) = x_0$. We furthermore generalize the path integral expressions by considering the product with some function $f(t,x)$. In the next section we utilize this result to construct a feedback controller. Here we proceed with the statement and the proof of the generalized path integral formula.

Notation 1. For any process $Y(t)$ we let $\langle Y(t)\rangle = \langle Y\rangle(t) = \mathbb{E}[\alpha^u Y(t)]$ denote the weighted average.

Theorem 3 (main theorem). Let $f : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$, and consider the process $f(t) = f(t,X(t))$. Then

$$\mathbb{E}[\psi(t)] = \mathbb{E}\left[e^{-S^u(t)}\right], \tag{13}$$

$$\langle (u^* - u) f\rangle(t) = \lim_{r\to t}\left\langle \frac{\int_t^r f(s)dW(s)}{r-t}\right\rangle(t). \tag{14}$$

Proof of (13). Consider the Feynman-Kac formula Eq. (11) and take the expectation with respect to $\mathcal{F}_{t_0}$. ■

Proof of (14). Consider Lemma 1 with $t = t_0$, multiply by $\int_t^r f(s)dW(s)$, and take the expected value:

$$\mathbb{E}\left[e^{-S^u(t_0)}\int_t^r f(s)dW(s)\right] = \mathbb{E}\int_t^r \phi(s)\psi(s)[u^*(s) - u(s)]f(s)ds.$$

On the left-hand side the term $\psi(t_0)\int f dW$ has disappeared because $\psi(t_0)$ is not random and the stochastic integral has zero mean. On the right-hand side we have used independent increments and the Itô isometry. Dividing by $r-t$ and taking the limit $r\to t$ we obtain

$$\begin{aligned}\lim_{r\to t}\frac{1}{r-t}\mathbb{E}\left[e^{-S^u(t_0)}\int_t^r f(s)dW(s)\right] &= \mathbb{E}[\phi(t)\psi(t)[u^*(t) - u(t)]f(t)] \\ &= \mathbb{E}\left[e^{-S^u(t_0)}[u^*(t) - u(t)]f(t)\right],\end{aligned}$$

where in the last line we used that $\phi(t) = e^{-S^u(t_0)+S^u(t)}$ and $\psi(t) = \mathbb{E}[e^{-S^u(t)}|\mathcal{F}_t]$ combined with the law of total expectation. Dividing both sides by $\mathbb{E}[e^{-S^u(t_0)}]$ gives Eq. (14). ■

VI. A PARAMETRIZED FEEDBACK CONTROLLER

In this section we illustrate how Theorem 3 can be used to construct a feedback controller. To this end we will assume that u^* is of the following parametrized from:

$$u^*(t,x) = A(t)h(t,x). \tag{15}$$

Here $h : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}^k$ will be referred to as the k "basis" functions which are assumed to be known. The "parameters" $A(t) \in \mathbb{R}^{m\times k}$ are assumed to be unknown. Note that the open-loop controller can be obtained by a parametrization with one basis function $h = 1$. The following corollary states that it is possible to estimate the optimal parameters from the equations in the main theorem.

Corollary 4 (path integral feedback). Let $f(t,x) \in \mathbb{R}^l$ be a function, and suppose that u^* is of the form Eq. (15); then

$$A(t)\langle hf'\rangle(t) = \langle uf'\rangle(t) + \lim_{r\to t}\left\langle \frac{\int_t^r f'(s)dW(s)}{r-t}\right\rangle. \tag{16}$$

Proof. This follows directly from Eq. (14) of the main theorem when the parametrized from of u^* is used. ■

Assuming that both the right-hand side and the cross correlations $\langle hf'\rangle(t)$ can be obtained by sampling methods, Eq. (16) gives for each time t a set of $m\times k$ linear equations in the $k\times m$ unknown parameters $A(t)$. These equations can be solved uniquely if the $k\times l$ matrix $\langle hf'\rangle$ is of rank k. Although we have to do these computations for each time t separately, only one set of paths is needed to get the sampling estimates for all times.

In general it will be impossible to check whether the optimal control is of the parametrized form. However, it seems plausible that if the parametrization can represent u^* quite well, it will be possible to estimate a good control function using Corollary 4. In the next section we perform a numerical experiment to support this statement.

Note that we can use any importance control u to estimate the optimal control u^*. In principle, we could use $u = 0$ and sample long enough to compute the u^* sufficiently accurately. However, we find it more efficient to use an iterative method where we use the optimal control estimate u_l that was computed at iteration l as an importance control for the computation of the optimal control u_{l+1}. According to Corollary 1 we know that improved controls have a higher fraction of effective samples and thus will make more efficient use of the sampling data. In particular, if u and u^* are parametrized with the same basis functions and time-dependent coefficients $A(t)$ and $A^*(t)$, respectively, this

results in an iterative update scheme for these coefficients. We refer to this method as iterative importance sampling.

We conclude that parametrized control functions can be obtained directly from path integral estimates, where the parameters can be computed using a single set of paths. Critically, these parametrized controls can be state-dependent functions. As a result, it is possible to construct (closed-loop) feedback controllers, which are more widely applicable than open-loop controllers.

VII. EXAMPLE

We consider the following control problem, of which we know the analytical solution.

Example 1 (geometric Brownian motion). For $t_0 \leqslant t \leqslant t_1$, the one-dimensional problem

$$dX^u(t) = X^u(t)\left(\frac{dt}{2} + u(t,X^u(t))dt + dW(t)\right),$$

$$S^u(t) = \frac{Q}{2}\ln[X^u(t_1)]^2 + \frac{1}{2}\int_t^{t_1} u(s,X^u(s))^2 ds + \int_t^{t_1} u(s,X^u(s))' dW(s),$$

has the solution

$$u^*(t,x) = \frac{-Q\ln(x)}{Q(t_1-t)+1}.$$

For the experiments we will take $x_0 = 1/2$, $t_0 = 0$, $t_1 = 1$, and $Q = 10$.

In a first experiment we visualize Theorem 2. To this end we consider a range of suboptimal importance controls $u^\epsilon(t,x) = u^*(t,x) + \sqrt{\epsilon}$. Each u^ϵ yields a path weight $\alpha^\epsilon := \alpha^{u^\epsilon}$. Because $\langle u^* - u\rangle' \langle u^* - u\rangle = \epsilon$, Theorem 2 implies that $\epsilon \leqslant \mathrm{Var}(\alpha^\epsilon) \leqslant \frac{\epsilon}{1-\epsilon}$. The results are reported in Fig. 1.

In a second experiment we construct feedback control functions based on various parametrizations. It is clear that a correct parametrization of the problem in Example 1 can be obtained with just one basis function: $\ln(x)$. In the experiment we also consider three parametrizations that cannot describe u^*: a constant, an affine, and a quadratic function of the state. The three controllers that we obtain in this way are denoted by $u^{(0)}$, $u^{(1)}$, and $u^{(2)}$, e.g., $u^{(2)}(t,x) = a(t) + b(t)x + c(t)x^2$.

We have used iterative importance sampling with $f = h$ as described in the previous section to estimate the parameters. The performance of the resulting control functions is given in Table I. The row $\mathbb{E}[S^u(t_0)]$ gives the expected cost, which we want to minimize. The row $\mathrm{Var}(\alpha^u)$ gives the variance of the path weight, which is directly related to the fraction of effective samples. Clearly the open-loop controller $u^{(0)}(t,x) = a(t)$ improves upon the zero controller $u(t,x) = 0$. The control further improves when the affine and quadratic basis functions are subsequently considered. The best result is obtained, unsurprisingly, with the logarithmic parametrization.

TABLE I. Performance estimates of various controllers based on 10^4 sample paths. Although for numerical consistency we used 10^4 sample paths to compute the parameters, only roughly 10^2 samples are required to obtain well-performing controllers.

	$u=0$	$u^{(0)}$	$u^{(1)}$	$u^{(2)}$	$a(t)\ln(x)$	u^*
$\mathbb{E}[S^u(t_0)]$	7.526	5.139	1.507	1.461	1.422	1.420
$\mathrm{Var}(\alpha^u)$	1.981	1.376	0.143	0.0506	0.0085	0.0071
λ^u (%)	34.3	42.08	87.5	95.2	99.1	99.3

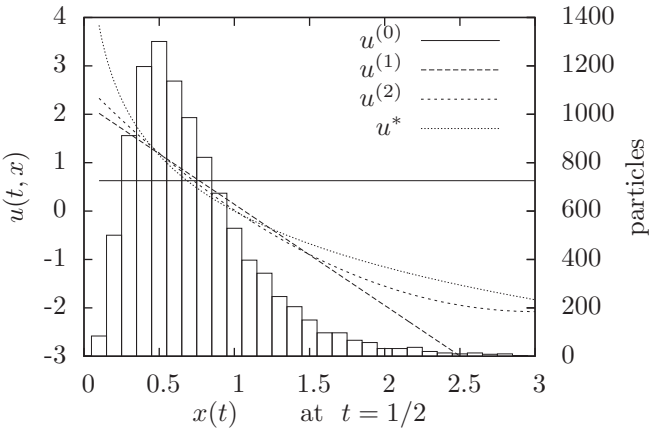

FIG. 2. The approximate controls calculated with 10^4 sample paths in two importance sampling iterations using a time discretization of $dt = 0.001$ for numerical integration. The histogram was created with 10^4 draws from $X^{u^*}(t)$ at $t = 1/2$.

In Fig. 2 we plot the state dependence of the feedback controllers at the intermediate time $t = 1/2$. Although the parametrized functions yield a control for all x, we are mainly interested in regions of the state space that are likely to be visited by the process X. This is visualized by a histogram of 10^4 particles that are drawn from $X^{u^*}(1/2)$. We observe that the optimal logarithmic shape is fitted, and that more complex parametrizations yield a better fit.

VIII. DISCUSSION

Most current feedback controllers that are used to stabilize systems are linear feedback controllers such as PID controllers. These are heuristic approaches that are optimal only if one assumes that the system dynamics is linear and the cost is quadratic. In this paper we have shown how to compute optimal feedback controllers for a class of nonlinear stochastic control problems. The optimality requires the use of the appropriate basis functions.

It should be noted that the optimal feedback is not necessarily a stabilizing term. Depending on the task it might be optimal to destabilize by amplifying the noise, for example, to create momentum efficiently.

Future work includes the development of methods for practical scenarios, based on the path integral feedback Eq. (16). An important aspect will be the selection of basis functions. A recent related work [6] discusses basis functions to obtain a solution of the linearized HJB equation (7).

ACKNOWLEDGMENTS

This work was supported by the European Community Seventh Framework Program (FP7) under Grant Agreement No. 270327 (CompLACS).

[1] E. Rombokas, E. Theodorou, M. Malhotra, E. Todorov, and Y. Matsuoka, Tendon-driven control of biomechanical and robotic systems: A path integral reinforcement learning approach, in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, New York, 2012), pp. 208–214.

[2] K. Kinjo, E. Uchibe, and K. Doya, Evaluation of linearly solvable Markov decision process with dynamic model learning in a mobile robot navigation task, Front. Neurorob. **7**, 7 (2013).

[3] P. Glasserman and P. Heidelberger, Asymptotically optimal importance sampling and stratification for pricing path-dependent options, Math. Finance **9**, 117 (1999).

[4] R. F. Stengel, *Optimal Control and Estimation* (Dover, New York, 1994).

[5] E. Theodorou and E. Todorov, Relative entropy and free energy dualities: Connections to path integral and *kl* control, in *Proceedings of the 51st Annual IEEE Conference on Decision and Control (CDC)* (IEEE, New York, 2012), pp. 1466–1473.

[6] M. B. Horowitz, A. Damle, and J. W. Burdick, Linear Hamilton Jacobi Bellman equations in high dimensions, in *Proceedings of the 53rd Annual IEEE Conference on Decision and Control (CDC), 2014* (IEEE, New York, 2014), pp. 5880–5887.

[7] H. J. Kappen, Linear theory for control of nonlinear stochastic systems, Phys. Rev. Lett. **95**, 200201 (2005).

[8] B. van den Broek, W. Wiegerinck, and H. J. Kappen, Graphical model inference in optimal control of stochastic multi-agent systems, J. Artif. Intell. Res. **32**, 95 (2008).

[9] R. Anderson and D. Milutinović, A stochastic optimal enhancement of feedback control for unicycle formations, in *Proceedings of the 11th International Symposium on Distributed Autonomous Robotic Systems (DARS)*, edited by M. Ani Hsieh and G. Chirikjian (Springer, New York, 2012).

[10] N. Sugimoto and J. Morimoto, Phase-dependent trajectory optimization for CPG-based biped walking using path integral reinforcement learning, in *Proceedings of the 11th International Conference on Humanoid Robots, IEEE-RAS* (IEEE, New York, 2011), pp. 255–260.

[11] E. Theodorou, J. Buchli, and S. Schaal, A generalized path integral control approach to reinforcement learning, J. Mach. Learn. Res. **11**, 3137 (2010).

[12] H. J. Kappen, Path integrals and symmetry breaking for optimal control theory, J. Stat. Mech.: Theory Exp. (2005) P11011.

[13] Bernt Øksendal, *Stochastic Differential Equations: An Introduction with Applications* (Springer, Berlin, 1985).

[14] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Stochastic Modelling and Applied Probability (Springer, Berlin, 2006).

[15] J. Bierkens and H. J. Kappen, Explicit solution of relative entropy weighted control, Syst. Control Lett. **72**, 36 (2014).

[16] Jun S. Liu, *Monte Carlo Strategies in Scientific Computing*, corrected ed. (Springer, Berlin, 2008).