A Hamiltonian Approach Using Partial Differential Equations for Open-Loop Stochastic Optimal Control

Aaron Palmer, Dejan Milutinović

Abstract— This paper utilizes a minimum principle for infinite dimensional systems for the optimal control of systems constrained by the Fokker-Plank equation governing the evolution of a state probability density function. From the backwards evolution of the corresponding adjoint system, we define a Hamiltonian and use its gradient to construct a numerical optimal control. The basic nature of the adjoint system allows for all of the necessary terms defining the control to be inferred from stochastic process samples which is exploited in provided examples. Solving stochastic optimal control problems utilizing stochastic processes is a promising approach for solving open-loop stochastic optimal control problems of non-linear dynamic systems with a multi-dimensional state vector.

I. INTRODUCTION

Since the development of the Pontryagin Minimum Principle [1], the Hamiltonian is a fundamental tool in the analysis of optimal control problems. Similar to Hamiltonian mechanics in Physics, the Hamiltonian for optimal control is defined based on a set of co-state variables obeying an adjoint system of equations. A standard approach to stochastic optimal control is to utilize Bellman's dynamic programming algorithm and solve the corresponding Hamilton-Jacobi-Bellman (HJB) equation. For incomplete state feedback information, the standard result considers problems with linear dynamics, quadratic costs and Gaussian noise, and the exact feedback solution is used along with a filter for the state estimation [2].

In this paper, we consider a Hamiltonian approach for solving general stochastic optimal control problems with no state feedback, i.e., open-loop control. The open-loop stochastic optimal control depends on terminal costs, the dynamics and the initial state probability distribution. Previous work on this problem has focused on constructing the Hamiltonian with solutions to forward-backward stochastic differential equations [3], [4]. The focus of our work is a novel method based on an infinite dimensional minimum principle [5] applied to partial differential operators for the state probability density function (PDF) and the co-state distribution.

In order to compute the optimal control, we can use partial differential equation (PDE) solution for the state PDF, the co-state distribution and an iterative numerical method similar to the one presented in [6]. However, the dimension of the dynamical system state can be the limiting factor to proceed

with these types of methods. The reason is that we need not only reliable multidimensional PDE solvers, but a reliable computation of the scalar product, which is important for the Hamiltonian evaluation. Therefore, it is beneficial to find an alternative way to evaluate the state PDF, the co-state distribution and the Hamiltonian.

In this work, we show that the state, co-state distributions and the Hamiltonian can be evaluated based on trajectory samples, where the co-state distribution is evaluated from trajectory samples by the Feynman-Kac formula [7]. This evaluation method is suitable for non-linear multi-dimensional systems and it provides rationale on how stochastic processes can be utilized to compute control.

Our work is illustrated by examples in which the Hamiltonian is used as a gradient in search for a time discretized optimal control, following work on successive approximation for optimal control from Mitter [8] and previous work of the author Milutinović, which uses the infinite dimensional minimum principle [6], [9]. In our examples, the co-state distribution is either solved exactly or evaluated from trajectory samples. Other manners of the co-state distribution evaluation can be done using a finite approximation of the eigenspace of the operators as done for an HJB example by Hongler, et. al. [10].

In Section II we introduce a non-linear stochastic optimal control problem. Section III explains the minimum principle we are dealing with, as well as its relation to stochastic trajectory samples. An example with controlled intensity of diffusion is presented in Section IV, and a problem involving a non-linear stochastic process and state constraints is provided in V. Section VI gives conclusions.

II. OPTIMAL CONTROL PROBLEM FORMULATION

The stochastic differential equation that we control is in the form

$$dX = b(X, t, u(t))dt + L(X, t, u(t))dw, \qquad (1)$$

where $X(t)$ is an n dimensional stochastic process and dw is the derivative of a m dimensional Wiener process. The control belongs to the set $u(t) : [0, T] \to U \subset \mathbb{R}^k$. The vector $b(X, t, u(t)) : D \times [0, T] \times U \to \mathbb{R}^n$, and, similarly, the matrix $L(X, t, u(t)) : D \times [0, T] \times U \to \subset \mathbb{R}^{n,m}$. In our approach, we deal with the probability density function (PDF) of the process $X(t)$, $\rho(X, t) : D \times [0, T] \to \mathbb{R}^+$, and consider its evolution in time. The PDF $\rho(X, t)$ evolves according to the Fokker-Planck equation [11], [12]. In sequel, to make notation shorter, we will omit explicit dependence on the space and time coordinates.

A. Palmer was an undergraduate student while he worked on this paper at University of California, Santa Cruz, USA, e-mail: apemalr@gmail.com

D. Milutinović is an Assistant Professor of Applied Mathematics and Statistics, University of California, Santa Cruz, 1156 High Street, USA, e-mail: dejan@soe.ucsc.edu

The Fokker-Planck equation for our process is

$$\frac{\partial \rho}{\partial t} = \sum_{i,j=1}^{n} \frac{\partial(-b_i(u)\rho)}{\partial x_i} + \frac{1}{2}\frac{\partial^2\left([LL^T]_{ij}(u)\rho\right)}{\partial x_i \partial x_j}, \quad (2)$$

$$= F(u)\rho. \quad (3)$$

where the second equality is the same equation written in the form of a differential operator $F(u)$ acting on the probability density function ρ, and x_i is the ith component of the state vector X. Let us introduce an inner product $\langle\cdot,\cdot\rangle$ between two functions as

$$\langle f, g \rangle = \int_D f(X)g(X)dX \quad (4)$$

where x belongs to the domain D. Then we can write

$$\langle \rho(t), 1 \rangle = 1, \quad \forall t \quad (5)$$

since under the Fokker-Plank evolution, the probability mass is conserved.

The optimal control we consider is to find the open-loop control sequence u that minimizes the cost function

$$J(u) = \langle \phi, \rho(T) \rangle + \int_0^T \langle f_0(X(t), u(t)), \rho(t) \rangle dt. \quad (6)$$

The first term is the terminal cost given in the form of a product between the weighting function ϕ and the PDF at the terminal time T, $\rho(T)$, which is the expected value of $\phi(X(T))$. The second term describes the expected costs acquired along the stochastic process (1) trajectory controller by the control sequence u.

III. Minimum Principle and Trajectory Samples

In the previous section, we defined the optimal control of the stochastic process in a probability density function space. Therefore, in order to solve the control, we employ the minimum principle for infinite dimensional systems [5].

According to the minimum principle, a necessary condition for optimality is that for any time point $t \in [0, T]$, the optimal control value $u^*(t)$ minimizes the Hamiltonian, i.e,

$$H(\rho^*(t), \pi^*(t), u^*(t)) = \min_{u(t)}\{H(\rho^*(t), \pi^*(t), u(t))\} \quad (7)$$

where $\rho^*(t)$ and $\pi^*(t)$ are the state PDF and the co-state distribution at the time t, respectively (Appendix I). The Hamiltonian is defined as

$$H(\rho, \pi, u) = \langle \rho, f_0(u) + F'(u)\pi \rangle. \quad (8)$$

while the co-state distribution evolution is defined by

$$-\frac{\partial \pi}{\partial t} = f_0(u(t)) + F'(u(t))\pi, \quad (9)$$

$$\pi(T) = \phi \quad (10)$$

with $F'(u)$ being the adjoint operator to the operator $F(u)$, such as that (see Appendix II)

$$F'(u)\pi = \sum_{i,j}^{n} b_i(u)\frac{\partial \pi}{\partial x_i} + \frac{1}{2}[LL^T]_{ij}(u)\frac{\partial^2 \pi}{\partial x_i \partial x_j}. (11)$$

Moreover, for $F'(u)$ to be the adjoint operator to $F(u)$, we also need to verify that at the boundary of the domain D, we have

$$\oint_{\partial D} \sum_{i,j}^{n} n_i \Big(b_i(u)\pi\rho + \pi\frac{\partial\left([LL^T]_{ij}(u)\rho\right)}{\partial x_j} \quad (12)$$

$$- [LL^T]_{ij}(u)\rho\frac{\partial \pi}{\partial x_j}\Big)dS = 0,$$

where n is the normal vector to the boundary surface and dS is the differential surface element.

An important use of the Hamiltonian is that it provides the gradient of the total cost with respect to control, in other words,

$$\frac{\partial J(u)}{\partial u(t)} = \frac{\partial H(\rho, \pi, u)}{\partial u(t)} \quad (13)$$

which is verified for discrete time by [2], [13] and is very useful in numerical iterative algorithms computing a discrete approximation of the optimal control. Clearly the right-hand side of (13) is easier to calculate than the left-hand side of the equation since the differentiation is only with respect to $u(t)$ and the Hamiltonian depends on the state PDF and the co-state distribution at the time t, i.e., $\rho(t)$ and $\pi(t)$, respectively.

If we sample trajectories of a stochastic process (1), the density of samples will evolve according to the Fokker-Planck equation. However, it is a less obvious fact that the co-state distribution at the point X^+ can be evaluated as

$$\pi(X^+, t) =$$
$$E\left[\int_t^T f_0(X(s), u)ds + \phi(X(T))|X(t) = X^+\right], \quad (14)$$

where $X(s)$ obeys the evolution of (1) given u and starting at the time point t with the initial condition $X(t) = X^+$. The proof of this relation is provided in Appendix III and the relation is also known as the Feyman-Kac equation from statistical mechanics.

IV. 2D Example With Controlled Noise Intensity

This example is inspired by the standard problem involving linear dynamics and quadratic cost. Here we use a numerical solution of the backward co-state distribution evolution (18-27) and samples of stochastic trajectories to evaluate the inner product of the Hamiltonian (28). The control is computed iteratively with update rules based on the gradient of the Hamiltonian (29). We use the gradient method with a constant step-size, the Fletcher-Reeves conjugate-gradient method and the Matlab $fmincon$ function for constrained non-linear minimization from the Matlab optimization toolbox. All three numerical procedures converge towards the same cost function value (Fig. 1) and the same control presented in Fig. 2.

The process under control is two-dimensional $X = [x \ y]$ and described by

$$d\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ u \end{bmatrix}dt + \begin{bmatrix} \xi_0 & 0 \\ 0 & \xi_1 u \end{bmatrix}\begin{bmatrix} dW_1 \\ dW_1 \end{bmatrix}. \quad (15)$$
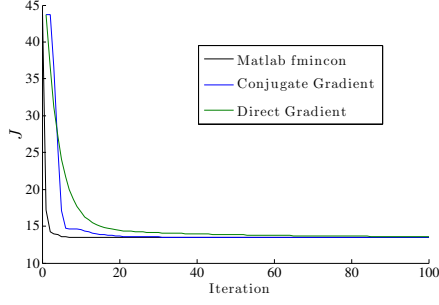
Fig. 1. A comparison of optimization methods for the first example. For the gradient methods, $\lambda = 0.005$ was used.
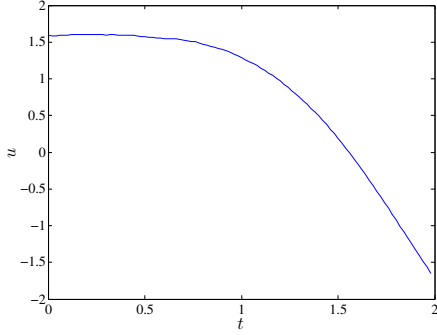

Fig. 2. Optimal control found for the 2D example. The control is nearly constant to limit diffusion and then decreases nearly linearly so the particle comes close to a stop near the origin. Iterations computing the optimal control were initialized with $u(t) = 0$, $\forall t$

where the control $u(t)$ influences both the drift, as well as the process noise intensity. We assume that the control is within the bounds $u \in [u_{\min}, u_{\max}]$ and the initial condition is $x(0) = x_0$, $y(0) = y_0$, i.e., the initial state PDF $\rho([x\ y], 0)$ is a two-dimensional Dirac function centered at x_0, y_0, i.e.,

$$\rho([x\ y], 0) = \rho_0([x\ y]) = \delta(x - x_0, y - y_0) \quad (16)$$

The cost function we consider is given by expression (6), the terminal time $T = 2$ and the instantaneous and terminal costs are given by

$$f_0(u) = \frac{1}{2}Ru^2 \text{ and } \phi(x, y) = \frac{a}{2}x^2 + \frac{b}{2}y^2. \quad (17)$$

Other parameters of the process and the cost function are $x_0 = -2$, $y_0 = -1$, $\xi_0 = 0.5$, $\xi_1 = 0.5$, $R = 0.5$, $a = 5$, $b = 5$, $u_{\min} = -10$, $u_{\max} = 10$.

The state PDF evolution $\rho(t)$ is given by

$$\frac{\partial \rho}{\partial t} = F(u)\rho = -y\frac{\partial \rho}{\partial x} - u\frac{\partial \rho}{\partial y} + \frac{\xi_0^2}{2}\frac{\partial^2 \rho}{\partial x^2} + \frac{\xi_1^2 u^2}{2}\frac{\partial^2 \rho}{\partial y^2}$$

The corresponding co-state distribution can be obtained from the adjoint equation as

$$-\frac{\partial \pi}{\partial t} = y\frac{\partial \pi}{\partial x} + u\frac{\partial \pi}{\partial y} + \frac{\xi_0^2}{2}\frac{\partial^2 \pi}{\partial x^2} + \frac{\xi_1^2 u^2}{2}\frac{\partial^2 \pi}{\partial y^2} \quad (18)$$

$$\pi(T) = \phi. \quad (19)$$

In the latter equation, π can be parameterized as a quadratic function, analogous to the commonly used Riccati equation of linear quadratic feedback control [2], and the parameters are found by solving ordinary differential equations. Thus, we look for a solution for π in the form

$$\pi = \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix} P(t)\begin{bmatrix} x \\ y \end{bmatrix} \quad (20)$$
$$+ \alpha_1(t)x + \alpha_2(t)y + \beta(t).$$
$$F'(u)\pi = \frac{R}{2}u^2 + yx[P]_{11} + (y^2 + xu)[P]_{12} + yu[P]_{22}$$
$$+ y\alpha_1 + u\alpha_2 + \frac{\xi_0^2}{2}[P]_{11} + \frac{\xi_1^2 u^2}{2}[P]_{22}. \quad (21)$$

Collecting similar terms, we get a system of differential equations,

$$[\dot{P}]_{11} = 0 \quad (22)$$
$$-[\dot{P}]_{22} = 2[P]_{12} \quad (23)$$
$$-[\dot{P}]_{12} = [P]_{11} \quad (24)$$
$$-\dot{\alpha}_1 = u[P]_{12} \quad (25)$$
$$-\dot{\alpha}_2 = \alpha_1 + u[P]_{22} \quad (26)$$
$$-\dot{\beta} = \frac{R}{2}u^2 + u\alpha_2 + \frac{\xi_0^2}{2}[P]_{11} + \frac{\xi_1^2 u^2}{2}[P]_{22}. \quad (27)$$

that should satisfy the terminal condition for $\pi(T)$, given by $\phi(x, y)$. The part of the Hamiltonian with control dependence can be expressed

$$H(\rho, \pi, u) = < \rho, \frac{R}{2}u^2 + \frac{u^2\xi_1^2}{2}\left(b + a(T-t)^2\right) > \quad (28)$$
$$+ u\left(a(T-t)x + (b + a(T-t)^2)y + \alpha_2\right)$$
$$\frac{\partial H(\rho, \pi, u)}{\partial u} = u\left(R + \xi_1^2\left(b + a(T-t)^2\right)\right) + \alpha_2 \quad (29)$$
$$+ < \rho, a(T-t)x + (b + a(T-t)^2)y >$$

While we do not know ρ exactly, the inner product can be computed by sampling trajectories and evaluating the necessary derivatives of the co-state at the sampled positions. In order to compute the numerical optimal control presented in Fig. 2, we discretized control with 100 time steps and we used 500 trajectory samples.

V. NON-LINEAR EXAMPLE

The purpose of this example is to show implementation of this method on a problem with non-linear dynamics and more difficult boundary conditions. Rather than attempting to solve the resulting PDE's, we evaluate the co-state by computing the expectation of costs along trajectory samples and infer all relevant information on the Hamiltonian from the statistical data. Once again an iterative method is employed that, despite the noise of the sampled data influencing the iterations (Fig. 3), converges to the optimal control presented in Fig. 4 after iterations of the control shown in Fig. 5.

The non-linear process under control is:

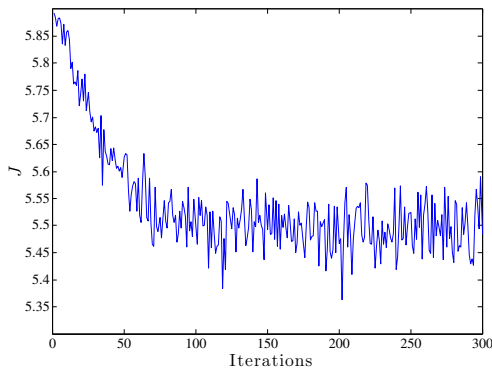$$dx = [u(t) - \alpha x]dt + \xi\sqrt{\frac{(3-x)^2}{2} + u^2(t)}dw \quad (30)$$

Fig. 3. The convergence plot of the cost for a conjugate gradient method with $\lambda = 0.05$. The use of stochastic process samples in calculations causes the cost function fluctuations.
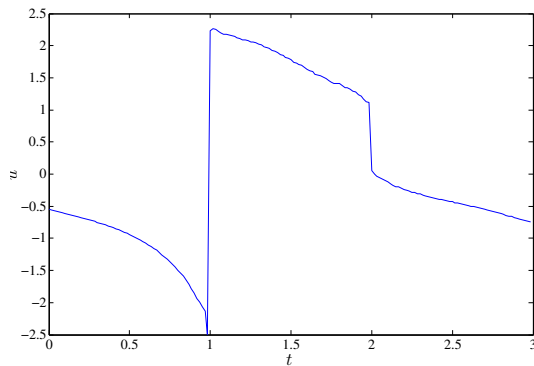

Fig. 4. Optimal control found for the non-linear example problem with diffusion proportional to the space. Iterations computing the optimal control were initialized with $u(t) = 0, \forall t$

with the initial condition $x(0) = 0$ and the non-linear state constraints for x,

$$x(t) \in [-3, 3], \forall t, \tag{31}$$

$$x(t) \in [-2, -1], \ t = 1, \tag{32}$$

$$x(t) \in [\frac{1}{2}, 2], \ t = 2 \tag{33}$$

the process terminates upon hitting the boundary or the 'walls' at two intermediate points, and accrues a cost upon termination. It also has a non-linear relation between the control and the noise intensity. The process is one-dimensional and we assume $\xi = 0.1$, $\alpha = 0.5$, $u(t) \in [-3, 3], \forall t$, the terminal time $T = 3$ and $\rho(0) = \delta(x)$. The cost function we use to control the process is given by

$$f_0(u) = \frac{R}{2}u^2, \ \phi(x) = \frac{4}{9}x^2, \ R = 0.2. \tag{34}$$

The Hamiltonian with dependance on u is

$$H(\rho, \pi, u) = \ <\rho, \frac{R}{2}u^2 + \frac{\partial \pi}{\partial x} + \frac{\xi_0^2 u^2}{2}\frac{\partial^2 \pi}{\partial x^2}> \tag{35}$$

and the corresponding gradient with respect to u is

$$\frac{\partial H}{\partial u}(\rho, \pi, u) = \ <\rho, Ru + u\frac{\partial \pi}{\partial x} + \xi_0^2 u\frac{\partial^2 \pi}{\partial x^2}> . \tag{36}$$

In order to compute π, we should use the expectation of (14). However, with the introduction of the boundaries, any time a trajectory, which we use to compute $\pi(x, t)$, hits the boundaries, i.e, forbidden states, it should be appropriately weighted in the expectation. Therefore, the expression for computing the expectation is

$$\pi(x^+, t) =$$
$$\frac{1}{K}\sum_{k=1}^{K}\left[\phi(x_k(T)) + \int_t^T f_0(x_k(s), u)ds\right](1 - I_k) \tag{37}$$
$$+ \frac{1}{K}\sum_{k=1}^{K}\left[\Gamma(r) + \int_t^r f_0(x_k(s), u)ds\right]I_k$$

where $x_k(t)$ is the sample of kth trajectory at the time point t and I_k is 1 if the trajectory k hits the boundary, otherwise is 0. The function $\Gamma(r) = 7 - r$ puts weight on the length of the trajectory hitting a boundary at time $r \in [0, T]$ and the weight is smaller for longer trajectories.

For a numerical solution instead of solving the PDE we evaluate $\pi(x, t)$ on a spatio-temporal grid using the expectation of (37). For all the points that are not at the grid point, we interpolate π using a piecewise cubic polynomial resulting from the Matlab's *interp* function for cubic spline approximation. To compute π, we iterate backwards in time, sampling paths forward and computing the expectation based on the instantaneous cost, as well as an interpolated value of the expected value from the time step ahead, for which π has already been computed. The piecewise cubic polynomial also gives us approximations of the first and second derivatives of π.

To compute the inner product to get the Hamiltonian, we generate multiple process trajectories from the initial position $x(0)$ and calculate the weighted expression for the gradient of the Hamiltonian using the interpolated values of the derivatives of π. In numerical calculations, we use 180 time steps and 20 space steps. For evaluation of $\pi(x, t)$, we sample 100 paths for each grid point and 1000 paths for the inner product. Starting from a control of zero, the conjugate
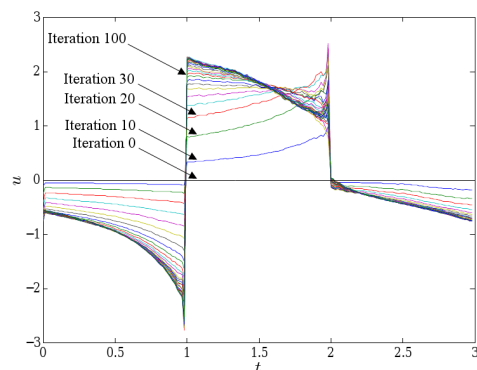

Fig. 5. A plot of the updated controls for every ten iterations. Although the calculation of the update is stochastic, we observe the convergence to the optimal control.
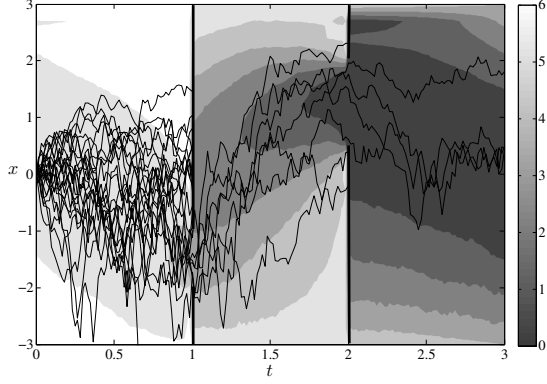
Fig. 6. Trajectories over a contour for π (gray scale), all calculated for the optimal control of the non-linear example.

gradient method finds the slits (31-33) and converges to the optimal control solution.

Fig. 6 shows the computed co-state distribution values as a contour plot on a gray scale, as well as examples of the stochastic process trajectories under the computed control. This helps to illustrate how the diffusion can actually help the optimization, as without the diffusion the slits (31-33) at $t = 1$ and $t = 2$ would never be found.

VI. CONCLUSIONS

While stochastic optimal control problems have been studied previously, their relation with the Pontryagin-like minimum principle has not been made clear and simple by the literature surveyed by the authors. Having in mind that the state PDF of stochastic processes under control obeys the PDE evolutions, we proposed a solution of control problems based on an infinite dimensional minimum principle.

The Hamiltonian we obtain from the minimum principle provides us a gradient we can use in numerical iterations computing the optimal control. In order to obtain the Hamiltonian, we can use the PDE solutions. However, we also showed that the Hamiltonian can be computed based on stochastic trajectory samples and we provided two illustrative examples.

The method for computing the Hamiltonian utilizing stochastic processes is general, it can work in many dimensions and deal with complex state constraints. Due to its intrinsic stochasticity, the optimization methods based on such computations have also a better chance of escaping from local minimums. For future research, methods of using importance sampling or sampling with larger diffusions could be considered as ways of improving the search for global rather then local minimums.

APPENDIX I

The work of this paper is based on Theorem 6.6.2 from H.O. Fattorini's book [5]. This theorem has the hypothesis that the evolution operator, which is in our case $F\rho$, is a Fréchet differentiable, and both the operator and its derivative are bounded. Since the differential operator we are dealing

with is linear, these conditions are met for our case when the operator is bounded, as shown in Appendix IV. The same has to be valid for the instantaneous cost and the terminal cost is assumed to be a Fréchet differentiable.

Assuming there are no restrictions on ρ at the terminal time T, the necessary conditions of optimality are that there exists a scalar $z_0 \geq 0$ and the co-state distribution $\pi^*(x, t)$ satisfying

$$-\frac{\partial \pi^*(X, t)}{\partial t} = F'(u^*)\pi^*(X, t) + z_0 f_0(X, t),$$
$$\pi(T) = z_0 \phi(T)$$
$$< \rho^*, z_0 f_0 > + < F(u^*)\rho^*, \pi^* > =$$
$$\min_{u \in U} \{ < \rho^*, z_0 f_0 > + < F(u)\rho^*, \pi^* > \}.$$

where u^* is the optimal control, while ρ^* and ϕ^* are the state PDF and co-state distribution under the optimal control, respectively. The Hamiltonian we are dealing with is defined by the expression inside the braces. Generally, z_0 is determined only up to multiplication by a positive scalar and in our work is set to be 1. In relation to classic optimal control, π can be interpreted as a Lagrange multiplier for the differential constraint.

APPENDIX II

We would like to find the adjoint differential operator F' satisfying the Green's identity,

$$< F\rho, \pi > - < \rho, F'\pi > = 0. \tag{38}$$

To determine the adjoint operator and the proper boundary conditions for π, we use the product rule to 'liberate' ρ [14],

$$\begin{aligned}
\pi F\rho &= \sum_{i,j=1}^{n} -\pi \frac{\partial}{\partial x_i}(b_i \rho) + \frac{1}{2}\pi \frac{\partial^2 \left([LL^T]_{ij}\rho\right)}{\partial x_i \partial x_j}, \\
&= \sum_{i,j=1}^{n} \left(-\frac{\partial}{\partial x_i}(\pi b_i \rho) + \rho b_i \frac{\partial}{\partial x_i}\pi \right) \\
&\quad + \frac{1}{2}\frac{\partial}{\partial x_i}\left(\pi \frac{\partial \left([LL^T]_{ij}\rho\right)}{\partial x_j} \right) \\
&\quad - \frac{1}{2}\frac{\partial}{\partial x_i}\left([LL^T]_{ij}\rho \frac{\partial \pi}{\partial x_j} \right) + \frac{1}{2}\rho[LL^T]_{ij}\frac{\partial^2 \pi}{\partial x_i \partial x_j}.
\end{aligned}$$

The "process of liberation" is shown in [14] on p. 196, and on pp. 171, 180 there are discussions on how the boundary conditions belong to the differential operator.

From the terms for which ρ has been 'liberated', we define the adjoint operator F' by

$$F'\pi = \sum_{i,j=1}^{n} b_i \frac{\partial}{\partial x_i}\pi + \frac{1}{2}[LL^T]_{ij}\frac{\partial^2 \pi}{\partial x_i \partial x_j} \tag{39}$$

and the additional condition that must be satisfied

$$\begin{aligned}
0 &= < \pi F\rho - \rho F'\pi > \\
&= \oint_{\partial D} \sum_{i,j=1}^{n} n_i \Big((\pi b_i \rho) + \frac{1}{2}\pi \frac{\partial \left([LL^T]_{ij}\rho\right)}{\partial x_j} \\
&\quad - \frac{1}{2}[LL^T]_{ij}\rho \frac{\partial \pi}{\partial x_j} \Big) dS
\end{aligned}$$

The last line is an application of the divergence theorem, where the term becomes the dot product with a normal vector, n, of the boundary surface.

APPENDIX III

The inner product of a function with a PDF represents the expectation of a function with regards to that PDF. Let ϕ be an arbitrary PDF,

$$
\begin{aligned}
\frac{\partial}{\partial t}(<\phi,\pi>) &= <\frac{\partial \phi}{\partial t},\pi> + <\phi,\frac{\partial \pi}{\partial t}>, \\
&= <F\phi,\pi> - <\phi, f_0 + F'\pi>, \\
&= -<\phi, f_0>.
\end{aligned}
$$

The adjoint operator is defined to satisfy the Green's identity leaving only the term of instantaneous cost on the right-hand side. Now, integrate both sides from $t \to T$

$$
<\phi,\pi>|_T - <\phi(t),\pi(t)>|_t = -\int_t^T <\phi, f_0>|_s ds,
$$

and let $\phi(X,t)$ be a delta function centered at X^+, i.e, $\phi(X,t) = \delta(X - X^+)$; then the expression above becomes

$$
<\phi,\pi>|_T - \pi(X^+,t) = -\int_t^T <\phi, f_0> ds,
$$

and finally we obtain

$$
\pi(X^+,t) = <\phi,\pi>_T + \int_t^T <\phi, f_0> ds.
$$

which is equivalent to the expectation of (14). If we have a boundary condition such that paths terminate earlier, we also need to include the end cost for those paths.

APPENDIX IV

In this appendix, we provide assumptions under which the operator F is bounded. We can use the product rule to rewrite (2) in terms of derivatives of ρ,

$$
\begin{aligned}
F\rho &= \sum_{i,j=1}^{n} \frac{\partial}{\partial x_i}(b_i \rho) \\
&\quad + \frac{1}{2}\frac{\partial}{\partial x_i}\left(\rho\frac{\partial [LL^T]_{ij}}{\partial x_j} + [LL^T]_{ij}\frac{\partial \rho}{\partial x_j}\right), \\
&= \sum_i^n \left(\rho\frac{\partial b_i}{\partial x_i} + b_i\frac{\partial \rho}{\partial x_i}\right) \\
&\quad + \sum_{i,j}^n \left(\frac{1}{2}\rho\frac{\partial [LL^T]_{ij}}{\partial x_i \partial x_j}\right. \\
&\quad \left. + \frac{\partial^2 [LL^T]_{ij}}{\partial x_i}\frac{\partial \rho}{\partial x_j} + \frac{1}{2}[LL^T]_{ij}\frac{\partial^2 \rho}{\partial x_i \partial x_j}\right).
\end{aligned}
$$

Naturally, we define the L_2 norm $\|\cdot\|$ using the scalar product (4). To show the boundedness, we must show that for $\rho \in L^2$, $\exists M \in \mathbb{R}^+$, such that for

$$
\|\rho\| \leq 1, \quad \|F\rho\| \leq M. \tag{40}
$$

Using the Cauchy-Schwartz inequality and triangle inequality, we see that

$$
\begin{aligned}
\|F\rho\| &\leq \sum_{i,j}^n \left|\left|\frac{\partial b_i}{\partial x_i}\rho\right|\right| + \frac{1}{2}\left|\left|\frac{\partial^2 [LL^T]_{ij}}{\partial x_i \partial x_j}\rho\right|\right| \\
&\quad + \sum_{i,j}^n \left|\left|b_i\frac{\partial \rho}{\partial x_i}\right|\right| + \left|\left|\frac{\partial [LL^T]_{ij}}{\partial x_i}\frac{\partial \rho}{\partial x_j}\right|\right| \\
&\quad + \sum_{i,j}^n \frac{1}{2}\left|\left|[LL^T]_{ij}\frac{\partial^2 \rho}{\partial x_i \partial x_j}\right|\right|.
\end{aligned}
$$

It is clear that $\|F\rho\|$ is bounded as long as b, LL^T, ρ and its first and second derivatives are bounded. In the case of discontinuities, extra care should be taken to define the operator F on a Sobolev space in such a way that the operator is bounded.

ACKNOWLEDGMENTS

REFERENCES

[1] L. S. Pontryagin, *L.S. Pontryagin Selected Works: The Mathematical Theory of Optimal Processes.* New York, New York: Gordon and Breach Science Publishers, 1986.
[2] A. E. Bryson and Y. C. Ho, *Applied Optimal Conrol.* New York, New York: Taylor & Francis, 1975.
[3] J. Yong and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations.* New York, New York: Springer-Verlag, 1999.
[4] L. Chen and Z. Wu, "Maximum principle for stochastic optimal control problem of forward-backward system with delay," in *Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, Shanghai, P. R. China, 2009.
[5] H. O. Fattorini, *Infinite Dimensional Optimization and Control Theory.* Cambridge, UK: Cambridge University Press, 1999.
[6] D. Milutinović and P. Lima, *Cells and Robots: Modeling and Control of Large-Size Agent Populations.* Springer Tracts in Advanced Robotics, Sptringer, 2007.
[7] H. J. Kappen, "A linear theory for control of non-linear stochastic systems," *Physical Review Letters*, vol. 95, Nov. 2005.
[8] S. K. Mitter, "Successive approximation methods for the solution of optimal control problems," *Automatica*, vol. 3, pp. 135–149, 1966.
[9] D. Milutinović and P. Lima, "Modeling and optimal centralized control of a large-soze robotic population," *IEEE Transactions on Robotics*, vol. 22, pp. 1280–1285, Nov. 2006.
[10] M. O. Hongler, O. Gallay, M. Hülsmann, P. Cordes, and R. Colmorn, "Centralized versus decentralized control–a solvable stylized model of transportation," *Physica A*, vol. 389, pp. 4162–4171, 2010.
[11] J. Q. Sun, *Stochastic Dynamics and Control.* Amsterdam, The Netherlands: Elsevier, 2006.
[12] P. Whittle, *Probability via Expectation.* New York, New York: Springer-Verlag, 2000.
[13] D. P. Bertsekas, *Nonlinear Programming.* Belmont, Massachusetts: Athena Scientific, 1995.
[14] C. Lanczos, *Linear Differential Operators.* Philadelphia, Pennsylvania: SIAM, 1961.