Finding the Most Likely Trajectories of Optimally-Controlled Stochastic Systems

Emanuel Todorov

University of Washington, Seattle WA, USA
todorov@cs.washington.edu

Abstract: Optimal trajectories of deterministic systems satisfy Pontryagin's maximum principle and can be computed efficiently. Related results for stochastic systems exist but they lack the simplicity and computational efficiency of the deterministic case. Here we show that a certain class of both discrete-time and continuous-time nonlinear stochastic control problems obey a classic maximum principle, in the sense that the most likely trajectory of the optimally-controlled stochastic system is the solution to a deterministic optimal control problem. Apart from their theoretical significance, our results yield new numerical methods for stochastic control.

Keywords: Stochastic optimal control, most likely trajectory, maximum principle

1. INTRODUCTION

The solutions to deterministic optimal control problems can be equivalently characterized by Bellman's equation and by Pontryagin's maximum principle. Bellman's equation characterizes the optimal cost-to-go function $v(x,t)$, which is the sum of immediate costs $\ell(x,u)$ accumulated by initializing the system at state $x \in \mathbb{R}^n$ and time t and controlling it optimally until a final time T. This characterization however is global, and computing v everywhere is numerically infeasible when the dimensionality n is large. In contrast, the maximum principle characterizes the gradient of v restricted to extremal trajectories (one of which is the optimal trajectory) and thereby avoids the need for global computation. A variety of numerical methods for computing extremal trajectories are available and are routinely used. They include gradient descent and ODE solvers, differential dynamic programing (DDP), iterative LQG (iLQG). See Bryson and Ho (1969); Jacobson and Mayne (1970); Todorov and Li (2005).

The stochastic case is quite different. Noise makes it difficult to guarantee any properties of an isolated trajectory without considering what happens in its neighborhood. More precisely, Bellman's equation

$$v(x,t) = \min_u \{\ell(x,u) + E[v(y,t+1)]\} \tag{1}$$

involves an expectation with respect to the probability distribution over next states y. This expectation couples the costs-to-go of nearby states in a way which is hard to disentangle. Maximum principles for stochastic control have been derived, however they take the form of backward stochastic differential equations (Yong and Zhou (1999)) whose computational utility remains to be clarified; indeed solving such equations seems to require global computations which are no more feasible than solving Bellman's equation. One could of course apply numerical methods for deterministic control and ignore the noise (as is often done in practice), but ideally this would be based on some understanding of how the resulting trajectories relate to the underlying stochastic problem.

1.1 The goal of the paper

Our goal here is as follows. Given a stochastic optimal control problem, we would like to construct a deterministic problem whose optimal trajectory is related to the stochastic problem in some sensible way. For example, this trajectory could be the mean or the mode of the distribution induced by the optimally-controlled stochastic dynamics. Currently this is only possible for linear systems; we would like to handle nonlinear systems.

1.2 An ideal but infeasible approach

One approach, specific to continuous-time controlled diffusions with drift $f(x,u)$ and noise covariance $\Sigma(x,u)$, would be to exploit the structure of the Hamilton-Jacobi-Bellman (HJB) equation

$$-v_t = \min_u \left\{ \ell + f^{\mathsf{T}} v_x + \frac{1}{2} \operatorname{tr}(\Sigma v_{xx}) \right\} \tag{2}$$

Subscripts denote partial derivatives. Similar to (1), this PDE couples nearby states and calls for a global solution. However, suppose we augmented the cost rate as

$$\widetilde{\ell} = \ell + \frac{1}{2} \operatorname{tr}(\Sigma v_{xx}) \tag{3}$$

and then made the problem deterministic by removing the noise. The HJB equation for this deterministic problem is identical to (2) by construction, therefore the optimal cost-to-go functions and the optimal control laws for the two problems are identical. Thus the optimal trajectory for the deterministic problem (starting at any initial state) coincides with the trajectory which the optimally-controlled stochastic system would follow in the absence of noise starting at the same state. This construction achieves the above-stated goal. It is of course infeasible computationally, because it requires access to the optimal cost-to-go v for the stochastic problem. Nevertheless it suggests that augmenting the cost rate with a term that depends on the noise amplitude, and then removing the noise, is a good idea which we will pursue.

1.3 Our approach

We will be able to achieve our goal for a restricted class of stochastic optimal control problems, which have been the focus of our recent work and are sufficiently general to include many systems of practical interest (see below). It will turn out that, for problems in this class, the probability of a trajectory under the optimally-controlled stochastic dynamics can be written down explicitly, without actually knowing what the optimal control law is. This probability will only depend on quantities that are given in the problem formulation (i.e. the dynamics and cost functions). The negative log of the probability will be interpreted as the total cost for a deterministic optimal control problem. The resulting augmentation of the cost rate will again depend on the noise amplitude, as in (3), but will not depend on the cost-to-go. The optimal deterministic trajectory will then coincide with the most likely trajectory of the optimally-controlled stochastic system.

1.4 Outline of the rest of the paper

Section 2 provides a self-contained summary of the problem class to which the present results apply. There are actually two classes, corresponding to discrete and continuous time, that are related but nevertheless differ significantly in the technical details. The discrete-time problem class was developed in our recent work; see Todorov (2006, 2008, 2009b,a, 2010). The continuous-time problem class was studied by Fleming and Mitter (1982); Mitter and Newton (2003); Kappen (2005); Todorov (2009b). These problem classes have a number of interesting properties, perhaps the most notable being the linearity of the exponentiated Bellman equation.

Section 3 develops the main results. In discrete time the derivation is straightforward and general, and yields an additional result: the same trajectory is the solution to a maximum a posteriori estimation problem. In continuous time we use the Onsager-Machlup functional to define the probability of a trajectory, and obtain results restricted to state-independent noise and smooth trajectories.

Section 4 explores the relation between the discrete and continuous-time results, using the fact that the continuous-time problem class can be obtained by specializing the discrete-time formulation and taking a certain limit. We find that, even though the continuous-time results require state-independent noise, discrete-time approximations can be obtained when the noise is state-dependent. We also find that explicit Euler discretization of the time axis is insufficient, and an implicit scheme is needed to make the discrete and continuous-time results agree in the limit.

Section 5 presents a numerical example where the noise amplitude has a large effect on the optimal behavior. As expected from the theory, the optimal trajectory for the deterministic problem is affected by the amount of noise in the stochastic problem, and accurately captures the typical behavior of the optimally-controlled stochastic system for any noise level.

The main results developed here were briefly stated without proof in Todorov (2009b), and the reader was then referred to an earlier unpublished version of the present manuscript (ref [12] in Todorov (2009b)).

2. THE PROBLEM CLASS

2.1 Discrete-time formulation (Problem S1)

Consider a discrete-time continuous-state system where $x \in \mathbb{R}^n$ is the state vector, and let $y \sim p(\cdot|x)$ be a one-step transition probability corresponding to the notion of passive dynamics. y denotes the next state. The controlled dynamics are

$$y \sim \pi(\cdot|x)$$

where π is a one-step transition probability distribution directly specified by the controller (instead of specifying a control signal which then yields a transition probability distribution). The only restriction we impose is that $\pi(y|x) = 0$ whenever $p(y|x) = 0$. The cost rate is

$$\ell_{S1}(x, \pi(\cdot|x)) = q(x) + D_{KL}(\pi(\cdot|x), p(\cdot|x)) \quad (4)$$

The first term is a state cost used to encode the task goal, while the second term (Kullback-Leibler divergence) is an energy cost used to penalize the controller for pushing the system away from the passive dynamics. There is also a final cost $b(x)$. Thus the total cost is

$$b(x_T) + \sum_{t=0}^{T-1} \ell_{S1}(x_t, \pi(\cdot|x_t, t))$$

The objective is to minimize the expectation of the total cost. The functions p, q, b can be arbitrary, making this problem class very general. Examples can be found in Todorov (2009b). We will define this as Problem S1.

The key property of this problem class is as follows. Define the function $z(x,t) = \exp(-v(x,t))$ where $v(x,t)$ is the optimal cost-to-go function. Then the optimal control law π^* can be found in closed form given z:

$$\pi^*(y|x,t) = \frac{p(y|x)\, z(y, t+1)}{\mathcal{G}[z(\cdot, t+1)](x)} \quad (5)$$

Here $\mathcal{G}$ is the linear integral operator

$$\mathcal{G}[w(\cdot)](x) = \int p(y|x)\, w(y)\, dy$$

This operator computes the expected value of its function argument at the next state under the passive dynamics. Bellman's equation now becomes linear in terms of z:

$$z(x,t) = \exp(-q(x))\, \mathcal{G}[z(\cdot, t+1)](x) \quad (6)$$

At the final time we have $z(x,T) = \exp(-b(x))$.

2.2 Continuous-time formulation (Problem S2)

Consider a control-affine Ito diffusion

$$dx = a(x)\, dt + B(x)\, u dt + C(x)\, d\omega \quad (7)$$

where $u \in \mathbb{R}^m$ is the control vector and $\omega \in \mathbb{R}^m$ is a standard Brownian motion process. Denote the noise covariance $\Sigma(x) = C(x)\, C(x)^{\mathsf{T}}$. Let the cost rate be

$$\ell_{S2}(x, u) = q(x) + \frac{1}{2} u^{\mathsf{T}} R(x)\, u \quad (8)$$

We assume that R is s.p.d, and C, B, R are such that

$$C(x)\, C(x)^{\mathsf{T}} = B(x)\, R(x)^{-1}\, B(x)^{\mathsf{T}} \quad (9)$$

This means that the noise and controls act within the same space (similar to the restriction on π). The total cost is

$$b(x(T)) + \int_0^T \ell_{S2}(x(t), u(t))\, dt$$

We will define this as Problem S2.

We will need the following properties of this problem class. It is well known (e.g. Stengel (1994)) that the optimal control law is

$$u^*(x,t) = -R(x)^{-1} B(x)^{\mathsf{T}} v_x(x,t)$$

Then, using (9), the drift in the optimally-controlled stochastic dynamics is

$$g(x,t) = a(x) - \Sigma(x) v_x(x,t) \tag{10}$$

Furthermore, substituting u^* in the HJB equation yields

$$-v_t = q + a^{\mathsf{T}} v_x + \frac{1}{2}\operatorname{tr}(\Sigma v_{xx}) - \frac{1}{2} v_x^{\mathsf{T}} \Sigma v_x \tag{11}$$

When expressed in terms of $z = \exp(-v)$, the latter PDE becomes linear:

$$-z_t = \mathcal{L}[z] - qz$$

Here $\mathcal{L}$ is the 2nd-order linear differential operator

$$\mathcal{L}[w] = a^{\mathsf{T}} w_x + \frac{1}{2}\operatorname{tr}(\Sigma w_{xx})$$

which is also the generator of the passive dynamics.

2.3 Relation between discrete and continuous time

While the above problem classes look very different, they both involve the key assumption that the noise and control act within the same space, and both yield a linear Bellman equation in terms of z. Indeed it can be shown that the continuous-time problem can be obtained by specializing the discrete-time problem and taking a limit, as follows. Using explicit Euler discretization of the time axis with time step h, (7) can be approximated as

$$\pi_{h,u}(\cdot|x) = \mathcal{N}(x + ha(x) + hB(x)u,\ h\Sigma(x))$$

where $\mathcal{N}$ denotes a Gaussian. The passive dynamics are then $p_h(\cdot|x) = \pi_{h,0}(\cdot|x)$. The generic KL divergence cost in (4) now specializes to

$$D_{KL}(\pi_{h,u}(\cdot|x), \pi_{h,0}(\cdot|x)) = \frac{h}{2} u^{\mathsf{T}} R(x) u \tag{12}$$

which is proportional to the quadratic control cost assumed in (8). It can further be shown that in the limit $h \to 0$ the continuous-time problem is recovered. This is done using the fact that

$$\mathcal{G}_h[z] = z + h\mathcal{L}[z] + o(h^2)$$

The details can be found in Todorov (2009b).

Remark. In the above problem formulations, the functions q, p, a, B, C, R can depend on time without affecting any of the derivations. We have omitted the time indices for clarity. Note that we are working with finite-horizion problems, so the cost-to-go function and the control law always depend on time.

3. THE MOST LIKELY TRAJECTORY

3.1 Discrete-time formulation

Theorem 1. The probability $p^*(x_1, x_2, \cdots x_T|x_0)$ that the optimally-controlled stochastic dynamics for Problem S1 follow trajectory $x_1, x_2, \cdots x_T$ starting at state x_0 is

$$p^* = \frac{z(x_T, T)}{z(x_0, 0)} \prod_{t=0}^{T-1} \exp(-q(x_t))\, p(x_{t+1}|x_t) \tag{13}$$

Proof. Since the controlled dynamics are Markov,

$$p^*(x_1, x_2, \cdots x_T|x_0) = \prod_{t=0}^{T-1} \pi^*(x_{t+1}|x_t, t)$$

Using (5) and (6), we can express π^* as

$$\pi^*(x_{t+1}|x_t, t) = \exp(-q(x_t))\, p(x_{t+1}|x_t) \frac{z(x_{t+1}, t+1)}{z(x_t, t)}$$

Substituting in the above expression for p^* and noting that all the z terms cancel (except for the first and the last), we obtain (13). ■

Now we can define a deterministic optimal control problem whose solution coincides with the maximum of p^*. The total cost for this problem will be the negative log of (13). Since x_0 is fixed, the terms dependent on x_0 can be omitted – which is crucial because we do not know $z(x_0, 0)$ and computing it would defeat the purpose of the paper. The only nontrivial step here is interpreting $-\log p(x_{t+1}|x_t)$ as a control cost, which can be done by defining arbitrary deterministic dynamics.

Definition. Let Problem D1 be a deterministic optimal control problem with dynamics $y = s(x, u)$ and admissible control set $\mathcal{U}(x)$, such that $p(s(x,u)|x) > 0$ if and only if $u \in \mathcal{U}(x)$. The cost rate is

$$\widetilde{\ell}_{D1}(x, u) = q(x) - \log p(s(x,u), x) \tag{14}$$

and the final cost is $b(x)$ evaluated at the final time T.

The above technicality is needed to ensure that the admissible trajectories in the deterministic problem have nonzero probability in the stochastic problem and vice versa. In practice p is often Gaussian (perhaps over some m-dimensional subspace), in which case $\mathcal{U}(x)$ is simply $\mathbb{R}^m$. We now state the main result of this subsection.

Theorem 2. For any initial state x_0, an optimal trajectory in Problem D1 is a most likely trajectory in Problem S1 and vice versa. Furthermore the local minima of the total cost in Problem D1 coincide with the local maxima of the trajectory probability in Problem S1.

Proof. Since log is a convex function, maximizing (13) is equivalent to minimizing its negative log, which is

$$b(x_T) - v(x_0, 0) + \sum_{t=0}^{T-1} q(x_t) - \log p(x_{t+1}|x_t)$$

Since x_0 is fixed, $v(x_0, 0)$ is constant and can be omitted. Using the fact that $x_{t+1} = s(x_t, u_t)$, the above expression becomes

$$b(x_T) + \sum_{t=0}^{T-1} \widetilde{\ell}_{D1}(x_t, u_t)$$

which coincides with the total cost for Problem D1. ■

Remark. The relation between the KL divergence cost in Problem S1 and the control cost $-\log p(x_{t+1}|x_t)$ in Problem D1 can be illuminated by evaluating the KL divergence at the deterministic dynamics, i.e. the Dirac delta function δ centered at x_{t+1}. We have

$$\begin{aligned} D_{KL}(\delta, p) &= \int \delta(x - x_{t+1}) \log \frac{\delta(x - x_{t+1})}{p(x|x_t)} dx \\ &= -H[\delta] - \log p(x_{t+1}|x_t) \end{aligned}$$

Here $H[\delta] = -\infty$ is the differential entropy of the delta function. This term is infinite however it does not depend on the state. The remaining term is the control cost in Problem D1.

The trajectory probability given by Theorem 1 can also be interpreted as the Bayesian posterior for a certain estimation problem, defined as follows.

Definition. Let Problem E1 be a Bayesian estimation problem with hidden state dynamics $y \sim p(\cdot|x)$. The initial state x_0 is given. A sequence of observations $o_1, o_2, \cdots o_T$ is drawn according to $o_t \sim p_o(\cdot|x_t)$.

Let $p^B(x_1, x_2, \cdots x_T|x_0, o_1, o_2, \cdots o_T)$ denote the Bayesian posterior over hidden state trajectories, which is

$$p^B \propto \prod_{t=0}^{T-1} p_o(o_{t+1}|x_{t+1})\, p(x_{t+1}|x_t) \tag{15}$$

This expression can be made equal to (13) by choosing a suitable observation model, as follows.

Theorem 3. If the observations o_t and emission probabilities $p_o(\cdot|x)$ in problem E1 are such that

$$p_o(o_t|x_t) = \exp(-q(x_t))$$

for all t, then the trajectory probability p^* in problem S1 equals the Bayesian posterior p^B in problem E1.

Proof. Expressions (15) and (13) are identical up to an unspecified normalization constant in (15) and the terms $z(x_0, 0)$ and $\exp(-q(x_0))$ in (13). However both p^* and p^B are probability densities and sum to 1, thus the normalization constants (which depend on the fixed x_0) are the same. ■

Corollary. The maximum a posteriori estimator (i.e. the trajectory maximizing p^B) in problem E1 equals the most likely trajectory in problem S1.

Note that in order to construct the estimation problem E1 equivalent to our stochastic control problem S1, we had to fix the measurements. Intuitively one can think of this as "observing" that the goals have been achieved, and estimating the control signals responsible for goal achievement. A number of additional results related to this estimation-control duality can be found in Mitter and Newton (2003); Todorov (2008).

3.2 Continuous-time formulation

In discrete time we were able to write down a formula for the probability distribution over trajectories of the optimally-controlled stochastic system. This was possible because the trajectories were finite-dimensional. In continuous time the trajectories are infinite-dimensional objects and a proper probability distribution is not available. So what does "most likely" mean? Our definition will be based on the Onsager-Machlup functional – which is the ratio of probability mass in the vicinity of two smooth trajectories. Assume for now that Σ is positive definite and does not depend on x. Consider $\mathbb{R}^n$ as a Riemannian manifold with metric Σ^{-1} and let $\rho(\cdot,\cdot)$ be the geodesic distance. Suppose $\varphi(\cdot)$ and $\psi(\cdot)$ are smooth trajectories, $\varphi(0) = \psi(0) = x(0)$, and $x(\cdot)$ is sampled from an Ito diffusion with drift $g(x)$ and noise covariance Σ. Then it is known (Takahashi and Watanabe (1981); Capitaine (2000)) that the limit

$$\lim_{\epsilon \to 0} \frac{p(\sup_t \rho(x(t), \varphi(t)) < \epsilon)}{p(\sup_t \rho(x(t), \psi(t)) < \epsilon)} \tag{16}$$

exists and is equal to

$$\exp\left(-\int_0^T L(\varphi(t), \dot{\varphi}(t)) - L\left(\psi(t), \dot{\psi}(t)\right) dt\right) \tag{17}$$

where the function L is defined as

$$L_{g,\Sigma}(x, w) = \frac{1}{2}(g(x) - w)^{\mathsf{T}} \Sigma^{-1}(g(x) - w) + \frac{1}{2}\operatorname{div}(g(x))$$

Instead of the supremum norm in (16) one could use other norms leading to the same result, see Capitaine (2000). This motivates the following definition.

Definition. The Onsager-Machlup relative probability of a smooth trajectory $\varphi(\cdot)$ under an Ito diffusion with drift g and noise covariance Σ is

$$p_{OM}(\varphi(\cdot); g, \Sigma) = \exp\left(-\int_0^T L_{g,\Sigma}(\varphi(t), \dot{\varphi}(t))\, dt\right) \tag{18}$$

Varying ψ in (17) has a scaling effect which does not affect the maximization with respect to φ, thus it makes sense to define p_{OM} only in terms of φ. This functional was previously used to characterize the most likely trajectories of uncontrolled diffusions, see Durr and Bach (1978). Here we will apply it to our optimally-controlled diffusion, with drift g given by (10), and use the notation p^*_{OM}. The dependence on g, Σ will be suppressed.

As in the discrete-time case, our goal now is to express p^*_{OM} in a way which does not depend on the unknown cost-to-go v appearing in (10), but only depends on quantities given in the problem formulation. The result is as follows.

Theorem 4. Let Σ in Problem S2 be constant and s.p.d. Then the Onsager-Machlup relative probability of a smooth feasible trajectory $x(\cdot)$ under the optimally-controlled stochastic dynamics is

$$p^*_{OM} = \frac{z(x(T), T)}{z(x_0, 0)} \exp\left(-\int_0^T \widetilde{\ell}_{D2}(x(t), u(t))\, dt\right) \tag{19}$$

where $u(\cdot)$ is any solution to

$$\dot{x}(t) = a(x(t)) + B(x(t))\, u(t) \tag{20}$$

and the augmented cost rate is

$$\widetilde{\ell}_{D2}(x, u) = \ell_{S2}(x, u) + \frac{1}{2}\operatorname{div}(a(x)) \tag{21}$$

Proof. First note that $u(\cdot)$ exists because $\Sigma = BR^{-1}B^{\mathsf{T}}$ is s.p.d. by assumption and thus B has full row-rank. Even if B were rank-defficient, (9) implies that the noise and the controls act in the same space thus $u(\cdot)$ exists as long as $x(\cdot)$ is feasible. Using (9) and the identity $\operatorname{div}(\Sigma v_x) = \operatorname{tr}(\Sigma v_{xx})$, we have

$$L(x, \dot{x}, t) =$$

$$\tfrac{1}{2}(Bu + \Sigma v_x)^{\mathsf{T}} \Sigma^{-1}(Bu + \Sigma v_x) + \tfrac{1}{2}\operatorname{div}(a - \Sigma v_{\mathbf{x}}) =$$

$$\tfrac{1}{2}u^{\mathsf{T}}Ru + \tfrac{1}{2}\operatorname{div}(a) + u^{\mathsf{T}}B^{\mathsf{T}}v_x + \tfrac{1}{2}v_x^{\mathsf{T}}\Sigma v_x - \tfrac{1}{2}\operatorname{tr}(\Sigma v_{xx})$$

Next we use the unique properties of our problem class to eliminate the dependence on v. From (11) it follows that the Σ-dependent expression on last line above equals $v_t + q + a^{\mathsf{T}}v_x$. Using (20) and the fact that $\dot{v} = v_t + \dot{x}^{\mathsf{T}}v_x$,

$$L(x, \dot{x}, t) = \widetilde{\ell}_{D2}(x, u) + \dot{v}(x, t)$$

The time-integral of $\dot{v}(x, t)$ is $b(x(T)) - v(x_0, 0)$, which yields (19). ■

Remark. The assumption that Σ is positive definite was needed because the treatments of the Onsager-Machlup functional we have seen rely on it. We suspect that this assumption can be relaxed, given that Σ^{-1} does not appear anywhere in the final results. Below we will approximate continuous-time diffusions with discrete-time problems and show how a singular Σ can be handled.

Remark. The assumption that Σ does not depend on x is more fundamental and appears to be unavoidable in continuous time (although the discrete-time approximations below will be able to avoid it). The Onsager-Machlup functional for state-dependent noise requires two extensions. The first is to define the Riemannian metric $\Sigma(x)^{-1}$ and add its scalar curvature to the function L. This extension is benign because it does not add terms that depend on v. The second extension is to augment the components g_i of the drift vector g as follows:

$$\widehat{g}_i(x,t) = g_i(x,t) + \frac{1}{2}\sum_{jk}\Gamma^i_{jk}(x)\,\Sigma_{jk}(x)$$

where $\Gamma^i_{jk}(x)$ are the corresponding Christoffel symbols. Plugging this $\widehat{g}$ in L yields a product between the new term and v_x, and so the dependence on v can no longer be removed. A further complication along the same lines is that $\operatorname{div}(\Sigma(x)\,v_x(x))$ now gives rise to an extra term which depends on v. Thus, when $\Sigma(x)$ is state-dependent, computing the most likely continuous-time trajectory requires access to v.

We can now define the deterministic optimal control problem corresponding to Problem S2 as follows.

Definition. Let Problem D2 be a deterministic optimal control problem with dynamics (20), cost rate (21) and final cost $b(x)$ evaluated at time T.

Theorem 5. If the optimal trajectory in Problem D2 for initial state x_0 is smooth, then it coincides with the most likely trajectory (in the sense of p^*_{OM}) in Problem S2.

Proof. The negative log of p^*_{OM} as given by (19) equals the total cost for Problem D2, up to a term that only depends on the fixed x_0. ■

Note that the augmented cost (21) involves $\operatorname{div}(a(x))$ rather than a term dependent on the noise amplitude. Since Σ is constant by assumption, such a term would not affect the optimal trajectory anyway. On the other hand, the divergence implicitly takes noise into account by penalizing states where disturbances are amplified by the passive dynamics. Furthermore, (9) implies that the cost rate (and thus the optimal deterministic trajectory) are affected by changes in Σ. If for example we scaled Σ by c, we would have to scale R by c^{-1}.

4. DISCRETIZING THE CONTINUOUS PROBLEM

In discrete time we obtained a very general result, while in continuous time we had to assume smooth trajectories and constant s.p.d. Σ. Yet, as explained in Section 2.3, the continuous-time problem can be recovered by discretizing the time axis, constructing a corresponding problem in our discrete-time class, and then taking a continuous-time limit. Can the same limiting procedure be used to extend our results regarding most likely trajectories, and in particular relax the assumptions on Σ? Below we show that although this is not possible in a strict sense, one can obtain approximations by using a small time step h instead of going all the way to the limit $h \to 0$.

4.1 State-dependent noise

Consider Problem S2 with $\Sigma(x)$ now a function of x, but still s.p.d. From Section 2.3, the discretized problem (S1) has passive dynamics

$$p(\cdot|x) = \mathcal{N}(x + ha(x),\ h\Sigma(x)) \tag{22}$$

and state-dependent cost rate $hq(x)$. We now seek to construct the corresponding deterministic Problem D1. Recall that we are free to choose the deterministic dynamics $s(x,u)$. The natural choice here is

$$s(x,u) = x + ha(x) + hB(x)\,u$$

The augmented cost rate $\widetilde{\ell}_{D1}$ for Problem D1 is then (14). Using the formula for a Gaussian density, we have

$$\begin{aligned}\widetilde{\ell}_{D1}(x,u) &= hq(x) + \frac{h}{2}u^{\mathsf{T}}R(x)\,u + \frac{1}{2}\log\left(\det\left(2\pi h\Sigma(x)\right)\right)\\ &= h\ell_{S2}(x,u) + \frac{n}{2}\log(2\pi h) + \frac{1}{2}\log\left(\det\left(\Sigma(x)\right)\right)\end{aligned}$$

The second term above does not depend on x and can be omitted. The last term depends on x and is not scaled by h. Since the continuous-time limit $h \to 0$ described in Section 2.3 is taken after dividing by h, the last term goes to infinity and so we cannot obtain a continuous-time result similar to Theorem 5 for state-dependent noise. Away from the limit however we obtain a useful approximation.

4.2 Degenerate diffusions

Here we consider the case when $\Sigma(x)$ is singular. This is important in modeling 2nd-order mechanical systems driven by noisy forces; in that case the state vector includes position and velocity, and only the velocity is directly affected by the noise, so $\Sigma(x)$ is singular. Assume that $m < n$ and the n-by-m matrices $B(x)$ and $C(x)$ have rank m. Then we can find an invertible m-by-m matrix $D(x)$ such that the matrix

$$V(x) = B(x)\,D(x)$$

is orthonormal, i.e. $V^{\mathsf{T}}V = I$. Define the s.p.d matrix

$$S(x) = \left(D(x)^{\mathsf{T}}\,R(x)\,D(x)\right)^{-1}$$

Then from (9) it follows that

$$\Sigma(x) = V(x)\,S(x)\,V(x)^{\mathsf{T}}$$

The Gaussian $p(\cdot|x)$ given by (22) is degenerate over $\mathbb{R}^n$, however it is a proper Gaussian within the m-dimensional subspace spanned by the columns of $V(x)$. So we can represent the next state y as

$$y = x + ha(x) + V(x)\,\xi \tag{23}$$

where $\xi \in \mathbb{R}^m$ has Gaussian distribution with mean 0 and covariance $hS(x)$. For any y such that $y - x - ha(x) \in \operatorname{span}(V(x))$ there is a unique value of ξ satisfying (23). Omitting some algebra, the modified cost rate for the discretized problem becomes

$$\widetilde{\ell}_{D1}(x,u) = h\ell_{S2}(x,u) + \frac{1}{2}\log\left(\det\left(S(x)\right)\right)$$

The results for the non-singular case can be recovered by setting $n = m$, $S = \Sigma$, $V = I$.

4.3 Implicit discretization

A somewhat surprising aspect of the above analysis is that we did not recover the cost augmentation term $\mathrm{div}\,(a\,(x))$ which showed up in our continuous-time results. This turns out to be an artefact of the explicit Euler approximation. Indeed there are better ways to approximate diffusion processes, including implicit methods, semi-implicit methods and Milstein methods, see Gardiner (2004). Particularly good results have been obtained using local linear approximations, see Shoji (1998). The simplest such approach is to approximate the increment within one time step as the stochastic process

$$d\varepsilon = \left(a\,(x) + a_x\,(x)^{\mathsf{T}}\,\varepsilon\right) dt + C\,(x)\,d\omega$$

where $\varepsilon\,(0) = 0$ and the next state is $y = x + \varepsilon\,(h)$. The covariance of $\varepsilon\,(h)$ is

$$\int_0^h \exp\left(\tau a_x^{\mathsf{T}}\right) \Sigma \exp\left(\tau a_x\right) d\tau$$

The above matrix integral cannot be evaluated in closed form. However consider the mid-point approximation

$$h \exp\left(\frac{h}{2} a_x^{\mathsf{T}}\right) \Sigma \exp\left(\frac{h}{2} a_x\right)$$

With this improved approximation to the noise covariance in place of $h\Sigma\,(x)$, and using the identity $\det\left(\exp\left(M\right)\right) = \exp\left(\mathrm{tr}\,\left(M\right)\right)$ for any M, the augmented cost for the above discretized problem is replaced with

$$\widetilde{\ell}_{D1}\,(x,u) = h\ell_{S2}\,(x,u) + \frac{h}{2}\,\mathrm{div}\,(a\,(x)) + \frac{1}{2}\log\left(\det\left(\Sigma\,(x)\right)\right)$$

Thus a more accurate discretization of the diffusion process yields an augmented cost combining the correction terms that arise from the continuous-time setting and from the explicit Euler approximation.

Another improvement we can make is as follows. Since the Jacobian $a_x\,(x)$ has to be evaluated anyway, we might as well use it to construct a better discrete-time approximation to the deterministic dynamics:

$$s\,(x,u) = x + h \exp\left(\frac{h}{2} a_x^{\mathsf{T}}\,(x)\right) \left(a\,(x) + B\,(x)\,u\right)$$

This actually correspond to an explicit Euler approximation for a modified problem where a, B, C have been premultiplied by the matrix $\exp\left(\frac{h}{2} a_x^{\mathsf{T}}\,(x)\right)$. The key relation (9) still holds in this modified problem.

5. NUMERICAL EXAMPLE

We now present a scalar example illustrating the relation between stochastic problems and their deterministic counterparts. The controlled Ito diffusion is

$$dx = \left(a\,(x) + u\right) dt + \sigma d\omega$$

The drift in the passive dynamics (Fig. 1A) is

$$a\,(x) = 2\exp\left(-\left(x-1\right)^2\right) - 2\exp\left(-1\right) - 0.2x$$

The cost rate is

$$\ell_{S2}\,(x,u) = \frac{1}{2\sigma^2} u^2$$

There is no state-dependent cost rate, i.e. $q\,(x) = 0$. The final cost is

$$b\,(x) = 5x^2$$

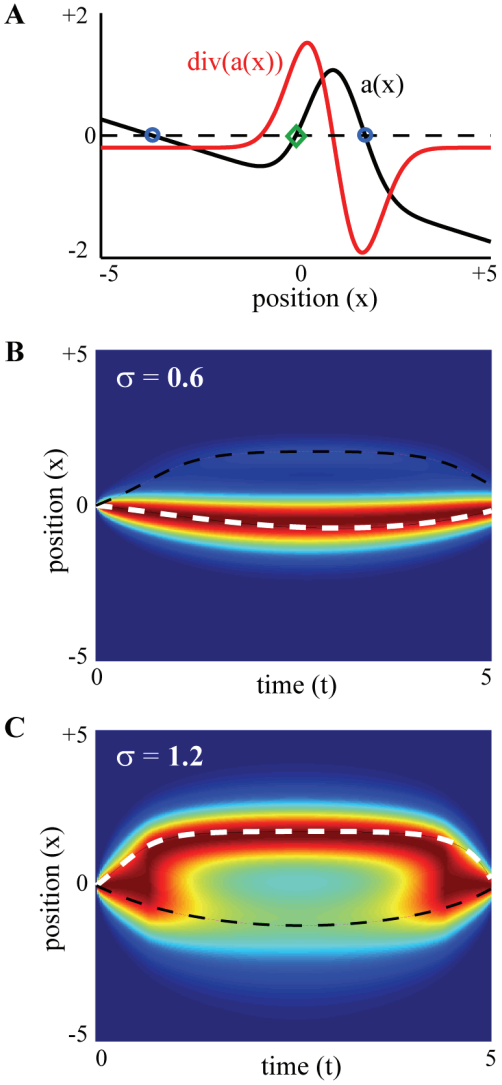

Fig. 1. **A**– The drift function $a\,(x)$ and its derivative $a'\,(x) = \mathrm{div}\,(a\,(x))$. The unstable equilibrium at $x = 0$ is marked with a diamond. The two stable equilibria are marked with circles. **B,C**– The intensity plot is the trajectory density under the optimally-controlled stochastic dynamics, marginalized at each point in time. The dashed lines are the optimal trajectory (white) and local minimum (black) for the corresponding deterministic problem. Note the effect of varying the noise amplitde σ between (B) and (C).

The final time is $T = 5$. The initial state is $x_0 = 0$. The augmented cost is

$$\widetilde{\ell}_{D2}\,(x,u) = \frac{1}{2\sigma^2} u^2 + a'\,(x)$$

The optimal solution in the absence of noise is $x\,(t) = u\,(t) = 0$. However the function $a\,(x)$ has been designed to make this solution unstable. Note that we have $\mathrm{div}\,(a\,(x)) = a'\,(x) > 0$ at unstable equilibria and $a'\,(x) < 0$ at stable equilibria. Thus the extra cost term $a'\,(x)$ has the effect of pushing the optimaly-controlled deterministic system away from unstable equilibria and towards stable equilibria. This effect (i.e. the relative importance of the extra cost) increases with σ.

Numerical results are shown in Fig. 1B for $\sigma = 0.6$ and in Fig. 1C for $\sigma = 1.2$. The globally-optimal solution to the stochastic problem is approximated on a 301-by-251 grid spanning $x \in [-5, 5]$ and $t \in [0, 5]$. The time step is $h = 0.02$. The discrete-time passive dynamics correspond to the explicit Euler approximation to the uncontrolled diffusion. For better visualization, all function values plotted in the figure have been scaled so that for every t the maximum over x is 1. The intensity plot is the trajectory probability p^* marginalized at each point in time. We denote this marginal with $m^*(x,t)$.

The inf1uece of σ can be understood as follows. For small noise it is relatively easy to keep the system near the unstable equilibrium $x = 0$. At the same time the controls are expensive, and if the system is allowed to deviate far from the desired final state $x = 0$, getting it back will be costly. Thus $m^*(x,t)$ is large near $x = 0$. For larger noise the situation is reversed. Staying at the unstable equilibrium is hard, while at the same time the controls are cheaper, thus getting the system back in the last minute (using large controls) is not so costly.

Optimal trajectories for the corresponding deterministic problems are also shown. These trajectories are found using our iLQG method developed in Todorov and Li (2005). Brief1y, each iteration starts with a state-control trajectory, computes a local linear approximation to the dynamics and a local quadratic approximation to the cost, and uses generalized Riccati equations combined with Levenberg-Marquardt optimization to find a better state-control trajectory. The method converges to an extremal trajectory in about 15 iterations. Different solutions are obtained for different initialization, however for the problems studies here we observed that all solutions fall in one of two clusters. The white line shows the best solution found; the black line shows the best solution in the other cluster.

There is good agreement between the optimal deterministic trajectories and the maxima of the marginal probability distribution $m^*(x,t)$. The latter distribution is bimodal over x. The optimal deterministic trajectory is near the peak of m^* while the second-best deterministic trajectory is near the smaller mode. Note that the optimal deterministic trajectory does not exactly coincide with the maximum of m^*, and indeed it is not guaranteed to. Instead it coincides with the maximum of p^*, which can be somewhat different due to the marginalization.

6. SUMMARY

In this paper we showed that, for certain problem classes, the most likely trajectory of the optimally-controlled stochastic system coincides with the optimal trajectory of a deterministic system with augmented cost. This makes it possible to apply deterministic numerical methods for optimal control and construct local solutions to stochastic control problems. Note that some of these methods (DDP and iLQG) construct not only an open-loop trajectory but also a linear feedback control law, which is particularly useful in the context of stochastic control. We also showed that, in discrete time, the most likely trajectory coincides with the maximum a posteriori estimator for a related Bayesian inference problem.

ACKNOWLEDGEMENTS

This work was supported by the US National Science Foundation.

REFERENCES

Bryson, A. and Ho, Y. (1969). *Applied Optimal Control*. Blaisdell Publishing Company, Massachusetts.

Capitaine, M. (2000). On the Onsager-Machlup functional for elliptic diffusion processes. *Seminaire de Probabilities (Strasbourg)*, 34, 313–328.

Durr, D. and Bach, A. (1978). The Onsager-Machlup function as Lagrangian for the most probable path of a diffusion process. *Communications in Mathematical Physics*, 60, 153–170.

Fleming, W. and Mitter, S. (1982). Optimal control and nonlinear filtering for nondegenerate diffusion processes. *Stochastics*, 8, 226–261.

Gardiner, C. (2004). *Handbook of Stochastic Methods, 3rd edition*. Springer, Berlin.

Jacobson, D. and Mayne, D. (1970). *Differential Dynamic Programming*. Elsevier, New York.

Kappen, H. (2005). Linear theory for control of nonlinear stochastic systems. *Physical Review Letters*, 95.

Mitter, S. and Newton, N. (2003). A variational approach to nonlinear estimation. *SIAM J Control Opt*, 42, 1813–1833.

Shoji, I. (1998). Approximation of continuous time stochastic processes by a local linearization method. *Mathematics of Computation*, 67(221), 287–298.

Stengel, R. (1994). *Optimal Control and Estimation*. Dover, New York.

Takahashi, Y. and Watanabe, S. (1981). The probability functional (Onsager-Machlup function) of stochastic processes. *Springer Lecture Notes in Mathematics*, 851, 433–463.

Todorov, E. (2006). Linearly-solvable Markov decision problems. *Advances in Neural Information Processing Systems*.

Todorov, E. (2008). General duality between optimal control and estimation. *IEEE Conference on Decision and Control*, 47, 4286–4292.

Todorov, E. (2009a). Compositionality of optimal control laws. *Advances in Neural Information Processing Systems*.

Todorov, E. (2009b). Efficient computation of optimal actions. *PNAS*, 106, 11478–11483.

Todorov, E. (2010). Policy gradient methods for linearly-solvable mdps. *Advances in Neural Information Processing Systems*.

Todorov, E. and Li, W. (2005). A generalized iterative LQG method for locally-optimal feedback control of constrained nonlinear stochastic systems. *American Control Conference*, 300–306.

Yong, J. and Zhou, X. (1999). *Stochastic controls: Hamiltonian systems and HJB equations*. Springer, New York.