Stochastic Optimal Control as Non-equilibrium Statistical Mechanics:
Calculus of Variations over Density and Current

Vladimir Y. Chernyak\textsuperscript{a,b}, Michael Chertkov\textsuperscript{b}, Joris Bierkens\textsuperscript{c}, and Hilbert J. Kappen\textsuperscript{c}

\textsuperscript{a}Department of Chemistry, Wayne State University, 5101 Cass Ave, Detroit, MI 48202, USA
\textsuperscript{b}Center for Nonlinear Studies and Theoretical Division, LANL, Los Alamos, NM 87545, USA and
\textsuperscript{c}SNN Adaptive Intelligence, Radboud University, Nijmegen PO. Box 9101, 6500 HB, Nijmegen, the Netherlands

(Dated: June 28, 2013)

In Stochastic Optimal Control (SOC) one minimizes the average cost-to-go, that consists of the cost-of-control (amount of efforts), cost-of-space (where one wants the system to be) and the target cost (where one wants the system to arrive), for a system participating in forced and controlled Langevin dynamics. We extend the SOC problem by introducing an additional cost-of-dynamics, characterized by a vector potential. We propose derivation of the generalized gauge-invariant Hamilton-Jacobi-Bellman equation as a variation over density and current, suggest hydrodynamic interpretation and discuss examples, e.g., ergodic control of a particle-within-a-circle, illustrating non-equilibrium space-time complexity.

In its standard setting the problem of the Stochastic Optimal Control (SOC) involves minimizing the average over stochastic trajectories of the cost-to-go, which consists of the cost of using the control field, the cost of arriving to a certain position $x$ in the configuration space $\mathcal{M}$, described by a potential function $\varphi(x)$, and the cost accumulated along the trajectory, described by a time-dependent potential $V(x, \tau)$. The potentials $\varphi(x)$ and $V(x, \tau)$ can be viewed as variables dual to the particle probability distributions at the arrival time and during the time evolution, respectively, and thus interpreted as Lagrange multipliers.

In this Letter, continuing the thread of \cite{1-6} where methods of statistical and quantum mechanics were applied to SOC, we extend the standard cost-to-go functional by adding a term, associated with a vector potential, $A(x, \tau)$, which leads to (i) a variational derivation of the stochastic Hamilton-Jacobi-Bellman (HJB) equation and its more general Gauge Invariant (GI) version, (ii) extend the capability of the HJB equation to non-contractible cycles in phase space.

Stochastic dynamics of a particle in a compact $m$-dimensional space $\mathcal{M}$, e.g., an $m$-dimensional torus, is described by the following Langevin equation

$$r^i = \frac{d}{d\tau} \eta^i = f^i(\tau, \eta) + u^i(\tau, \eta) + \xi^i(\tau, \eta),$$

$$\langle \xi^i(\tau, \eta) \rangle = 0, \quad \langle \xi^i(\tau, \eta) \xi^j(\tau', \eta) \rangle = \kappa g^{ik} \delta(\tau - \tau'),$$

where $\tau \in [t; T]$; $f$ is the “force” field, deterministic and assumed known; $u$ is the “control” field which as we will see below is subject to our optimization/choice; and $\xi$ is a $\delta$-correlated in time, zero-mean Gaussian random field, whose correlations are fully expressed via a strictly positive symmetric matrix, $\kappa g(\eta)$, where $\kappa$ measures the noise strength, and $g$ can be viewed as a (generally space-time dependent) metric in the configuration space $\mathcal{M}$, with $g_{ij} g^{jk} = g^{ik} g_{ji} = \delta^i_j$, where we use standard in theoretical physics covariant notations, i.e., assuming summation over repeating pairs of sub/superscripts, and applying the metric to relate vectors to co-vectors, e.g., $f_i = g_{ij} f^j$.

We consider a problem describing the optimal choice of the control vector field, $u$, in Eq. (1)

$$\mathcal{C}(t, x; T) = \min_{\{u\}} \mathcal{C}(\{u\}; t, x; T), \quad \mathcal{C}(\{u\}; t, x; T) \equiv \left\langle \varphi(\eta(T)) + \int_t^T d\tau \left( \frac{1}{2} h_{ij} u^i u^j + V + A_j \eta^j \right) \right\rangle,$$  

$$B(t', \eta(t')) = \frac{\int_{\eta(t') = x} d\eta \exp \left( -\kappa \int_t^{t'} d\tau (\dot{\eta}^i - f^i - u^i) g_{ij} (\dot{\eta}^j - f^j - u^j) \right) B(t', \eta(t'))}{\int_{\eta(t) = x} d\eta \exp \left( -\kappa \int_t^{t'} d\tau (\dot{\eta}^i - f^i - u^i) g_{ij} (\dot{\eta}^j - f^j - u^j) \right)},$$  

where Eq. (1) defines averaging over stochastic trajectories in terms of a path integral; all co-vector and tensor fields in the integrand of Eq. (3) may depend explicitly on $\tau$ and $x$; minimization/variation over $\{u\}$ is functional, i.e. we minimize over all $u(\tau, \eta(\tau))$; and Eq. (4) stated as a path integral over $\eta(\tau)$ defines averaging over the stochastic trajectories evolving according to Eq. (1).

The meaning of the four terms under the average in the cost-to-go $\mathcal{C}(\{u\}; t, x; T)$ is as follows: the first local term describes the target cost, i.e. the cost for the system to arrive...
at the final moment of time \( T \) at \( \eta(T) \); the second term (which is also the first integral term) defines the cost-of-control; the third term stands for the cost-of-space, as it measures the cost depending on where the system stays in the phase space during the entire interval; finally, the last term in \( C(\{u\}; t, x; T) \), as shown in Eq. (5), represents the cost-of-dynamics, i.e. it is sensitive to how the system is moving in phase space during the period of interest (the cost is zero if the system does not move). The first three terms in Eq. (5) are standard in control theory, while the fourth term is new. It is also natural (exploiting theoretical physics jargon and intuition) to refer to \( V \) and \( A \) in Eq. (5) as the scalar and vector potentials, respectively. Obviously, the average cost-to-go \( C \) depends functionally on \( V \) and \( A \).

Following [7] we define the so-called average density and average current-density (hereafter referred to as just density and current) for the Langevin dynamics given by Eq. (1):

\[
\rho(\{u\}; t, x; T) = \int_{t}^{T} \frac{d\tau}{T-t} (\delta(\eta(\tau) - x)),
\]

(5)

\[
J(\{u\}; t, x; T) = \int_{t}^{T} \frac{d\tau}{T-t} (\eta\delta(\eta(\tau) - x)).
\]

(6)

In what follows we will simplify the notations, dropping the responding term to Eq. (8), we compute the variations over \( \rho \) for the first equation in (7), adding the corresponding term to Eq. (8), we compute the variations over \( J(\tau, x) \) and \( \rho(\tau, x) \), under condition \( \delta \rho(t; x) = 0 \) (the initial density is fixed). Combining the two variation equations results in the following closed form equation

\[
\partial_{\tau} \Phi \; = \; \kappa \sqrt{\gamma} g_{ij} \left( \partial_{j} \Phi + A_{j} \right) + g_{ij} f_{i}(\partial_{j} \Phi + A_{j}) + V - \frac{1}{2} g^{ik} h_{kl} g^{lj} (\partial_{l} \Phi + A_{i})(\partial_{j} \Phi + A_{j}),
\]

(10)

which should be solved backwards in time with the “initial” condition \( \Phi(T, x) = \varphi(x) \) that originates from the contact term in the variation over \( \rho(T) \).

A number of remarks with regards to Eq. (10) is in order. First, combining the equation emerging in the result of variation over \( J \) with the second relation in Eq. (7) one arrives at the following explicit expression of the optimal control field \( u \) via the optimal \( \Phi \):

\[
u_{i} = -\hat{h}_{ik}(\partial_{k} \Phi + A_{k}), \quad \hat{h}_{ij} h^{jk} = \delta^{k}, \quad \hat{h}_{ij} = \hat{h}_{ij} g^{jk}.
\]

Second, Eq. (10) is gauge invariant under simultaneous transformation of the scalar and vector potentials: \( A_{i} \rightarrow A_{i} + \partial_{t} \phi, V \rightarrow V + \partial_{t} \phi \), where \( \phi(\tau, x) \) is an arbitrary scalar function. Third, the Lagrangian multiplier solving Eq. (10) actually coincides with the optimal average cost-to-go function, \( C(t, x; T) = \Phi(t, x) \). The relation follows from multiplying Eq. (10) by \( \rho \), integrating the result over the \( dx/\sqrt{\gamma} \) and also over \( \tau \) in the \( [t, T] \) interval, and then comparing the final expression with Eq. (8). Fourth, the terms in Eq. (10) that contain the gauge field can be obviously absorbed into the force field \( f \) and scalar potential \( V \), resulting in the celebrated stochastic Hamilton-Jacobi-Bellman (HJB) equation of the control theory, which means that Eq. (10) is a particular case of the standard HJB equation, and access to an efficient solver of the latter provides a way to solve Eq. (10). On the other hand, the equivalence between \( C \) and \( \Phi \) means that one can also replace \( \Phi \) in Eq. (10), thus discovering that Eq. (10) can be viewed as a Gauge Invariant (GI) generalization of the HJB, rather than a particular case, since it allows to consider control over a broader set of phenomena (e.g., work/entropy generation, fluxes, etc.). Finally, combining Eqs. (10,11) we arrive at the following version of the GI-HJB equation (10) stated in terms of the control field

\[
\partial_{\tau} u_{i} = -\hat{h}_{ik}(\partial_{k} \Phi + A_{k}), \quad \hat{h}_{ij} h^{jk} = \delta^{k}, \quad \hat{h}_{ij} = \hat{h}_{ij} g^{jk}.
\]

where, to avoid tedious expressions, we assumed that both \( h \) and \( g \) metrics are time-independent.

All terms in Eq. (12) allow for a very natural hydrodynamic interpretation, where the optimal control field, \( u \), is interpreted up to a metric-dependent re-normalization as the “velocity” field that evolves backwards in time in a compact space with curvature \( g \). The second term on the l.h.s. of
Eq. (12), also the only nonlinear term in the equation, describes the “self-advection of velocity by itself”. Continuing the hydrodynamic analogy, one interprets the second term on the l.h.s as “advection by an external field $f$”. Then, the first term on the r.h.s. stands for the dissipation/viscosity, induced by the noise in the original Langevin equation (1). Finally, the last term on the r.h.s. represents pumping/injection, it may also represent constraints, e.g. expressing relations between pressure and density, the phase space hydrodynamics. Details and consequences of the ultimate relation between the control and hydrodynamics will be discussed elsewhere [8].

Scalar and vector potentials, as well as $\varphi(x)$, can also be viewed as functional Lagrangian multipliers used to fix specific forms of the density and current functions. (Note, however, that in this formulation $\rho$ and $J$ are not fully arbitrary but consistent with each other through the continuity equation, i.e. the first equation of Eqs. (7).) Therefore, under fixed and consistent scalar and vector potentials no additional optimization in Eq. (3) is needed, the control field is completely defined by the second equation of (7), and then the average cost is just the cost-of-control, $C_0$, given by Eq. (9).

This GI approach also allows to consider less restrictive cases with constraints which are linear in the density and current. For example, an interesting problem is: find the least $C_0/T$ over a non-contractible cycle $C$, i.e., the principal capability to optimize over the fluxes. Flux over a non-contractible cycle is defined as the number of times the system goes over the cycle divided by the (long) time $T$, or equivalently as the integral over the current density $J$ over the corresponding non-contractible $(m-1)$-dimensional surface (see [7] for details). This can be done by solving the stationary version of Eq. (10) with $V = 0$ and curvature free, i.e., $\partial_i \underline{A}_j - \partial_j \underline{A}_i = 0$, vector potential, which is globally still not a gradient $\underline{A}_i \neq \partial_i \varphi$.

Next we discuss our enabling ergodic case example of a particle moving along a simple circle of length $L = 2\pi$ with constant $g = h = 1$. Note, however, that in this special (and not fully representative) case the flux density and the current density coincide ($J(x) = \text{const}$). This case is analyzed by combining the stationary version of Eq. (12) with the second expression in Eq. (7), resulting in

$$u^2/2 + fu + \kappa \partial_x u = -E, \quad -\kappa \partial_x \rho + (f + u)\rho = J,$$

where $E$ and $J$ should be treated as constants, with the periodic boundary conditions $u(x + 2\pi) = u(x)$ and $\rho(x + 2\pi) = \rho(x)$ imposed. It is convenient to perform analysis implicitly by fixing the value of $E$, solving Eqs. (15), and thus determining the value of the flux $J$. This analysis, illustrated in Figs. 1a-c, suggests a few observations. First, the cases with and without flux-fixing control are significantly different.
Since the bare (without control) flux was smaller (simply zero in the cases of Figs. 1a,b) than the resulting flux under control, the density distribution is significantly more spread out in the control case, also showing appearance of some additional structure (two local maxima in density). Second, comparing Fig. 1a and Fig. 1b, different in diffusion only, we observe that increase in diffusion spreads up the density distribution, resulting in the average decrease of the cost-to-go. We observe that the extra diffusion helps advection to boost the particle transport (flux) with less cost. Third, in the case of Fig. 1c the increasing flux leads to the control field splitting into two components, one modifying the potential (divergence free) component of the force, \( f(x) \), and the other enhancing the constant/flux contribution to the force. We conclude that the optimal flux control cannot be explained simply as adjusting only the constant contribution leaving the potential intact.

In this letter we focused on the physics analysis and interpretations of the density/current variational formulation of the SOC. A reader interested in related mathematically rigorous results is advised to consult with [14]. The research of VYC has received support from the NSF under grant agreement no. CHE-1111350. The work at LANL was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy under Contract No. DE-AC52-06NA25396. The research of JB and HK was funded by the FP7/2007-2013 program under the grant no. 231495.

---

**FIG. 1:** Three illustrative examples of the 1d (particle on the circle) ergodic control with fixed flux (zero in the cases (a,b) and nonzero in the case (c)). The color coding of the curves is as follows: bright green and dark green curves show 10 + \( \rho(x) \) in the bare (without control) and optimal control cases respectively; blue and purple curves show \( f(x) \) and \( f(x) + u(x) \) respectively.

(a) \( f(x) = -2 \cos(x), \kappa = 0.5, J = 0.2 \) and \( \lim_{T \to \infty} C/T \approx 2.32 \)

(b) \( f(x) = -2 \cos(x), \kappa = 2.5, J = 0.2 \) and resulting in \( \lim_{T \to \infty} C/T \approx 1.81 \)

(c) \( f(x) = 1 - 2 \cos(x), \kappa = 0.5, J = 0.09 \) and resulting in \( \lim_{T \to \infty} C/T \approx 0.36 \)