Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

LINEAR PDES AND EIGENVALUE PROBLEMS CORRESPONDING TO ERGODIC STOCHASTIC OPTIMIZATION PROBLEMS ON COMPACT MANIFOLDS

JORIS BIERKENS, VLADIMIR Y. CHERNYAK, MICHAEL CHERTKOV, AND HILBERT J. KAPPEN

ABSTRACT. Long term average or 'ergodic' optimal control problems on a compact manifold are considered. The problems exhibit a special structure which is typical of control problems related to large deviations theory: Control is exerted in all directions and the control costs are proportional to the square of the norm of the control field with respect to the metric induced by the noise. The long term stochastic dynamics on the manifold will be completely characterized by the long term density ρ and the long term current density J. As such, control problems may be reformulated as variational problems over ρ and J. The density ρ is paired in the cost functional with a state dependent cost function V, and the current density J is paired with a vector potential or gauge field A. We discuss several optimization problems: the problem in which both ρ and J are varied freely, the problem in which ρ is fixed and the one in which J is fixed. These problems lead to different kinds of operator problems: linear PDEs in the first two cases and a nonlinear PDE in the latter case. These results are obtained through a variational principle using infinite dimensional Lagrange multipliers. In the case where the initial dynamics are reversible the optimally controlled diffusion is also reversible. The particular case of constraining the dynamics to be reversible of the optimally controlled process leads to a linear eigenvalue problem for the square root of the density process.

Key words and phrases: Stochastic optimal control, ergodic theory, calculus of variations, differential geometry, flux, current, gauge invariance

AMS Subject Classification (2010): Primary 49K20; Secondary 93E20, 58A25

1. INTRODUCTION

In this paper we discuss stochastic, long term average optimal, or 'ergodic' control problems on compact orientable manifolds, in which control is exerted in all directions, and where the control costs are proportional to the square of the norm of the control field with respect to the metric induced by the noise. As such, our emphasis is not on the solution of applied control problems. However this setting has strong connections with the general theory of large deviations of ergodic Markov processes [DV75b, Var84]. We place special emphasis on the characterization of control solutions in terms of density and current, in relation to the so called Level 2.5 large deviations theory [CCMT09, BC15, BFG15], discussed in more detail below.

The general theory of ergodic control in continuous spaces has been developed rigorously relatively recently; see works by Borkar and Gosh (e.g. [BG88]) and the recent monograph [ABG12]. The special case of compact manifolds has been extensively studied in relation to the theory of large deviations. A brief historic overview of this field will be provided in this introduction (Section 1.1).

The 'squared control cost' case is further motivated by recent attention to stochastic optimal control for finite time horizon problems with relative entropy determining control cost [Kap05]. Typically the solution of stochastic optimal control problems can be rephrased as the solution of a

non-linear partial differential equation called the Hamilton-Jacobi-Bellman equation [FR75, FS09]. If the control cost can be interpreted as a relative entropy (in continuous settings, problems with squared control cost) this often yields elegant simplifications of the non-linear Hamilton-Jacobi-Bellman (HJB) equation through an exponential transform, known as the Cole-Hopf transform in fluid dynamics [Hop50]. In the finite time horizon case the HJB equation is transformed into a linear equation, see e.g. [Kap05, BK14]. In the ergodic setting it leads typically to operator eigenvalue problems.

The reader interested in the statistical physics interpretation of this material is referred to the related publication by the same authors [CCBK14] which provides a brief overview in physical terms of some of the main results in this paper, including the expression of ergodic behaviour in terms of current and density. The current paper can be seen as a more detailed and precise mathematical exposition of these results, and at some appropriate places we will point to related discussions in [CCBK14]. Perhaps the main achievement of this paper, in comparison with [CCBK14], is the precise mathematical characterization of the necessary and sufficiency conditions for optimality for the optimal control problem formulated over density and current (Section 4). Also we have a more extensive description of the related optimization problems in which one of the variables (current or density) is held fixed (Sections 5, 6).

1.1. **Historical overview of connections between stochastic control theory and the theory of large deviations.**

Let us first remark that we will take some care to rephrase the historical results using the same notation as the one which is used in subsequent sections of this paper. As a basic introductions to the theory large deviations we recommend [Var84, dH00].

In [Kac51] the classical expression

$$(1) \qquad \lambda^\star = -\lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}\left[\exp\left(\int_0^t V(X_s)\, ds\right)\right]$$

(with expectations over Brownian sample paths) was obtained for the principal eigenvalue $\lambda^\star$ of the eigenvalue problem

$$(2) \qquad \tfrac{1}{2}\psi'' - V\psi = \lambda\psi,$$

where $V : \mathbb{R} \to [0, \infty)$.

As is well known, the representation (1) has a direct connection to the theory of large deviations. Specifically, for real random variables Y_n a *large deviation principle (LDP)* is said to hold with *rate function I* if

$$(3) \qquad \mathbb{P}\left(Y_n \in A\right) \approx \exp\left(-n \inf_{y \in A} I(y)\right),$$

in which case the *rate function I* satisfies, by Varadhan's Lemma (or in physics nomenclature, the Laplace method),

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}\left[\exp\left(nF(Y_n)\right)\right] = \sup_{y \in \mathbb{R}}[F(y) - I(y)].$$

A converse result known as the Gärtner-Ellis theorem allows one, under certain conditions, to deduce an LDP from the Legendre transform of the large n limit of the exponential expectations $\frac{1}{n} \log \mathbb{E}\left[\exp\left(nF(Y_n)\right)\right]$. Bryc's formula [Bry90] is an extension to infinite dimensions of this result.

In our case of interest, taking $Y_n = \frac{1}{n} \int_0^n V(X_s)\, ds$ and $F(y) = y$, the connection between (1) and the theory of large deviations becomes clear. At a higher level of abstraction, one could let Y_n assume values in the space of probability measures, and let Y_n denote the empirical distribution of (X_s), i.e. $Y_n(A) = \frac{1}{n} \int_0^n \mathbb{1}_{\{X_s \in A\}}\, ds$. In this case the functional F would be $F(\mu) := \int V\, d\mu$. Such a large deviation principle on the level of empirical distributions is called a *Level 2 LDP*. At an even higher level of abstraction, the *Level 3 LDP* concerns the asymptotic behaviour of the empirical process, i.e. the empirical distributions of finite sequences $(Y_1, \ldots Y_n)$.

Recently there has been an emergence of interest in the so called *Level 2.5 LDP*, which concerns the empirical distribution of a Markov sequence of random variables along with the empirical distribution of its current or flow. See e.g. [CCMT09, CCBK14] for relations with non-equilibrium statistical mechanics, and [BC15, BFG15] for large deviations of currents in Markov processes. It

is the aim of this paper to contribute to the Level 2.5 large deviation theory of diffusion processes in connection with a) the theory of stochastic control and b) certain partial differential equations and eigenvalue problems.

We hope that the above discussion has clarified to the reader the strong connection between an expression of the form (1) and the theory of large deviations, and will now sketch the historical developments concerning the principal eigenvalue in relation to stochastic control and large deviations below.

The seminal papers [DV75a, DV75b] extended (1) to general Markov processes on compact metric spaces, in the following sense. Let L be the infinitesimal generator of a Feller-Markov process in a compact metric space E and let $V : E \to \mathbb{R}$ be a continuous mapping. In [DV75a] a variational characterization of the principal eigenvalue $\lambda^\star$ of the operator $L - V$ is established as

$$(4) \qquad \lambda^\star = - \inf_{\mu \in \mathcal{M}} \left[\int_E V \, d\mu + I(\mu) \right],$$

where μ is the space of probability measures on E and I is defined by

$$(5) \qquad I(\mu) = - \inf_{\substack{\psi \in D(L) \\ \psi > 0}} \int_E \left(\frac{L\psi}{\psi} \right) \, d\mu.$$

The function I is the rate function describing the large deviations of the empirical distribution from the invariant probability distribution. In [DV75b] the relation to exponential expectations is obtained,

$$(6) \qquad \lim_{t \to \infty} \frac{1}{t} \log \mathbb{E} \left[\exp \left(-tF(L_t) \right) \right] = - \inf_{\mu \in \mathcal{M}} [F(\mu) + I(\mu)],$$

with

$$L_t(A) = \frac{1}{t} \int_0^t \mathbb{1}_A(X_s) \, ds,$$

i.e. $L_t(A)$ denotes the proportion of time up to time t that a sample path spends in A, and $F : \mathcal{M} \to \mathbb{R}$. Taking $F(\mu) := \int_E V \, d\mu$ formally recovers (1) in the special case of a Brownian motion. An extension to non-compact spaces is given in [DV76] but it seems that the stated conditions are hard to check in practice; see also [Var84]. In [G77] an alternative characterization of the Donsker-Varadhan rate function for non-degenerate diffusions on compact manifolds was obtained independently, given by

$$(7) \qquad I(\mu) = \tfrac{1}{2} \int_E \|\nabla \Phi\|^2 \, d\mu,$$

for μ having density ρ with respect to Riemann volume, and where Φ is the unique (up to a constant) solution of

$$\Delta \Phi + \frac{1}{\rho} \langle \nabla \rho, \nabla \Phi \rangle = \frac{1}{\rho} L^\star \rho,$$

with $L^\star$ the adjoint of L. In [Hol78] a similar representation for the principal eigenvalue of a diffusion generator is obtained, and in this paper the relation to stochastic control theory seems to be discussed for the first time. The connection with the control theoretic formulation is explored more extensively for the one-dimensional Brownian case in [Kar80]. It is noted that a change of variables $\psi = \exp(\Psi)$ in (2) yields

$$\lambda^\star = \tfrac{1}{2} \frac{d^2 \Psi}{dx^2} + \min_u \left(u \frac{d\Psi}{dx} + \tfrac{1}{2} u^2 \right) - V(x),$$

with the minimum being attained in $u^\star = -\frac{d\Psi}{dx}$. The above equation can be recognized as the Hamilton-Jacobi-Bellman equation [FR75, FS09] corresponding to a long term average cost problem with dynamics

$$(8) \qquad dX_t^u = u_t(X^u) \, dt + dW_t, \quad X_0 = x,$$

where W is a standard Brownian motion and where $u_t(X_t^u)$ is a control depending on $t \in [0, \infty)$ and the path of X_t^u up to time t. The cost function of the associated control problem is

$$\mathcal{C}(u, x) = \lim_{T \to \infty} \frac{1}{T} \mathbb{E}_x \left[\int_0^T V(X_t) + \tfrac{1}{2} u_t(X)^2 \ dt \right].$$

Also in [Kar80] it is discussed how the principal eigenvalue $\lambda^\star$ admits the representation

$$\lambda^\star = -\lim_{T \to \infty} \frac{\Psi(x; T)}{T},$$

where $\Psi(\cdot; T)$ is the value of the optimal control problem with finite time horizon T, with dynamics (8) and cost functional

$$\mathcal{C}(x, T; u) = \mathbb{E}_x^u \int_0^T \left\{ V(X_t^u) + \tfrac{1}{2} U_t^2(X^u) \right\} \ dt.$$

Subsequently Fleming, Sheu and Soner clarified the relation between stochastic control and the principal eigenvalue for general Markov operators [Fle82, She84], and provided an alternative proof of Donsker and Varadhan's result on the relation of the principal eigenvalue to the exponential expectation (6) in [FSS87]; see also the book chapter [Fle85].

The connections between the theory of stochastic optimal control and large deviation theory extend beyond the ergodic setting which is considered in this paper. For large deviations theory of Markov chains in continuous time see [dH00, For01]. A modern research monograph on the connection between stochastic optimal control and large deviation theory is provided by [FK06]. A different more commonly encountered stochastic control problem over an infinite time horizon is the control problem with discounted cost [FS09, Section III.9]. The ergodic control problem can be thought of as the limiting case where the discount factor approaches zero [ABG12]. For applied control theoretic papers see also e.g. [RBW08] for the diffusion case and [Tod06] for the Markov chain setting.

1.2. **Vector potentials and current density.** On a compact manifold, a few phenomena play a special role. The most important aspect of this setting is that transient behaviour cannot occur. Therefore, an invariant measure is necessarily unique and ergodicity follows immediately. The long term stochastic dynamics on the manifold will be completely characterized by the long term (particle) density ρ and the long term current density J (see Section 3). As such, control problems may be reformulated as variational problems over ρ and J. We will consider a cost functional where the density ρ is paired with the *scalar cost* or *(scalar) potential V*, and the current density J is paired with a *vector potential* or *gauge field A*.

It has been well established long ago in quantum field theory that the vector potential A plays the role of the variable conjugate to the current density J, just as the scalar potential V is related to the charge density ρ. Conservation of current, also known as the continuity equation, can then be viewed as a dual formulation of gauge invariance. This duality is not specific to the quantum world and can be applied to currents in a more general setting, e.g., for stochastic currents [CCMT09, BC15]. In the case of stochastic processes adding the current density variables to the more customary particle density variable is often referred to as the 2.5 level of theory [BC15], and has shown several advantages.

Besides the obvious benefit of an ability to efficiently treat observables depending on current, such as performed work, generated heat or entropy, as well as fluxes (see Section 3.1 and Appendix A), the combination of scalar density and current density turns out to constitute the 'right' set of variables in the sense that the large deviations of these quantities can be identified explicitly both in the continuous and discrete settings (such as stochastic dynamics on graphs). Even without the inclusion of a vector potential (i.e. taking $A = 0$), the formulation in terms of current and density provides a new and clear perspective on the optimal control problem in relation to the large deviations theory. The corresponding rate function is known in the physics literature as the current-density functional. Computation of the large deviations of the generalized observables can be reduced to solving an optimization problem, similar in spirit to the discussion in Section 1.1 for the scalar potential above.

In this manuscript we demonstrate that the gauge invariant approach, which has shown its capability in the context of stochastic processes (as briefly discussed above) can be extended to the stochastic optimal control setting while maintaining the same advantages of (i) considering optimal control of current density, in addition to classical particle density, which is achieved by introducing the gauge invariant extension of the standard Bellman equation, and (ii) formulating the gauge invariant Bellman equation as a solution of an optimization problem that involves a functional of current and particle densities. In other words, the approach outlined in this manuscript can be viewed as building the 2.5 level stochastic optimal control theory.

1.3. **Outline and aims of this paper.** In Section 2 the long term average stochastic optimal control problem for a non-degenerate diffusion over a compact manifold is formulated. In Section 3 some preliminary operations, mostly based upon the ergodic properties of a compact diffusion, are performed which allow us to remove all reference to probability theory from the problem formulation, resulting in a simplified optimization problem.

We then obtain conditions related to optimality for the formulated ergodic optimization problem. In Section 4 the problem in which both density ρ and current density J are varied freely is solved, in the sense that necessary as well as sufficient conditions for optimality are obtained. The sufficient conditions are obtained by considering the dual problem (Section 4.1) and the necessary conditions are derived using the method of Lagrange multipliers (Section 4.2). These conditions can be phrased as the solution of a linear eigenvalue problem. As a side result we derive a variational formula for the principal eigenvalue of an elliptic operator.

Then we consider two further, closely related optimization problems. In Section 5 the optimization problem in which J is varied for fixed density ρ is discussed. It turns out that a necessary condition for optimality can be phrased as a linear elliptic PDE. In Section 6 we obtain necessary conditions for optimality of the optimization problem for fixed current density J and varying ρ. Here the necessary condition for optimality can be formulated as a non-linear eigenvalue problem. In the reversible case this reduces to a linear eigenvalue problem.

In the case where the initial dynamics are reversible we obtain the result that the optimally controlled diffusion is also reversible (Section 4.3). The particular case of insisting $J = 0$ coincides with demanding reversible dynamics of the optimally controlled process. Interestingly, this optimization problem leads to a linear eigenvalue problem for the square root of the density process, just as we see in quantum mechanics (but note that our setting is entirely classical). We conclude this paper with a brief discussion (Section 7). In the appendices a detailed discussion of the use of a vector potential to quantify flux is provided (Appendix A), as well as a derivation of the relation between long term average drift, density, and current density (Appendix B).

1.4. **Notation.** When $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $H(\omega, x_1, x_2, \dots)$ is a measurable function of $\omega \in \Omega$ and other variables $x_1, x_2, \dots$, we often omit the dependence on ω, making $H(x_1, x_2, \dots)$ into a random quantity which also depends on $x_1, x_2, \dots$.

Throughout the paper we will work on a smooth orientable Riemannian manifold (M, g) and use similar notation as may be found e.g. in [IW89, Chapter V] or [War83]. By smooth we mean infinitely often differentiable, unless stated otherwise. As usual we employ the Einstein summation convention, i.e. in local coordinates summation over one upper and one lower index is automatic, e.g. $g^{ij} f_j = \sum_j g^{ij} f_j$. $C^\infty(M)$ denotes the space of smooth functions from M into $\mathbb{R}$, $C^k(M)$ the space of k times continuously differentiable functions (where $k = 0, 1, \dots$), $\mathfrak{X}(M)$ denotes the space of smooth vectorfields on M, and $\Lambda^p(M)$ denotes the space of smooth differential forms of order p on M, for $p = 0, 1, \dots, n$. The volume form is denoted dx, in local coordinates $dx = \sqrt{\det g_{ij}(x)} dx^1 \dots dx^n$. The Riemannian metric induces a local inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$ on tensors of arbitrary covariant and contravariant orders. E.g. if $S^{i_1, \dots, i_p}_{j_1, \dots, j_q}$ and $T^{i_1, \dots, i_p}_{j_1, \dots, j_q}$ are (p, q)-tensors, we have

$$\langle S, T \rangle = g_{i_1 i_1'} \dots g_{i_p i_p'} g^{j_1 j_1'} \dots g^{j_q j_q'} S^{i_1, \dots, i_p}_{j_1, \dots, j_q} T^{i_1', \dots, i_p'}_{j_1', \dots, j_q'}$$

and $\|T\| = \sqrt{\langle T, T \rangle}$. We employ the usual notions of exterior derivative $d : \Lambda^{p-1}(M) \to \Lambda^p(M)$, which as $L^2(M, g)$ adjoint $\delta : \Lambda^p(M) \to \Lambda^{p-1}(M)$, i.e. for differential forms $\alpha \in \Lambda^p$ and $\beta \in \Lambda^{p-1}$,

we have

$$\int_M \langle \alpha, d\beta \rangle \ dx = \int_M \langle \delta\alpha, \beta \rangle \ dx.$$

Partial derivatives in local coordinates are denoted by $\partial_i = \frac{\partial}{\partial x^i}$. If $\xi \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$ then in local coordinates $(\xi f)(x) = \xi^i \partial_i f(x)$. The symbol ∇ is the covariant derivative corresponding to the Levi-Civita connection of the Riemannian metric g. The Christoffel symbols corresponding to the Levi-Civita connection are denoted by Γ^i_{jk}. The gradient of a function $f \in C^\infty(M)$ is the vector field $\operatorname{grad} f = \nabla f$ with components $g^{ij} \partial_j f$. The divergence of a vectorfield $\xi \in \mathfrak{X}(M)$ is, in local coordinates, $\operatorname{div} \xi = \frac{1}{\sqrt{\det G}} \partial_i(\xi^i \sqrt{\det G})$, where $G = (g_{kl})$, and satisfies $\operatorname{div} \xi = -\delta\alpha$, where $\alpha \in \Lambda^1(M)$ is given by $\alpha_i = g_{ij} \xi^j$. On $\Lambda^p(M)$ an inner product is defined by $\langle \alpha, \beta \rangle_{\Lambda^p(M)} = \int_M \langle \alpha, \beta \rangle \ dx$, where dx denotes the volume form corresponding to g. The inner product $\langle \cdot, \cdot \rangle_{\Lambda^0(M)}$ is also denoted by $\langle \cdot, \cdot \rangle_{L^2(M)}$. Let $L^2(M, g) = L^2(M)$ denote the usual Hilbert space obtained by completing $C^\infty(M)$ with respect to the $L^2(M)$-inner product and considering, where appropriate, equivalence classes of functions. The Hodge star operator is denoted by $\star : \Lambda^p(M) \to \Lambda^{n-p}(M)$, $p = 0, \dots, n$.

2. Problem setting

Throughout this paper let (M, g) denote a smooth compact connected oriented n-dimensional Riemannian manifold. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ denote a filtered probability space on which is defined a d-dimensional standard Brownian motion. Consider a stochastic process X defined on M by the SDE, given in local coordinates by

$$(9) \qquad dX_t = b(X_t) + \sum_{\alpha=1}^d \sigma_\alpha \circ dB_t^\alpha, \quad t \geq 0,$$

or, in local coordinates,

$$dX_t^i = b^i(X_t) \ dt + \sum_{\alpha=1}^d \sigma_\alpha^i(X_t) \circ dB_t^\alpha, \quad t \geq 0,$$

where, for $\alpha = 1, \dots, d$, $\sigma_\alpha \in \mathfrak{X}(M)$, and it is assumed without loss of generality that the noise vectorfields σ_α are related to the Riemannian metric through the relation

$$g^{ij} = \sum_{\alpha=1}^d \sigma_\alpha^i \sigma_\alpha^j, \quad i, j = 1, \dots, n.$$

The notation $\circ dB^\alpha$ indicates that we take Stratonovich integrals with respect to the Brownian motion. One can think of b as a force field, resulting from a potential, some external influence, or a combination of both.

The SDE (9) is referred to as the *uncontrolled dynamics*. These dynamics may be altered by exterting 'control' vectorfield $u \in \mathfrak{X}(M)$ in the following way,

$$(10) \qquad dX_t = [b(X_t) + u(X_t)] \ dt + \sum_{\alpha=1}^d \sigma_\alpha(X_t) \circ dB_t^\alpha, \quad t \geq 0.$$

For any initial condition $X_0 = x_0 \in M$ and control vectorfield $u \in \mathfrak{X}(M)$, a unique solution to (10) exists [IW89, Chapter V] and will be denoted by $X^{x_0, u}$. The SDE (10) is referred to as the *controlled dynamics*.

Consider the random functional $\mathcal{C} : \Omega \times M \times \mathfrak{X}(M) \to \mathbb{R}$ denoting pathwise long term average cost,

$$(11) \qquad \mathcal{C}(\omega, x_0, u) := \limsup_{T \to \infty} \frac{1}{T} \left[\int_0^T V(X_s^{x_0, u}) + \frac{1}{2} \|u(X_s^{x_0, u})\|^2 \ ds + \int_0^T \langle A(X_s^{x_0, u}), \circ dX_s^{x_0, u} \rangle \right]$$

where $V \in C^\infty(M)$ is a *potential* or *state dependent cost function*, $\|u(\cdot)\|^2$ represents the *(instantaneous) control cost* corresponding to a control vectorfield $u \in \mathfrak{X}(M)$, and $A \in \mathfrak{X}(M)$. The final

term in (10) is of course shorthand notation for $\int_0^T [g_{ij}A^i](X_s^{x_0,u}) \circ d(X_s^{x_0,u})^j$ and may represent a *flux*, as explained in Section 3.1 and more extensively in Appendix A. The vectorfield A is often called a *vector potential* or *gauge field* in physics.

Remark 2.1. From a physics perspective, it would be better to let the vector potential A take the form of a differential form. On a mathematical level this distinction is irrelevant and we choose A to be a vectorfield for notational convenience, unless stated otherwise.

Remark 2.2. The notation 'lim sup' in (11) is used to avoid any discussion at this point about the existence of the limit. Instead of the *pathwise formulation* in (11), we could alternatively consider the weaker *average formulation*, in which case the cost function would be the long term average of the expectation value $\mathbb{E}^{x_0,u}$ of the integrand in (11). We will see in Section 3 that the limit of (11) exists (and not just the 'lim sup'). Furthermore this limit will turn out to be equal to a deterministic quantity (after excluding a a set of measure zero with respect to the invariant distribution), so that the pathwise formulation and the average formulation may be considered equivalent.

It is the main aim of this paper to consider the following problem.

Problem 2.3. For every $x_0 \in M$, find a control vectorfield $\hat{u} \in \mathfrak{X}(M)$ such that

$$\mathcal{C}(x_0, \hat{u}) = \inf_{u \in \mathfrak{X}(M)} \mathcal{C}(x_0, u), \quad \text{almost surely.}$$

In Sections 5 and 6 we will also discuss other variants of the control problem, where we will respectively fix the invariant density and the current density, which will be defined in Section 3.

3. Ergodic reformulation of the optimization problem

In this section we will derive two equivalent formulations of Problem 2.3. These reformulations, Problem 3.6 and Problem 3.12 below, are better suited to the analysis in the remaining sections. Also some notation will be established that will be used throughout this paper.

Let $\Omega^X = C([0,\infty); M)$ denote the space of sample paths of solutions to (10). We equip Ω^X with the σ-algebra $\mathcal{F}^X$ and filtration $(\mathcal{F}_t^X)_{t \geq 0}$ generated by the cylinder sets of X. Furthermore let probability measures $\mathbb{P}^{x_0,u}$ on Ω^X be defined as the law of $X^{x_0,u}$, for all $x_0 \in M$ and $u \in \mathfrak{X}(M)$. For every $u \in \mathfrak{X}(M)$ the collection of probability measures $\mathbb{P}^{\cdot,u}$ defines a Markov process on Ω^X, i.e. for every $x_0, x_1 \in M$,

$$\mathbb{P}^{x_0,u}\left((X(t_1+s), \ldots, X(t_k+s)) \in F\right) \mid \mathcal{F}_s^X\right) = \mathbb{P}^{X(s),u}\left((X(t_1), \ldots, X(t_k)) \in F\right).$$

For the moment let $u \in \mathfrak{X}(M)$ be fixed. By [IW89, Theorem V.1.2], the Markov generator corresponding to (10) is given by

$$L_u f(x) = \tfrac{1}{2} \sum_{\alpha=1}^{d} \sigma_\alpha \sigma_\alpha f(x) + (b+u)f(x).$$

Lemma 3.1. L_u *may be written as*

$$L_u f = \tfrac{1}{2}\Delta f + (\widetilde{b} + u)f, \quad f \in C^2(M),$$

where $\widetilde{b} := b + \sum_{\alpha=1}^{d} \nabla_{\sigma_\alpha} \sigma_\alpha$. *The adjoint of* L_u *with respect to the* $L^2(M)$ *inner product is given by*

$$L_u^\star \rho = \tfrac{1}{2}\Delta \rho - \operatorname{div}\left(\rho(\widetilde{b} + u)\right), \quad \rho \in C^2(M).$$

Proof. The Laplace-Beltrami operator may be expressed as (see [IW89, p. 285, eqn. (4.32)])

$$(12) \qquad \Delta f = g^{ij}\partial_i\partial_j f - g^{ij}\Gamma_{ij}^k \partial_k f.$$

Using this expression, we compute

$$
\begin{aligned}
\sum_{\alpha=1}^{d} \sigma_\alpha \sigma_\alpha f &= \sum_{\alpha=1}^{d} \sigma_\alpha^i \partial_i \left(\sigma_\alpha^j \partial_j f \right) = \sum_{\alpha=1}^{d} \left(\sigma_\alpha^i \sigma_\alpha^j \partial_i \partial_j f + \sigma_\alpha^i \left(\partial_i \sigma_\alpha^j \right) (\partial_j f) \right) \\
&= \sum_{\alpha=1}^{d} \left(g^{ij} \partial_i \partial_j f + \sigma_\alpha^i \left(\partial_i \sigma_\alpha^j \right) (\partial_j f) \right) = \Delta f + g^{ij} \Gamma_{ij}^k \partial_k f + \sum_{\alpha=1}^{d} \sigma_\alpha^i \left(\partial_i \sigma_\alpha^j \right) \partial_j f \\
&= \Delta f + \sum_{\alpha=1}^{d} \left(\nabla_{\sigma_\alpha} \sigma_\alpha \right) f,
\end{aligned}
$$

where the last equality is a result of the definition of the Levi-Civita connection and the corresponding Christoffel symbols. The expression for $L_u^\star$ is immediate from its definition. $\qquad\square$

In the remainder of this work, we will assume that all advection terms are absorbed in the drift b so that we may omit the tilde in $\widetilde{b}$. This can alternatively be interpreted as assuming $\sum_{\alpha=1}^{d} \nabla_{\sigma_\alpha} \sigma_\alpha = 0$. This is further equivalent to demanding that $L_u = \frac{1}{2}\Delta + b + u$. This assumption is without loss of generality (on the level of the probability law on trajectories of X) by the above lemma and the fact that the Markov generator L^u uniquely determines the law of the trajectories of X.

Lemma 3.2. *Let $x_0 \in M$ and $u \in \mathfrak{X}(M)$. The expectation of the trajectory of X over the vector potential may be expressed as*

$$
\mathbb{E}^{x_0,u} \int_0^T \langle A(X_t), \circ dX_t \rangle = \mathbb{E}^{x_0,u} \int_0^T \left[\langle A, (b+u) \rangle + \tfrac{1}{2} \operatorname{div} A \right] (X_t) \; dt.
$$

Proof. Using the usual transformation rule between Itô and Stratonovich integrals [IW89, Equation (1.4), p. 250], we may write

$$
dX_t^i = \overline{b}_u^i(X_t) \; dt + \sum_{\alpha=1}^{d} \sigma_\alpha^i(X_t) dB^\alpha(t), \tag{13}
$$

where $\overline{b}_u(x)$ is given by

$$
\overline{b}_u^i(x) := b^i(x) + u^i(x) + \tfrac{1}{2} \sum_{\alpha=1}^{d} \left(\partial_k \sigma_\alpha^i(x) \right) \sigma_\alpha^k(x).
$$

By the definition of the Stratonovich integral, $Z \circ dY = Z \; dY + \frac{1}{2} d[Z, Y]$ for semimartingales Y and Z [IW89, Equation (1.10), p.100], with $Z \; dY$ denoting the Itô integral. Therefore

$$
\begin{aligned}
&[g_{ij}A^i](X_t) \circ dX_t^j \\
&= [g_{ij}A^i](X_t) \; dX_t^j + \tfrac{1}{2} d[(g_{ij}A^i)(X_t), X_t^j] \\
&= [g_{ij}A^i(X_t)] \; dX_t^j + \tfrac{1}{2} \partial_k(g_{ij}A^i)(X_t) d[X_t^k, X_t^j] \\
&= [g_{ij}A^i(X_t)] \left[\overline{b}_u^j(X_t) \; dt + \sum_{\alpha=1}^{d} \sigma_\alpha^i(X_t) \; dB^\alpha(t) \right] + \tfrac{1}{2} \partial_k(g_{ij}A^i)(X_t) \sum_{\alpha=1}^{d} \sigma_\alpha^k(X_t)\sigma_\alpha^j(X_t) \; dt.
\end{aligned}
$$

Integrating over t and taking expectations gives

$$
\begin{aligned}
&\mathbb{E}^{x_0,u} \int_0^T A(X_s) \circ dX_s \\
&= \mathbb{E}^{x_0,u} \int_0^T \left\{ (g_{ij}A^i)(X_t) \left[b_u^j(X_t) + \tfrac{1}{2} \left(\partial_k \sigma_\alpha^j(X_t) \right) \sigma_\alpha^k(X_t) \right] + \tfrac{1}{2} \partial_k (g_{ij}A^i)(X_t) \sum_{\alpha=1}^{d} \sigma_\alpha^k(X_t)\sigma_\alpha^j(X_t) \right\} \; dt \\
&= \mathbb{E}^{x_0,u} \int_0^T \left\{ g_{ij}A^i(b^j + u^j)(X_t) + \tfrac{1}{2} \nabla_i A^i(X_t) \right\} \; dt.
\end{aligned}
$$

In the last expression we recognize the divergence of the vectorfield A^i, resulting in the stated expression. $\qquad\square$

Let $\mathcal{B}(M)$ denote the Borel σ-algebra on M. Let $u \in \mathfrak{X}(M)$. A probability measure $\mu_u \, dx$ on M is called an *invariant probability distribution* for (10), if

$$\int_M \mathbb{P}^{x,u}(X_t \in B)\mu_u \, dx(dx) = \mu_u(B), \quad \text{for all } t \geq 0 \text{ and } B \in \mathcal{B}(M).$$

The following result on invariant measures for non-degenerate diffusions [IW89, Proposition V.4.5] is essential for our purposes.

Proposition 3.3 (Existence and uniqueness of invariant probability measure). *For any $u \in \mathfrak{X}(M)$ there exists a unique invariant probability measure $\rho_u \, dx$ on M for (10) which is absolutely continuous with respect to the Riemannian volume measure dx. The density $\rho_u \in C^\infty(M)$ is a solution of*

$$(14) \qquad\qquad\qquad\qquad L_u^\star \rho = 0.$$

Furthermore $\rho_u > 0$ on M.

We will refer to (14) as the *Fokker-Planck equation*, in agreement with common physics terminology. In the remainder of this work let ρ_u as defined by Proposition 3.3.

In the physics literature, the *empirical density* and *empirical current density* are defined respectively as (see [CCMT09]):

$$\rho_t(x,\omega) = \frac{1}{t}\int_0^t \delta(x - X_s(\omega)) \, ds, \quad J_t(x,\omega) = \frac{1}{t}\int_0^t \dot{X}_s\delta(x - X_s(\omega)) \, ds.$$

Here (and only here) δ denotes the Dirac delta function. These fields, which have a clear intuitive meaning, will be very relevant in the remainder of this work and we will make these precise from a mathematical point of view.

Let $B_b(M)$ denote the set of bounded Borel-measurable functions on M. We will work with the set of empirical average measures $(\nu_t(dx,\omega))_{t>0}$ on $\mathcal{B}(M) \times \Omega^X$, defined by

$$(15) \qquad\qquad \nu_t(B) := \frac{1}{t}\int_0^t \mathbb{1}_B(X_s) \, ds, \quad t > 0, \ B \in \mathcal{B}(M),$$

where $\mathbb{1}_B$ denotes the indicator function of the set B. Our primary interest is in the infinite time horizon limit.

Proposition 3.4. *Let $u \in \mathfrak{X}(M)$. For all $\varphi \in L^2(M, \rho_u \, dx)$ and ρ_u-almost all $x_0 \in M$,*

$$(16) \qquad\qquad \lim_{t\to\infty} \int_M \varphi \, d\nu_t = \int_M \varphi \, \rho_u \, dx, \quad \mathbb{P}^{x_0,u}\text{-almost surely.}$$

Proof. For $u \in \mathfrak{X}(M)$, we define a stationary probability measure $\mathbb{P}^u$ on Ω^X by

$$\mathbb{P}^u(G) = \int_M \mathbb{P}^{x,u}(G) \, \rho_u(x) \, dx, \quad G \in \mathcal{F}^X.$$

For $\varphi \in L^2(M, \rho_u \, dx)$ we then have, by the ergodic theorem, see e.g. [DZ96, Theorem 3.3.1], that $\lim_{t\to\infty} \int_M \varphi \, d\nu_t = \int_M \varphi \, \rho_u \, dx$, $\mathbb{P}^u$-almost surely. Since $\rho_u > 0$ on M, this implies that

$$\lim_{t\to\infty} \int_M \varphi \, d\nu_t = \int_M \varphi \, \rho_u \, dx, \quad \mathbb{P}^{x_0,u}\text{-a.s. } \rho_u\text{-a.a. } x_0 \in M.$$

$\qquad\square$

Hence $\lim_{t\to\infty} \nu_t$ is μ_u-almost everywhere equal to a constant.

Corollary 3.5. *Let $u \in \mathfrak{X}(M)$. For ρ_u-almost every $x_0 \in M$,*

$$(17) \qquad \lim_{T\to\infty} \frac{1}{T}\int_0^T \langle A(X_t), \circ dX_t \rangle = \int_M \left[\langle A, (b+u) \rangle + \tfrac{1}{2}\operatorname{div} A\right] \rho_u \, dx, \quad \mathbb{P}^{x_0,u}\text{-a.s.}$$

Proof. This is an immediate corollary of Lemma 3.2, Proposition 3.3 and Proposition 3.4. $\qquad\square$

The above results provide sufficient motivation to rephrase Problem 2.3 as follows.

Problem 3.6. Minimize

$$(18) \qquad \mathcal{C}(\rho, u) := \int_M \left\{ V + \tfrac{1}{2}\|u\|^2 + \langle A, (b+u)\rangle + \tfrac{1}{2}\operatorname{div} A \right\} \rho \ dx.$$

with respect to $(\rho, u) \in C^\infty(M) \times \mathfrak{X}(M)$ subject to the constraints $L_u^\star \rho = 0$ and $\int_M \rho \ dx = 1$.

Lemma 3.7. *Suppose* $u, \widetilde{u}, b, \widetilde{b} \in \mathfrak{X}(M)$ *and* $V, \widetilde{V} \in C^\infty(M)$ *are related by*

$$(19) \qquad \widetilde{b} = b - A, \quad \widetilde{u} = u + A, \quad and \quad \widetilde{V} = V - \tfrac{1}{2}\|A\|^2 + \langle A, b\rangle + \tfrac{1}{2}\operatorname{div} A.$$

Let $\widetilde{L}_{\widetilde{u}}$ *denote the generator corresponding to drift* $\widetilde{b}$ *and control* $\widetilde{u}$. *Then* $L_u^\star \rho = 0$, $\int_M \rho \ dx = 1$, *and* (ρ, u) *are optimal for Problem 3.6, if and only if* $\widetilde{L}_{\widetilde{u}}^\star \rho = 0$ *and* $(\rho, \widetilde{u})$ *are optimal for Problem 3.6, with* $\mathcal{C}$ *replaced by*

$$\widetilde{\mathcal{C}}(\rho, \widetilde{u}) = \int_M \left\{ \widetilde{V} + \tfrac{1}{2}\|\widetilde{u}\|^2 \right\} \rho \ dx.$$

Proof. First note that $\widetilde{b} + \widetilde{u} = b + u$, so that $L_u^\star = \widetilde{L}_{\widetilde{u}}^\star$, and hence $\widetilde{L}_{\widetilde{u}}^\star \rho = 0$ if and only if $L_u^\star \rho = 0$. Furthermore $\widetilde{C}(\rho, \widetilde{u}) = C(\rho, u)$ by direct computation. $\qquad\square$

Remark 3.8. As a consequence of Lemma 3.7 the vector potential A may be completely removed from the problem by redefining b, V and u by (19). Using this observation simplifies the derivation of some of the results in subsequent sections.

Note that $\dot{X}_s$ is not defined, a.s., so our mathematical analogue of the empirical current density requires more care. In Appendix B, we derive the vector field $J \in \mathfrak{X}(M)$ denoting current density, as

$$(20) \qquad J = -\tfrac{1}{2}\nabla\rho + \rho\,(b+u),$$

for $\rho = \rho_u$.

By rearranging (20), we can express a control u in terms of J and ρ as

$$(21) \qquad u = -b + \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla\rho\right).$$

The following lemma is an immediate consequence of the equality $\operatorname{div} J = -L_u^\star \rho$.

Lemma 3.9. *Suppose* u, ρ *and* J *are related by* (20). *Then* $\operatorname{div} J = 0$ *if and only if* $L_u^\star \rho = 0$.

In other words, $\operatorname{div} J = 0$ if and only if $\rho \ dx$ is invariant for (10). Similar observations may be found throughout the physics literature, e.g. [Ris89, BC15]. Recall Lemma 3.2, where the expectation of the vector potential A over the trajectory was expressed as an expectation over a Lebesgue integral. For the long term average of the gauge field this leads to the following result.

Lemma 3.10. *Suppose* $u \in \mathfrak{X}(M)$ *and* $x_0 \in M$, *and* J *satisfy* (20). *Then*

$$(22) \qquad \lim_{T\to\infty} \frac{1}{T}\int_0^T \langle A(X_t), \circ dX_t\rangle = \int_M \langle A, J\rangle \ dx, \quad \mathbb{P}^{x_0, u}\text{-a.s.}$$

Proof. From Corollary 3.5 we have (17). By (21), and partial integration, this equals

$$\int_M \left(\left\langle A, \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla\rho\right)\right\rangle + \tfrac{1}{2}\operatorname{div} A \right) \rho \ dx = \int_M \langle A, J\rangle \ dx.$$

$\qquad\square$

Because of the above observations, instead of varying ρ and u in the optimization problem 3.6, we may as well vary $\rho \in C^\infty$ and $J \in \mathfrak{X}(M)$, while enforcing $\operatorname{div} J = 0$ (equivalent to the Fokker-Planck equation for ρ by Lemma 3.9) and $\int_M \rho \ dx = 1$. The uniqueness of the solution to the Fokker-Planck equation ensures that ρ is positive. The control u is then determined uniquely by

(21). Combining (21) and (22), we may alternatively express the cost functional (18) as a function of ρ and J, namely

$$(23) \qquad \mathcal{C}(\rho, J) = \int_M \left[\left(V + \frac{1}{2} \left\| \frac{1}{\rho} \left(J + \tfrac{1}{2}\nabla\rho \right) - b \right\|^2 \right) \rho + \langle A, J \rangle \right] \, dx.$$

Remark 3.11. Strictly speaking the use of $\mathcal{C}$ for different cost functionals is an abuse of notation but we trust this will not lead to confusion.

Problem (2.3) can thus be rephrased as the following problem:

Problem 3.12. Minimize $\mathcal{C}(\rho, J)$ with respect to $\rho \in C^\infty(M)$ and $J \in \mathfrak{X}(M)$, subject to the constraints $\operatorname{div} J = 0$ and $\int_M \rho \, dx = 1$, where $\mathcal{C}(\rho, J)$ is given by (23).

Although we will not directly make use of the following fact, it seems relevant enough to mention here.

Proposition 3.13. $\mathcal{C}(\rho, J)$ *is convex.*

Proof. First consider the integrand pointwise, in the form of the map $h : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by

$$h(x, y, z) = \left(a + \frac{1}{2} \left\| c + \frac{y + \tfrac{1}{2}z}{x} \right\|^2 \right) x + \langle m, y \rangle = ax + \frac{1}{2}\|c\|^2 x + \langle c, y + \tfrac{1}{2}z \rangle + \frac{\|y + \tfrac{1}{2}z\|^2}{2x} + \langle m, y \rangle.$$

This is a summation of convex functions in (x, y, z), so it follows that h is convex. Note that $\mathcal{C}$ is obtained by integrating h over M, taking $x = \rho$, $y = J$ and $z = \nabla\rho$. $\qquad\square$

3.1. **Interpretation of the vector potential as flux.** In view of Remark 2.1, for the discussion in this section let $A \in \Lambda^1(M)$ be a vector potential in differential form. A particular use or interpretation of A is that $A(J)$ may quantify flux of J through a submanifold of M. In particular, for a given $(n-1)$-cycle α (roughly speaking, an $(n-1)$-dimensional submanifold of M without boundary), there exists a unique harmonic $A \in \Lambda^1(M)$ (i.e. $\Delta A = 0$, where Δ denotes the Laplace-Beltrami operator), which depends only on the singular homology class $[\alpha] \in \mathcal{H}_{m-1}(M; \mathbb{R})$ of α, such that $\int_\alpha \star J = \int_M A(J) \, dx$ for all $J \in \mathfrak{X}(M)$ satisfying $\operatorname{div} J = 0$.

Example 3.14 (S^1). The divergence free 1-forms J on S^1 are constant, say $J = J_0 \, d\theta$ for some $J_0 \in \mathbb{R}$. A 0-cycle α of S^1 consists of a collection of points $\theta_1, \ldots, \theta_k \subset [0, 2\pi)$ with multiplicities $\alpha_1, \ldots, \alpha_k$. The flux of J through α is then simply given by $\sum_{i=1}^k \alpha_i J(\theta_i) = \sum_{i=1}^k \alpha_i J_0$. By defining a differential form $A = A_0 \, d\theta$, with constant component $A_0 := \sum_i \alpha_i \frac{1}{2\pi}$, we find that

$$\int_{S^1} \langle A, J \rangle \, d\theta = \int_{S^1} A_0 J_0 \, d\theta = \sum_i \alpha_i J_0 = \int_\alpha \star J.$$

We see that this choice of A is the constant (and therefore harmonic) representative in $\mathcal{H}^1_{\text{deRham}}(S^1)$ corresponding to $[\alpha] \in \mathcal{H}_0(S^1; \mathbb{R})$.

A more extensive discussion of this topic may be found in Appendix A.

4. Unconstrained optimization – the HJB equation

In this section we will find necessary conditions for a solution of Problem 3.6 or equivalently Problem 3.12. In fact, for technical reasons we will work with the the formulation in terms of ρ and J, i.e. Problem 3.12. The main reason for this is the convenient form (in particular, the linearity) of the constraint $\operatorname{div} J = 0$. This may be compared to the equivalent constraint $\frac{1}{2}\Delta\rho - \operatorname{div}(\rho(b+u)) = 0$, which is non-linear as a function in (ρ, u).

The approach to Problem 3.6 or Problem 3.12 is to use the method of Lagrange multipliers to enforce the constraints. Since the constraint $\operatorname{div} J(x) = 0$ needs to be enforced for all $x \in M$, the corresponding Lagrange multiplier is an element of a function space. A purely formal derivation of the necessary conditions using Lagrange multipliers is straightforward, but we wanted to be precise in proving necessary and sufficient conditions for optimality.

4.1. **Sufficient condition for optimality.** Define $\mathcal{U} = \{(\rho, J) \in C^\infty \times \mathfrak{X}(M) : \rho(x) \geq 0 \text{ for all } x \in M\}$. Associated to Problem 3.12 we may define a Lagrangian

$$(24) \qquad \mathcal{L}(\rho, J, \Psi, \lambda) := \mathcal{C}(\rho, J) + \int_M \Psi(\operatorname{div} J) \, dx + \lambda \left(\int_M \rho \, dx - 1 \right),$$

defined for $(\rho, J) \in \mathcal{U}$, $\Psi \in C^\infty(M)$ and $\lambda \in \mathbb{R}$. The dual cost functional is as usual defined as

$$\mathcal{C}^\star(\Psi, \lambda) := \inf_{(\rho, J) \in \mathcal{U}} \mathcal{L}(\rho, J, \Psi, \lambda).$$

Write $\mathcal{V} := C^\infty(M) \times \mathbb{R}$ for the domain of the dual cost functional $\mathcal{C}^\star$. It is immediate from the definition that for every $(\rho, J) \in \mathcal{U}$ satisfying the constraints

$$(25) \qquad \operatorname{div} J = 0 \quad \text{and} \quad \int_M \rho \, dx = 1,$$

and $(\Psi, \lambda) \in \mathcal{V}$ we have

$$\mathcal{C}^\star(\Psi, \lambda) \leq \mathcal{L}(\rho, J, \Psi, \lambda) = \mathcal{C}(\rho, J).$$

The following lemma is therefore immediate.

Lemma 4.1. *Suppose $(\rho^\star, J^\star) \in \mathcal{U}$, $(\Psi^\star, \lambda^\star) \in \mathcal{V}$. If the constraints (25) are satisfied for $(\rho^\star, J^\star)$ and $\mathcal{C}^\star(\Psi^\star, \lambda^\star) = \mathcal{C}(\rho^\star, J^\star)$, then $(\rho^\star, J^\star)$ is optimal for Problem 3.12.*

Lemma 4.2. *Suppose $A = 0$. The dual cost functional $\mathcal{C}^\star$ can be expressed as*

$$(26) \qquad \mathcal{C}^\star(\Psi, \lambda) = \begin{cases} -\lambda & \text{if } I[\Psi, \lambda](x) \geq 0 \text{ for all } x \in M, \\ -\infty & \text{if } I[\Psi, \lambda](x) < 0 \text{ for some } x \in M, \end{cases}$$

where

$$(27) \qquad I[\Psi, \lambda] := V - \tfrac{1}{2}\|\nabla\Psi\|^2 - \langle b, \nabla\Psi \rangle - \tfrac{1}{2}\Delta\Psi + \lambda.$$

Proof. We have

$$\mathcal{L}(\rho, J, \Psi, \lambda) = \int_M \left[\left\{ V + \tfrac{1}{2} \left\| \tfrac{1}{\rho}(J + \tfrac{1}{2}\nabla\rho) - b \right\|^2 + \lambda \right\} \rho - \langle J, \nabla\Psi \rangle \right] dx - \lambda.$$

For fixed ρ the choice $J^\star = -\tfrac{1}{2}\nabla\rho + \rho(b + \nabla\Psi)$ determines the pointwise minimum of the integrand. For this choice of J, we obtain

$$\mathcal{L}(\rho, J^\star, \Psi, \lambda) = \int_M \left[\left\{ V + \tfrac{1}{2}\|\nabla\Psi\|^2 + \lambda \right\} \rho + \langle (\tfrac{1}{2}\nabla\rho - \rho(b + \nabla\Psi)), \nabla\Psi \rangle \right] dx - \lambda$$

$$= \int_M \left\{ V - \tfrac{1}{2}\|\nabla\Psi\|^2 + \tfrac{1}{2}\langle \nabla(\log\rho), \nabla\Psi \rangle - \langle b, \nabla\Psi \rangle + \lambda \right\} \rho \, dx - \lambda$$

$$= \int_M \left[\left\{ V - \tfrac{1}{2}\|\nabla\Psi\|^2 - \langle b, \nabla\Psi \rangle + \lambda \right\} \rho + \tfrac{1}{2}\langle \nabla\rho, \nabla\Psi \rangle \right] dx - \lambda$$

$$= \int_M \left\{ V - \tfrac{1}{2}\|\nabla\Psi\|^2 - \langle b, \nabla\Psi \rangle - \tfrac{1}{2}\Delta\Psi + \lambda \right\} \rho \, dx - \lambda.$$

Now if $I(\Psi, \lambda) < 0$ on some open subset of M, then $\lambda(\rho, J^\star, \Psi, \lambda)$ can assume arbitrarily large negative values by letting ρ approach a Dirac delta peak centered within that subset, so that in this case $\mathcal{L}(\rho, J^\star, \Psi, \lambda) \to -\infty$. If $I(\Psi, \lambda) \geq 0$ on M, then $\mathcal{L}(\rho, J^\star, \Psi, \lambda)$ is minimized by taking $\rho^\star = 0$. $\qquad\square$

Lemma 4.3. *Suppose $A = 0$. Suppose $(\Psi, \lambda) \in \mathcal{V}$ satisfies $I[\Psi, \lambda] = 0$, i.e.*

$$\tfrac{1}{2}\Delta\Psi + \tfrac{1}{2}\|\nabla\Psi\|^2 + \langle b, \nabla\Psi \rangle - V = \lambda.$$

Let $u = \nabla\Psi$ and let $\rho > 0$ be the unique solution to the Fokker-Planck equation $L_u^\star \rho = 0$ with $\int_M \rho \, dx = 1$. Define $J = -\tfrac{1}{2}\nabla\rho + \rho(b + \nabla\Psi)$. Then (ρ, J) solves Problem 3.12, and $C(\rho, J) = -\lambda$.

Proof. By Lemma 3.9, $\operatorname{div} J = 0$, so that (ρ, J) satisfy the constraints (25). Furthermore

$$\mathcal{C}(\rho, J) = \int_M \left\{ V + \tfrac{1}{2} \left\| \tfrac{1}{\rho}(J + \tfrac{1}{2}\nabla\rho) - b \right\|^2 \right\} \rho \, dx = \int_M \left\{ 2V - \tfrac{1}{2}\Delta\Psi - \langle b, \nabla\Psi \rangle \right\} \rho \, dx + \lambda$$

$$= \int_M \left[2V\rho + (\operatorname{div}(b\rho) - \tfrac{1}{2}\Delta\rho)\Psi \right] \, dx + \lambda$$

$$= \int_M \left[2V\rho - (L_u^\star \rho)\Psi - \operatorname{div}(\rho u)\Psi \right] \, dx + \lambda = \int_M \left\{ 2V + \|\nabla\Psi\|^2 \right\} \rho \, dx + \lambda = 2\mathcal{C}(\rho, J) + \lambda.$$

Consequently, $\mathcal{C}(\rho, J) = -\lambda$. Also, for this choice of Ψ and λ we have by Lemma 4.2 that $\mathcal{C}^\star(\Psi, \lambda) = -\lambda$. The optimality of (ρ, J) now follows from Lemma 4.1. $\square$

The following proposition is now a direct consequence of Remark 3.8, Lemma 4.3 and a brief computation.

Proposition 4.4 (Gauge invariant Hamilton-Jacobi-Bellman equation – sufficiency)**.** *Suppose* $(\Psi, \lambda) \in \mathcal{V}$ *satisfies*

$$(28) \qquad \tfrac{1}{2}\operatorname{div}(\nabla\Psi - A) + \tfrac{1}{2}\|\nabla\Psi - A\|^2 + \langle b, (\nabla\Psi - A) \rangle - V = \lambda.$$

Let $u = \nabla\Psi - A$, *let* ρ *denote the unique solution to* $L_u^\star \rho = 0$ *with* $\int_M \rho \, dx = 1$, *and let* $J = -\tfrac{1}{2}\nabla\rho + \rho(b + \nabla\Psi - u)$. *Then* (ρ, J) *is optimal for Problem 3.12 and* $\mathcal{C}(\rho, J) = -\lambda$.

Equation (28), the *gauge-invariant Hamilton-Jacobi-Bellman equation* was already discussed in our related physics paper, see [CCBK14, equation (10)]. As remarked there, the equation may be linearised by an exponential transformation. The verification of this result is straightforward.

Corollary 4.5 (Linear gauge invariant Hamilton-Jacobi-Bellman equation – sufficiency)**.** *Suppose* $(\psi, \lambda) \in C^\infty(M) \times \mathbb{R}$, *satisfy* $\psi > 0$ *and*

$$(29) \qquad \tfrac{1}{2}\operatorname{div}(\nabla\psi - A\psi) + \langle (b - A), (\nabla\psi - A\psi) \rangle - (V + \tfrac{1}{2}\|A\|^2)\psi = \lambda\psi.$$

Then (28) *is satisfied for* $\Psi = \log\psi$, *so that the conclusion of Proposition 4.4 applies. In terms of* ψ *the optimal control and associated current are given by*

$$(30) \qquad u = \nabla\log\psi - A \quad \text{and} \quad J = -\tfrac{1}{2}\nabla\rho + \rho(b + \nabla\log\psi - A).$$

Essentially (29) is an eigenvalue problem for a non-degenerate second order elliptic differential equation on a compact manifold, for which it is well established that a solution (ψ, λ) exists such that ψ is smooth and for which $\psi > 0$; see [War83].

4.2. **Necessary condition for optimality.** In order to obtain necessary conditions for optimality of (ρ, J) for Problem 3.12 we will relax the problem to an optimization problem over Sobolev spaces. In particular, we will rephrase it as the following abstract optimization problem. Let X and Z be Banach spaces and let $\mathcal{U}$ be an open set in X. Let $\mathcal{C} : \mathcal{U} \subset X \to \mathbb{R}$ and $\mathcal{H} : \mathcal{U} \subset X \to Z$.

Problem 4.6. Minimize $\mathcal{C}(x)$ over $\mathcal{U}$ subject to the constraint $\mathcal{H}(x) = 0$.

The *Fréchet derivative* [Lue69] of a mapping $T : D \subset X \to Y$ in $x \in D$ will be denoted by $T'(x) \in L(X; Y)$. We will need the following notion.

Definition 4.7 (Regular point)**.** Let T be a continuously Fréchet differentiable function from an open set D in a Banach space X into a Banach space Y. If $x_0 \in D$ is such that $T'(x_0)$ maps X onto Y, then the point x_0 is said to be a *regular point* of the transformation T.

For the abstract Problem 4.6 the following necessary condition holds for a local extremum [Lue69, Theorem 9.3.1].

Lemma 4.8 (Lagrange multiplier necessary conditions)**.** *Suppose* $\mathcal{C}$ *and* $\mathcal{H}$ *are continuously Fréchet differentiable on* $\mathcal{U}$. *If* $\mathcal{C}$ *has a local extremum under the constraint* $\mathcal{H}(x) = 0$ *at the regular point* $x_0 \in \mathcal{U}$, *then there exists an element* $z_0^\star \in Z^\star$ *such that*

$$\mathcal{C}'(x_0) + \langle \mathcal{H}'(x_0), z_0^\star \rangle = 0.$$

Here $\langle \cdot, \cdot \rangle$ *denotes the pairing between* Z *and* $Z^\star$.

We will define Sobolev spaces of functions and vector fields as follows. For $k \in \mathbb{N} \cup \{0\}$ let $H^k(M)$ be the completion of $C^\infty(M)$ with respect to the norm

$$\|\varphi\|^2_{H^k(M)} := \sum_{l=0}^{k} \int_M \|\nabla^l \varphi\|^2 \, dx,$$

equipped with Hilbert space structure induced by $\|\cdot\|_{H^k(M)}$. Similarly, let $H^k_{\mathfrak{X}}(M)$ be the completion of $\mathfrak{X}(M)$ with respect to the norm

$$\|\xi\|^2_{H^k_{\mathfrak{X}}(M)} := \sum_{l=0}^{k} \int_M \|\nabla^l \xi\|^2 \, dx,$$

equipped similarly with Hilbert space structure. Note that the Laplace-Beltrami operator Δ is a bounded mapping $\Delta : H^k(M) \to H^{k-2}(M)$. Also div $: H^k_{\mathfrak{X}}(M) \to H^{k-1}(M)$ and $\nabla : H^k(M) \to H^{k-1}_{\mathfrak{X}}(M)$ are bounded linear mappings. Recall Sobolev's Lemma [Tay96, Proposition 4.3.3].

Lemma 4.9 (Sobolev embedding). *Suppose $\varphi \in H^k(M)$ and suppose $m \in \mathbb{N} \cup \{0\}$ satisfies $k > n/2 + m$. Then $\varphi \in C^m(M)$.*

For $k \in \mathbb{N}$, $k \geq n/2 + 1$, define spaces as follows. Let $X^k := H^{k+1}(M) \times H^k_{\mathfrak{X}}(M)$. Since $k \geq n/2 + 1$, we have by the Sobolev Lemma that $(\rho, J) \in X^k$ satisfies $\rho \in C(M)$. Furthermore define

$$
\begin{aligned}
H^k_+(M) &:= \left\{ \rho \in H^k(M) : \rho > 0 \text{ on } M \right\}, \\
\mathcal{U}^k &:= H^{k+1}_+(M) \times H^k_{\mathfrak{X}}(M), \\
Y^k &:= \left\{ \psi \in H^{k-1}(M) : \psi = \operatorname{div} \xi \text{ for some } \xi \in H^k_{\mathfrak{X}}(M) \right\}, \qquad \text{and} \\
Z^k &:= Y^k \times \mathbb{R}.
\end{aligned}
$$

The condition $\rho > 0$ defines an open subset $\mathcal{U}^k \subset X^k$.

Lemma 4.10. *Suppose $k \geq n/2 + 1$. The mapping $\mathcal{C}(\rho, J)$, as given by (23), may be continuously extended to a mapping $\mathcal{C} : \mathcal{U}^k \to \mathbb{R}$. Moreover, the mapping $\mathcal{C}$ is continuously differentiable on $\mathcal{U}^k$ with Fréchet derivative $\mathcal{C}'(\rho, J) \in L(X^k; \mathbb{R})$ given for $(\rho, J) \in \mathcal{U}^k$ by*

$$
\begin{aligned}
\mathcal{C}'(\rho, J) : (\zeta, G) \mapsto & \int_M \left\{ V + \tfrac{1}{2}\|b\|^2 - \frac{1}{2\rho^2}\|J + \tfrac{1}{2}\nabla\rho\|^2 + \tfrac{1}{2}\operatorname{div}\left(b - \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla\rho \right) \right) \right\} \zeta dx \\
& + \int_M \left\langle \left(-b + \frac{1}{\rho}(J + \tfrac{1}{2}\nabla\rho) \right) + A, G \right\rangle \, dx.
\end{aligned}
$$

Proof. We compute the directional (Gateaux) derivative $\mathcal{C}'(\rho, u)$ to be the linear functional on X given by

$$(\zeta, G) \mapsto$$

$$
\int_M \left\{ V + \frac{1}{2}\left\| -b + \frac{1}{\rho}(J + \tfrac{1}{2}\nabla\rho) \right\|^2 \right\} \zeta + \left\langle -b + \frac{1}{\rho}(J + \tfrac{1}{2}\nabla\rho), -\frac{1}{\rho^2}\left(J + \tfrac{1}{2}\nabla\rho \right)\zeta + \frac{1}{2\rho}\nabla\zeta \right\rangle \, \rho \, dx
$$

$$
+ \int_M \left\{ \left\langle -b + \frac{1}{\rho}(J + \tfrac{1}{2}\nabla\rho), \frac{1}{\rho}G \right\rangle \rho + \langle A, G \rangle \right\} dx,
$$

This is after rearranging, and partial integration of the term containing $\nabla\zeta$, equal to the stated expression. The derivative $\mathcal{C}'(\rho, u)$ is a bounded functional on X^k since $V, 1/\rho, d\rho, b, J$ and A are bounded on M. Since the derivative depends continuously on (ρ, u), it is in fact the Fréchet derivative of $\mathcal{C}$. $\qquad\square$

We define the constraint mapping $\mathcal{H} : \mathcal{U}^k \to Z^k$ as

$$(31) \qquad \mathcal{H}(\rho, J) := \left(\operatorname{div} J, \int_M \rho \, dx - 1 \right), \quad (\rho, J) \in \mathcal{U}^k.$$

The following lemma is now immediate.

Lemma 4.11. *The mapping $\mathcal{H}$ is continuously differentiable on X^k, with Fréchet derivative $\mathcal{H}'(\rho, J) \in L(X^k; Z^k)$ given for $(\rho, J) \in X^k$ by*

$$\mathcal{H}'(\rho, J) : (\zeta, G) \mapsto \left(\operatorname{div} G, \int_M \zeta \, dx \right), \quad (\zeta, v) \in X^k.$$

Every $(\rho, J) \in X^k$ is regular for $\mathcal{H}$, thanks to our choice of the function space Z^k.

Lemma 4.12. *Any $(\rho, J) \in X^k$ is a regular point of $\mathcal{H}$ (in the sense of Definition 4.7).*

Proof. Let $(\Psi, \kappa) \in Z^k = Y^k \times \mathbb{R}$. In particular there exists a $\xi \in H^k_{\mathfrak{X}}(M)$ such that $\Psi = \operatorname{div} \xi$. We may pick $G = \xi$, and ζ a constant function such that $\int_M \zeta \, dx = \kappa$. Then $\mathcal{H}'(\rho, J)(\zeta, G) = (\Psi, \kappa)$, showing that $\mathcal{H}'(\rho, J)$ is onto. $\square$

In order to apply the abstract Lagrange multiplier theorem (Lemma 4.8) in a useful manner, we need to give interpretation to the dual spaces $(Z^k)^\star$, and in particular to $(Y^k)^\star$. Recall that the spaces $(H^l(M))^\star$, for $l \in \mathbb{N} \cup \{0\}$ may be canonically identified through the $L^2(M)$-inner product with spaces of distributions, denoted by $H^{-l}(M)$ [Tay96, Proposition 4.3.2]. In other words, if $z \in (H^l(M))^\star$, then there exists a distribution $\Phi \in H^{-l}(M)$ such that $z(\Psi) = \int_M \Phi \Psi \, dx$. Now in case $\Psi \in Y^k$, i.e. $\Psi = \operatorname{div} \xi$ for some $\xi \in H^k_{\mathfrak{X}}(M)$, then $z(\Psi) = \int_M \Phi \operatorname{div} \xi \, dx = -\int_M \langle \nabla \Phi, \xi \rangle \, dx$ for some $\Phi \in H^{-(k-1)}(M)$. Therefore the choice of $\Phi \in H^{-(k-1)}(M)$ representing $z \in (Y^k)^\star$ is fixed up to the addition of a "constant" distribution: if $\nabla \gamma \in H^{-(k-1)}(M)$ vanishes in a weak sense, then $\int_M (\Phi + \gamma) \operatorname{div} \xi \, dx = \int_M \Phi \operatorname{div} \xi \, dx$. We summarize this in the following lemma.

Lemma 4.13. $(Y^k)^\star \cong H^{-(k-1)}(M)/\left\{ \gamma \in H^{-(k-1)}(M) : \nabla \gamma = 0 \right\}.$

We may now apply Lemma 4.8 to obtain the following preliminary result.

Proposition 4.14 (Gauge invariant Hamilton-Jacobi-Bellman equation – necessity)**.** *Suppose $(\rho, J) \in C^\infty \times \mathfrak{X}(M)$ is a local extremum of $\mathcal{C}$, defined by (23), under the constraint that $\mathcal{H}(\rho, J) = 0$. Then there exists $\Psi \in C^\infty(M)$ and $\lambda \in \mathbb{R}$ satisfying (28) and such that $J = -\frac{1}{2}\nabla \rho + \rho(b + \nabla \Psi - A)$. The corresponding control field $u \in \mathfrak{X}(M)$ is given by $u = \nabla \Psi - A$.*

Proof. Let $k \geq n/2 + 1$. Then $(\rho, J) \in \mathcal{U}^k$. By Lemma 4.8, there exists an element $(\Psi, \lambda) \in H^{-(k-1)}(M) \times \mathbb{R}$ such that the following equations hold.

$$V + \frac{1}{2}\|b\|^2 - \frac{1}{2\rho^2}\|J + \tfrac{1}{2}\nabla \rho\|^2 + \frac{1}{2} \operatorname{div}\left(b - \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla \rho \right) \right) + \lambda = 0, \quad \text{and}$$

$$\left(-b + \frac{1}{\rho}(J + \tfrac{1}{2}\nabla \rho) \right) + A - \nabla \Psi = 0.$$

(By Lemma 4.13 Ψ is defined up to a constant). Substituting the second equation into the first, and making some rearrangements, gives the sytem (28). Then $\Psi \in C^\infty(M)$ as a result of the expression for J. The expression for u is an immediate result of (21). $\square$

By letting $\psi = \exp(\Psi)$, we immediately obtain the following result.

Corollary 4.15 (Linear gauge invariant HJB equation – necessity)**.** *Suppose $(\rho, J) \in \mathcal{U}$ is a local extremum of $\mathcal{C}$, defined by (23), under the constraints (25). Then there exists a $\psi \in C^\infty(M)$, $\psi > 0$ on M, and $\lambda \in \mathbb{R}$ such that (29) holds. Furthermore $\rho \in C^\infty(M)$, $J \in \mathfrak{X}(M)$ and the associated control field $u \in \mathfrak{X}(M)$ are related by (30).*

Since the problem considered in Corollary 4.5 is a relaxed version of Problem 3.12, and smoothness of J and ρ is established in the relaxed case, we immediately have the following corollary.

Corollary 4.16. *Suppose (ρ, u) is a solution of Problem 3.6, or equivalently that (ρ, J) a solution of Problem 3.12. Then the results of Corollary 4.15 hold.*

Remark 4.17. Combining Corollaries 4.5 and 4.15, taking $A = 0$, gives a variational character-
ization of the principal eigenvalue of an elliptic differential operator $L = \frac{1}{2}\Delta + b\nabla - V$, given
by

$$\lambda^\star = - \inf_{\substack{(\rho, J) \in C^\infty \times \mathfrak{X}(M) \\ \int_M \rho \ dx = 1 \text{ and } \mathrm{div} \ J = 0}} \mathcal{C}(\rho, J).$$

This result may be compared to [DV75a], in which a variational principle is derived for the maximal
eigenvalue of an operator L satisfying a maximum principle.

4.3. **Reversible solution.** In the reversible case we can represent the optimally controlled in-
variant measure in terms of ψ and the uncontrolled invariant measure. Let $(T(t))_{t\geq 0}$ denote the
transition semigroup of a diffusion on M. The diffusion is said to be *reversible* if there exists a
Borel measure $\nu(dx)$ such that

$$\int_M (T(t)f)(x)g(x) \ \nu \ dx = \int_M f(x)(T(t)g)(x) \ \nu \ dx \quad \text{for all } f, g \in C(M), t \geq 0.$$

Other equivalent terminology is that the Markov process is *symmetrizable* or that the invariant
measure satisfies *detailed balance* [IW89, Section V.4]. In case a diffusion is reversible with respect
to a measure ν, this measure is an invariant measure for the diffusion.

   The following results hold for any control field $u \in \mathfrak{X}(M)$.

Lemma 4.18. *Let X denote a diffusion with generator given by $Lh = \frac{1}{2}\Delta h + \langle (b + u), \nabla h \rangle$,
$h \in C^2(M)$. The following are equivalent.*

   (i) *X is reversible, with invariant density $\rho_u = \exp(-U)$ for some $U \in C^\infty(M)$;*
   (ii) *$b + u = -\frac{1}{2}\nabla U$ for some $U \in C^\infty$;*
   (iii) *The long term current density J, given by (20), vanishes.*

Proof. The equivalence of (i) and (ii) is well known, see e.g. [IW89, Theorem V.4.6]. The equiva-
lence of (ii) and (iii) is then immediate from (20). $\qquad\square$

Proposition 4.19. *Let ρ_0 and $J_0 = -\frac{1}{2}\nabla\rho_0 + \rho_0 b$ denote the density and current corresponding
to the uncontrolled dynamics (9). The following are equivalent.*

   (i) *The diffusion corresponding to the optimal control u is reversible, with density $\rho = \psi^2\rho_0$,
        where ψ is as in Corollary 4.5 (normalized such that $\int_M \psi^2\rho_0 \ dx = 1$);*
   (ii) *$J_0 = \rho_0 A$;*
   (iii) *$b = -\frac{1}{2}\nabla U + A$, for some $U \in C^\infty(M)$.*

   *In particular, if the uncontrolled diffusion is reversible and $A = 0$, then the controlled difussion
is reversible and the density admits the expression given under (i).*

Proof. Setting $\rho = \psi^2\rho_0$, we have

$$J = -\tfrac{1}{2}\nabla\rho + \rho\left(b - A + \nabla\log\psi\right) = -\tfrac{1}{2}\psi^2\nabla\rho_0 - \rho_0\psi\nabla\psi + \psi^2\rho_0\left(b - A + \nabla\log\psi\right)$$
$$= -\tfrac{1}{2}\psi^2\nabla\rho_0 + \psi^2\rho_0(b - A) = \psi^2(J_0 - \rho_0 A),$$

which establishes the equivalence of (i) and (ii). Representing the density ρ_0 by $\exp(-U)$ and
using (20) with $u = 0$ gives the equivalence between (ii) and (iii). $\qquad\square$

4.4. **Gauge invariance.** For a special choice of A, the solution of Problem 3.12 may be related
to the solution corresponding to $A = 0$ in a simple way.

Proposition 4.20. *Let $A_0 \in \mathfrak{X}(M)$. For $\varphi \in C^\infty(M)$ and $A = A_0 + \nabla\varphi$ let $(\rho_\varphi, J_\varphi)$ denote the
solution of Problem 3.12 with corresponding solutions ψ_φ and u_φ of (29). Then, for $\varphi \in C^\infty(M)$,*

(32) $$\rho_\varphi = \rho, \quad u_\varphi = u, \quad \psi_\varphi = \exp(\varphi)\psi, \quad and \quad J_\varphi = J,$$

where ρ, J, ψ and u denote the solution of Problem 3.12 and (29) corresponding to $A = A_0$.

Proof. This is a matter of straightforward computation. $\qquad\square$

   In other words, the solution of Problem 3.12 depends (essentially) only on the equivalence class
of A, under the equivalence relation $A \sim B$ if and only if $A = B + \nabla\varphi$ for some $\varphi \in C^\infty(M)$.

Remark 4.21. A standard way in physics to obtain gauge invariant differential operators is to replace the gradients with 'long' derivatives. This phenomenon also occurs in the linear Bellman equation. In differential form notation (interpreting A as differential form; see Remark 2.1) the linear operator on the left hand side of (29) can be written as

$$H\psi = \tfrac{1}{2} \star (d - A\wedge) \star (d - A\wedge) \psi + \star \left(f \star (d - A\wedge)\right)\psi - V\psi.$$

The operator $\psi \mapsto d\psi - A \wedge \psi$ is called a *'long' derivative* operator. This result is easily verified using the standard relations

$$\star(\alpha \wedge \star\beta) = \langle \alpha, \beta \rangle, \quad \star d \star d\phi = \Delta\phi, \quad \star d \star \alpha = -\delta\alpha,$$

for $\alpha, \beta \in \Lambda^1(M)$ and $\phi \in C^\infty(M)$.

See also the directly related expression [BC15, Equation (31)], where a similar twisted second order generator occurs in relation to a tilting of a Markov process to accommodate for currents.

5. Fixed density

In this section we consider the problem of fixing the density function ρ, and finding a force u that obtains this density function, at minimum cost. Let $\rho \in C^\infty(M)$ be fixed, with $\rho > 0$ on M and $\int_M \rho \, dx = 1$. Then for some constant c_ρ we have

$$\mathcal{C}(\rho, u) = c_\rho + \int_M \left\{ \frac{1}{2}\|u\|^2 + \langle A, u \rangle \right\} \rho \, dx.$$

Therefore we will consider the following problem.

Problem 5.1. Minimize $\mathcal{C}(u)$ over $\mathfrak{X}(M)$, subject to the constraint $L_u^\star \rho = 0$, where $\mathcal{C}(u)$ is defined by

$$\mathcal{C}(u) = \int_M \left\{ \frac{1}{2}\|u\|^2 + \langle A, u \rangle \right\} \rho \, dx.$$

The corresponding problem in terms of the curent density is the following. As for $\mathcal{C}(u)$, terms that do not depend on J are eliminated from the cost functional.

Problem 5.2. Minimize $\mathcal{C}(J)$ over $\mathfrak{X}(M)$, subject to the constraint $\operatorname{div} J = 0$, where $\mathcal{C}(J)$ is defined by

$$(33) \qquad \mathcal{C}(J) = \int_M \left(\left\langle -b + \tfrac{1}{2}\nabla \log \rho + A, J \right\rangle + \frac{1}{2\rho}\|J\|^2 \right) \, dx.$$

By a standard variational argument we obtain the following result, which states that a relaxed version of Problem 5.2 may be transformed into an elliptic PDE. Essentially, this is obtained through variation of the Lagrangian functional

$$\mathcal{L}(J, \Phi) = \mathcal{C}(J) + \int_M \Phi \operatorname{div} J \, dx.$$

Theorem 5.3. *Suppose $J \in \mathfrak{X}(M)$ is a local extremum of $\mathcal{C}(J)$ given by (33) under the constraint that $\operatorname{div} J = 0$. Then there exists a $\Phi \in C^\infty(M)$ such that*

$$(34) \qquad \Delta\Phi + \langle \nabla \log \rho, \nabla\Phi \rangle = \frac{1}{\rho} \left(\tfrac{1}{2}\Delta\rho - \operatorname{div}(\rho(b - A)) \right).$$

For this Φ, J is given by

$$J = \rho f - \tfrac{1}{2}\nabla\rho + \rho(\nabla\Phi - A),$$

and the corresponding control field u is given by

$$u = \nabla\Phi - A.$$

Remark 5.4. It appears that the result of Theorem 5.3 can be thought of as a gauge invariant extension of the characterization (7) of the Donsker-Varadhan functional by Gärtner [G77]. Taking $A = 0$ and comparing with (7) we may rephrase the Donsker-Varadhan rate functional (5), as

$$I(\mu) = \inf_{\substack{u \in \mathfrak{X}(M) \\ L_u^\star \rho = 0}} \frac{1}{2} \int_M \|u\|^2 \, \rho \, dx$$

for μ absolutely continuous with density ρ. (Note that strictly speaking, we need to be careful since we have only obtained Theorem 5.3 under the condition that a local minimum exists.)

Remark 5.5 (Solution in the reversible case). If $b - A = -\frac{1}{2}\nabla U$ for some $U \in C^\infty(M)$, it may be checked that $\Phi = \frac{1}{2}(\ln \rho + U)$ solves (34), so that the optimal control field $u = \frac{1}{2}\nabla(\ln \rho) - b$. In other words, the optimal way to obtain a particular density function ρ if $b - A$ is in 'gradient form' is by using a control u so that the resulting force field $b + u$ is again in gradient form, $b + u = \frac{1}{2}\nabla(\ln \rho)$, resulting in reversible dynamics; see also Section 4.3.

Example 5.6 (Circle). On S^1 every differential 1-form β, and in particular $\beta = b - A$, may be written as $\beta = -\frac{1}{2} \, dU + \frac{1}{2}k \, d\theta$, where θ represents the polar coordinate function, $U \in C^\infty(S^1)$ and $k \in \mathbb{R}$; see e.g. [War83, Example 4.14]. Equation (34) then reads

$$\Phi''(\theta) + \frac{1}{\rho}\Phi'(\theta)\rho'(\theta) = \frac{1}{2\rho}\left(\rho''(\theta) - \frac{d}{d\theta}(-\rho(\theta)U'(\theta) + k\rho)\right).$$

Based on Remark 5.5, we try a solution of the form $\Phi = \frac{1}{2}(\ln \rho + U - \varphi)$. Inserting this into the differential equation, we obtain for φ the equation

$$\varphi''(\theta) + \gamma(\theta)\varphi'(\theta) = k\gamma(\theta),$$

where $\gamma(\theta) = \frac{d}{d\theta}\ln \rho(\theta) = \frac{1}{\rho(\theta)}\rho'(\theta)$.

Up to an arbitrary additive constant (which we put to zero), there exists a unique periodic solution φ to this differential equation, given by

$$\varphi(\theta) = \frac{k\left(\theta \int_0^{2\pi} \rho(\xi)^{-1} \, d\xi - 2\pi \int_0^\theta \rho(\xi)^{-1} \, d\xi\right)}{\int_0^{2\pi} \rho(\xi)^{-1} \, d\xi}, \quad \theta \in [0, 2\pi].$$

Remark 5.7 (Gauge invariance). As in Section 4.4, it is straightforward to check that a solution to Problem 5.2 for $A_\varphi = A_0 + \nabla\varphi$ is given by

$$\Phi_\varphi = \Phi + \varphi, \quad \rho_\varphi = \rho, \quad u_\varphi = u, \quad J_\varphi = J,$$

in terms of the solution (Φ, ρ, u, J) corresponding to the gauge field A_0.

6. FIXED CURRENT DENSITY

In this section we approach the problem of minimizing the average cost, under the constraint that J is fixed. In light of the remark just below (20), it will be necessary to demand that $\operatorname{div} J = 0$, otherwise we will not be able to obtain a solution. Hence in the remainder of this section let $J \in \mathfrak{X}(M)$ satisfying $\operatorname{div} J = 0$ be fixed. By (21), we may express u in terms of J and ρ by $u = -b + \frac{1}{\rho}\left(J + \frac{1}{2}\nabla\rho\right)$. Note that by Lemma 3.9, the Fokker-Planck equation (14) is satisfied. This leads to the following problem.

Problem 6.1. Minimize $\mathcal{C}(\rho)$ subject to the constraint $\int_M \rho \, dx = 1$, where

$$(35) \qquad \mathcal{C}(\rho) = \int_M \left(V + \frac{1}{2}\left\|-b + \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla\rho\right)\right\|^2\right)\rho \, dx.$$

Remark 6.2. The constraint $\rho \geq 0$ on M does not need to be enforced, since if we find ρ solving Problem 6.1 without this constraint, we may compute u by (21). Then ρ satisfies $L_u^\star \rho = -\operatorname{div} J = 0$ by Lemma 3.9, so by Proposition 3.3 and the constraint $\int_M \rho \, dx = 1$, it follows that $\rho > 0$ on M.

Remark 6.3. Note that by Lemma 3.10, the contribution of A is determined once we fix J. Therefore we may put $A = 0$ in the current optimization problem.

Necessary conditions for the solution of Problem 6.1 may be obtained rigorously in a similar manner as in Section 4, to obtain the following result.

Theorem 6.4. *Suppose* $\rho \in C^\infty(M)$ *minimizes* $\mathcal{C}(\rho)$ *given by* (35) *under the constraint that* $\int_M \rho \, dx = 1$. *Then there exists a* $\mu \in \mathbb{R}$ *such that*

$$(36) \qquad \qquad \tfrac{1}{2}\Delta\phi - (W + \mu)\phi = -\frac{\|J\|^2}{2\phi^3},$$

holds, where $\phi = \sqrt{\rho}$ *and* $W = V + \tfrac{1}{2}\|b\|^2 + \tfrac{1}{2}\operatorname{div} b$.

Remark 6.5. Equation (36) is known (at least in the one-dimensional case) as *Yermakov's equation* [PZ03].

Instead of proving Theorem 6.4 rigorously (which may be done analogously to Section 4) we provide an informal derivation, which we hope provides more insight to the reader. We introduce the Lagrangian $\mathcal{L} : C^\infty(M) \times \mathbb{R} \to \mathbb{R}$ by

$$\mathcal{L}(\rho, \mu) = \int_M \left(V + \frac{1}{2}\left\| -b + \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla\rho\right) \right\|^2 \right) \rho \, dx + \mu \left(\int_M \rho \, dx - 1 \right).$$

Varying $\mathcal{L}(\rho, \mu)$ with respect to ρ in the direction $\zeta \in C^\infty(M)$ gives

$\mathcal{L}'(\rho, \mu)\zeta$

$$= \int_M \left(V + \frac{1}{2}\left\| -b + \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla\rho\right) \right\|^2 + \mu \right)\zeta + \left\langle \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla\rho\right) - b, \frac{1}{2\rho}\nabla\zeta - \frac{1}{\rho^2}\left(J + \tfrac{1}{2}\nabla\rho\right)\zeta \right\rangle \rho \, dx$$

$$= \int_M \left(V + \mu + \tfrac{1}{2}\|b\|^2 - \frac{1}{2\rho^2}\|J\|^2 + \tfrac{1}{2}\operatorname{div} b + \frac{1}{8\rho^2}\|\nabla\rho\|^2 - \frac{1}{4\rho}\Delta\rho \right)\zeta \, dx,$$

where we used $\operatorname{div} J = 0$. We require that for any direction ζ the above expression equals zero, which is the case if and only if

$$(37) \qquad \qquad -\frac{1}{4\rho}\Delta\rho + \frac{1}{8\rho^2}\|\nabla\rho\|^2 - \frac{1}{2\rho^2}\|J\|^2 + W + \mu = 0.$$

Note that we need to solve this equation for both ρ and μ, in combination with the constraint that $\rho > 0$ on M and $\int_M \rho \, dx = 1$. By substituting $\phi = \sqrt{\rho}$, equation (37) transforms into the equation (36).

We may then compute the cost corresponding to $\rho = \phi^2$ as

$$\mathcal{C}(\rho) = \int_M \left(V + \frac{1}{2}\left\| -b + \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla\rho\right) \right\|^2 \right) \rho \, dx$$

$$= \int_M \frac{1}{4}\Delta\rho - \frac{1}{8\rho}\|\nabla\rho\|^2 + \frac{1}{2\rho}\|J\|^2 - \frac{\rho}{2}\|b\|^2 - \frac{\rho}{2}\operatorname{div} b - \mu\rho + \frac{\rho}{2}\left\| -b + \frac{1}{\rho}\left(J + \tfrac{1}{2}\nabla\rho\right) \right\|^2 \, dx$$

$$(38) \qquad = \int_M \frac{1}{4}\Delta\rho - \frac{\rho}{2}\operatorname{div} b - \mu\rho - \langle b, J \rangle - \rho\left\langle b, \frac{1}{2\rho}\nabla\rho \right\rangle + \frac{1}{\rho}\langle J, \tfrac{1}{2}\nabla\rho \rangle \, dx$$

$$= -\left(\mu + \int_M \langle b, J \rangle \, dx \right),$$

where we used (37) in the first equality, and the following relations for the last equality:

$$\int_M \Delta\rho \, dx = 0, \qquad\qquad \int_M \rho(\operatorname{div} b) \, dx = -\int_M \langle \nabla\rho, b \rangle \, dx,$$

$$\int_M \frac{1}{\rho}\langle J, \nabla\rho \rangle \, dx = -\int_M (\operatorname{div} J)\ln\rho \, dx = 0, \qquad\qquad \int_M \rho \, dx = 1.$$

We can only influence the first term in (38) by choosing ρ or μ, so we see that minimizing $\mathcal{C}$ therefore corresponds to finding the largest value of μ such that (37), or, equivalently, (36), admits a solution.

6.1. Reversible solution – stationary Schrödinger equation.
In this section we consider the special case of the above problem for zero current density, $J = 0$. By Lemma 4.18, this is equivalent to $u + b = -\frac{1}{2}\nabla\Psi$ for some unknown $\Psi \in C^\infty(M)$, with $\rho = \exp(-\Psi)$. In other words, we demand the net force field (including the control) to be in gradient form, and the corresponding diffusion to be reversible; see Section 4.3.

In this case (36) transforms into the linear eigenvalue problem,

$$(39) \qquad \frac{1}{2}\Delta\phi - W\phi = \mu\phi.$$

This is intriguing since this is in fact a time independent Schrödinger equation for the square root of a density function, analogous to quantum mechanics; even though our setting is entirely classical. See also [CCBK14, equation (13)] and the discussion following it.

By (38), we are interested in the largest value of μ so that (39) has a solution ϕ. The optimal control field is then given by $u = \frac{1}{\phi}\nabla\phi - b$.

Remark 6.6. It is straightforward to check that if $b = -\frac{1}{2}\nabla U$ for some $U \in C^\infty(M)$, then $\phi = \exp(-\frac{1}{2}U)$ satisfies (39) with $V = 0$, $\mu = 0$, resulting in $u = 0$. This corresponds to the intuition that, if b is already a gradient, no further control is necessary to obtain a reversible invariant measure.

Remark 6.7. We may also compare the case $b = -\frac{1}{2}\nabla U$ with the result of Section 4.3. There we obtained that, in case $A = 0$ and $b = -\frac{1}{2}\nabla U$, the optimization problem for unconstrained J resulted in a reversible solution. In other words, the constraint $J = 0$ does not need to be enforced, and the solution of this section should equal the solution obtained in Proposition 4.19. Apparently, with ψ as in Proposition 4.19, we have that $\phi^2 = \psi^2 \exp(-U)$.

7. Discussion

In this paper we showed how stationary long term average control problems are related to eigenvalue problems (for the unconstrained problem and the problem constrained to a reversible solution, Sections 4 and 6.1), elliptic PDEs (for the problem with fixed density, Section 5) or a non-linear eigenvalue problem (for the problem with fixed current density, Section 6). For this we fruitfully used the representation of an optimal control field u in terms of the density function ρ and the current density J.

The theory on existence of solutions and spectrum of operators is classical and we refer the interested reader to e.g. [Eva10, Tay96, War83].

An interesting problem for further research is the following. One may ask the question whether we may obtain solutions when we constrain a certain flux $\int_M \langle A, J\rangle \, dx$ (see Section 3.1) to a given value. In this case, one may use $\widetilde{A} = \mu A$ as a Lagrange multiplier and use the results of Section 4 (for constrained flux) and Section 5 (for constrained flux and density) to obtain necessary conditions.

Apparently, in the case of a diffusion, density and current density completely characterize the ergodic behaviour. It seems an interesting problem to formulate a general theory which extracts from a given Markov processes the precise quantities which specify the long term behaviour of the process.

APPENDIX A. FLUX

In this section we will give a natural interpretation of the term $\int_M \langle A, J \rangle \; dx$, namely as the flux of J through a cross-section α, or equivalently, the long term average intersection index of the stochastic process $(X(t))_{t \geq 0}$ with a cross-section. The section motivates the gauge field A in the cost function. The remainder of this paper does not refer to this section. For background reading in differential geometry, see [Arn89, War83]. See also [CCMT09, CCBK14] where the considerations below are discussed from a physics perspective.

Let M be a compact, oriented, Riemannian manifold of dimension n. Recall the notion of a *singular p-chain in M (with real coefficients)* as a finite linear combination $c = \sum a_i \sigma_i$ of smooth p-simplices σ_i in M where the a_i are real numbers. Let $S_p(M; \mathbb{R})$ denote the real vector space of singular p-chains in M. On $S_p(M; \mathbb{R})$, $p \in \mathbb{Z}$, $p \geq 0$, are defined boundary operators $\partial_p : S_p(M; \mathbb{R}) \to S_{p-1}(M; \mathbb{R})$. The *p-th singular homology group of M with real coefficients* is defined by

$$\mathcal{H}_p(M; \mathbb{R}) = \ker \partial_p / \operatorname{im} \partial_{p+1}.$$

Elements of $\ker \partial_p$ are called p-cycles, and elements of $\operatorname{im} \partial_{p+1}$ are called p-boundaries. The *deRham cohomology classes* $\mathcal{H}^p_{\mathrm{deRham}}(M; \mathbb{R})$ are defined, for $0 \leq p \leq m = \dim(M)$ as

$$\mathcal{H}^p_{\mathrm{deRham}}(M) = \ker d_p / \operatorname{im} d_{p-1},$$

where $d_p : \Lambda^p(M) \to \Lambda^{p+1}(M)$ denotes exterior differentiation.

Let α be a p-cycle. The functional $p_\alpha \in \Lambda^p(M) \to \int_\alpha \beta \in \mathbb{R}$ depends, by Stokes' theorem, only on the homology class $[\alpha]$ of α, and the deRham cohomology class $[\beta]$ of β. An element $[\alpha] \in \mathcal{H}_p(M; \mathbb{R})$ may therefore be considered an element of $(\mathcal{H}^p_{\mathrm{deRham}}(M))^\star$. Now let M be oriented and of dimension n. The mapping $q_\beta : [\gamma] \in \mathcal{H}^p_{\mathrm{deRham}}(M) \to \int_M \beta \wedge \gamma \in \mathbb{R}$ is an element of $(\mathcal{H}^p_{\mathrm{deRham}}(M))^\star$. Poincaré duality states that the mapping $[\beta] \in \mathcal{H}^{m-p}_{\mathrm{deRham}}(M) \to q_\beta \in (\mathcal{H}^p_{\mathrm{deRham}}(M))^\star$ is an isomorphism for compact M, i.e. $\mathcal{H}^{m-p}_{\mathrm{deRham}}(M) \cong (\mathcal{H}^p_{\mathrm{deRham}}(M))^\star$. Therefore, for compact, oriented M, we have

$$\mathcal{H}_p(M; \mathbb{R}) \cong (\mathcal{H}^p_{\mathrm{deRham}}(M))^\star \cong \mathcal{H}^{m-p}_{\mathrm{deRham}}(M).$$

Since an equivalence class in $\mathcal{H}^{m-p}_{\mathrm{deRham}}(M)$ has a unique harmonic representative, we conclude that for a p-cycle α, there exists a unique harmonic $r_\alpha \in \Lambda^{m-p}(M)$ such that

$$\int_\alpha \beta = p_\alpha(\beta) = \int_M r_\alpha \wedge \beta, \quad [\beta] \in \mathcal{H}^p_{\mathrm{deRham}}(M).$$

In particular, for $p = m - 1$, we may interpret $\int_\alpha \star J$ (with $\star$ the *Hodge star operator*) as the flux of J through α. This quantity may further be interpreted as the long term average intersection index of the stochastic trajectory $(X(t))_{t \geq 0}$ with respect to α, i.e. the long term average of the number of intersections (with ± 1 signs depending on the direction); see e.g. [GH94, Section 0.4]. Specializing the above result to this situation, we obtain the following proposition.

Proposition A.1. *For a given $(n-1)$-cycle α, there exists a unique harmonic $A \in \Lambda^1(M)$, which depends only on the singular homology class $[\alpha] \in \mathcal{H}_{m-1}(M; \mathbb{R})$ of α, such that $\int_\alpha \star J = \int_M \langle A, J \rangle \; dx$ for all $J \in \Lambda^1(M)$ satisfying $\delta J = 0$.*

APPENDIX B. DERIVATION OF EXPRESSION FOR LONG TERM AVERAGE OF CURRENT DENSITY

In the physics literature (see e.g. [CCMT09]), the current density is defined formally as

$$(40) \qquad J^i_t(x) = \frac{1}{t} \int_0^t \dot{X}^i_s \delta(X_s - x) \; ds,$$

for $x \in M$, where δ is the Dirac delta function. We will derive an alternative expression for this quantity, using the model (10) for the dynamics. Note that (40) formally defines a vector field that acts on functions $f \in C^\infty(M)$ as

$$J_t f(x) = \frac{1}{t} \int_0^t \dot{X}^i_s \delta(X_s - x) \partial_i f(x) \; ds = \frac{1}{t} \int_0^t \dot{X}^i_s \delta(X_s - x) \partial_i f(X_s) \; ds = \frac{1}{t} \int_0^t \delta(X_s - x) \partial_i f(X_s) \circ dX^i_s.$$

The δ-function is still problematic. We may however formally compute the $L^2(M, g)$ inner product of the above expression with any $h \in C^\infty(M)$ with support in a coordinate neighbourhood U containing x. This results in

$$\int_M h(x) J_t f(x) \; dx = \frac{1}{t} \int_U h(x) \int_0^t \delta(X_s - x) \partial_i f(X_s) \circ dX_s^i \; dx = \frac{1}{t} \int_0^t h(X_s) \partial_i f(X_s) \circ dX_s^i.$$

Using the relation $Y \circ dZ = Y \; dZ + \frac{1}{2} d[Y, Z]$ and (13), we compute

$$\int_U h(x) J_t f(x) \; dx = \frac{1}{t} \int_0^t h(X_s) \partial_i f(X_s) \left(\overline{b}_u^i(X_s) ds + \sigma_\alpha^i(X_s) \; dB_s^\alpha \right) + \frac{1}{2} \frac{1}{t} \int_0^t \sigma_\alpha^i \sigma_\alpha^j \partial_j (h \partial_i f)(X_s) \; ds$$

$$\rightarrow \int_U \left\{ h(x)(\partial_i f)(x) \left(\overline{b}_u^i(x) \right) + \frac{1}{2} g^{ij} \partial_j (h \partial_i f)(x) \right\} \rho_u \; dx \quad \text{(almost surely as } t \rightarrow \infty)$$

$$= \int_U h(x) \left\{ \rho_u (\partial_i f) \left(\overline{b}_u^i \right) - \frac{1}{2} \frac{1}{\sqrt{|g|}} (\partial_i f) \partial_j \left(\rho_u \sqrt{|g|} g^{ij} \right) \right\}(x) \; dx,$$

using Proposition 3.4 and the law of large numbers for martingales [NP82]. We find that the long term average vector field J has components

$$J^i = \rho_u \overline{b}_u^i - \frac{1}{2} \frac{1}{\sqrt{|g|}} \partial_j \left(\rho_u \sqrt{|g|} g^{ij} \right) = \rho_u \left(b_u^i + \frac{1}{2} \sigma_\alpha^k (\partial_k \sigma_\alpha^i) \right) - \frac{1}{2} g^{ij} (\partial_j \rho_u) - \frac{1}{2} \rho_u \frac{1}{\sqrt{|g|}} \partial_j \left(\sqrt{|g|} g^{ij} \right)$$

$$= \rho_u b_u^i - \frac{1}{2} g^{ij} (\partial_j \rho_u),$$

where the last equality is a result of the identity

$$(\nabla_{\sigma_\alpha} \sigma_\alpha)^i = \sigma_\alpha^k (\partial_k \sigma_\alpha^i) - \frac{1}{\sqrt{|g|}} \partial_j \left(\sqrt{|g|} g^{ij} \right),$$

which may be verified by straightforward calculation.

References

[ABG12]   A Arapostathis, VS Borkar, and MK Ghosh. *Ergodic control of diffusion processes.* 2012.

[Arn89]   V.I. I Arnol'd. *Mathematical Methods of Classical Mechanics (Graduate Texts in Mathematics)*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1989.

[BC15]   Andre C. Barato and Raphael Chetrite. A Formal View on Level 2.5 Large Deviations and Fluctuation Relations. *Journal of Statistical Physics*, 160(5):1154–1172, May 2015.

[BFG15]   L. Bertini, A. Faggionato, and D. Gabrielli. Flows, currents, and cycles for Markov Chains: large deviation asymptotics. *Stochastic Processes and their Applications*, 125(7):38, 2015.

[BG88]   Vivek S Borkar and Mrinal K Ghosh. Ergodic control of multidimensional diffusions. I. The existence results. *SIAM J. Control Optim.*, 26(1):112–126, 1988.

[BK14]   Joris Bierkens and Hilbert J. Kappen. Explicit solution of relative entropy weighted control. *Systems & Control Letters*, 72:36–43, May 2014.

[Bry90]   Wlodzimierz Bryc. Large deviations by the asymptotic value method. In *Diffusion processes and related problems in analysis, Vol. I (Evanston, IL, 1989)*, volume 22 of *Progr. Probab.*, pages 447–472. Birkhäuser Boston, Boston, MA, 1990.

[CCBK14]   Vladimir Y. Chernyak, Michael Chertkov, Joris Bierkens, and Hilbert J. Kappen. Stochastic Optimal Control as Non-equilibrium Statistical Mechanics: Calculus of Variations over Density and Current. *Journal of Physics A: Mathematical and Theoretical*, 47(2):4, June 2014.

[CCMT09]   Vladimir Y. Chernyak, Michael Chertkov, Sergey V. Malinin, and Razvan Teodorescu. Non-Equilibrium Thermodynamics and Topology of Currents. *Journal of Statistical Physics*, 137(1):109–147, September 2009.

[dH00]   Frank den Hollander. *Large deviations*, volume 14 of *Fields Institute Monographs*. American Mathematical Society, Providence, RI, 2000.

[DV75a]   M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, I. *Communications on Pure and Applied Mathematics*, XXVIII:1–47, 1975.

[DV75b]   M. D. Donsker and S. R. S. Varadhan. On a Variational Formula for the Principal Eigenvalue for Operators with Maximum Principle. *Proceedings of the National Academy of Sciences*, 72(3):780–783, March 1975.

[DV76]   M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, III. *Communications on Pure and Applied Mathematics*, 29(4):389–461, July 1976.

[DZ96]   G Da Prato and J Zabczyk. *Ergodicity for infinite-dimensional systems*, volume 229 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1996.

[Eva10]   Lawrence C Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[FK06]   Jin Feng and Thomas G. Kurtz. *Large Deviations for Stochastic Processes (Mathematical Surveys and Monographs)*. American Mathematical Society, 2006.

[Fle82]   Wendell H Fleming. Logarithmic transformations and stochastic control. In *Advances in filtering and optimal stochastic control (Cocoyoc, 1982)*, volume 42 of *Lecture Notes in Control and Inform. Sci.*, pages 131–141. Springer, Berlin, 1982.

[Fle85]   Wendell H Fleming. A stochastic control approach to some large deviations problems. In *Recent mathematical methods in dynamic programming (Rome, 1984)*, volume 1119 of *Lecture Notes in Math.*, pages 52–66. Springer, Berlin, 1985.

[For01]   A De La Fortelle. Large Deviation Principle for Markov Chains in Continuous Time. *Problems of Information Transmission*, 37(2):120–139, 2001.

[FR75]   WH Fleming and RW Rishel. *Deterministic and stochastic optimal control*. 1975.

[FS09]   Wendell H. Fleming and Halil Mete Soner. *Controlled Markov Processes and Viscosity Solutions (Stochastic Modelling and Applied Probability)*. Springer, 2009.

[FSS87]   W.H. Fleming, S.J. Sheu, and H.M. Soner. A remark on the large deviations of an ergodic markov process. *Stochastics*, 22(3-4):187–199, November 1987.

[G̈77]   J Gärtner. On large deviations from the invariant measure. *Theory of Probability & Its Applications*, pages 24–40, 1977.

[GH94]   Phillip Griffiths and Joseph Harris. *Principles of algebraic geometry*. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1994.

[Hol78]   Charles J. Holland. A minimum principle for the principal eigenvalue for second-order linear elliptic equations with natural boundary conditions. *Communications on Pure and Applied Mathematics*, 31(4):509–519, July 1978.

[Hop50]   Eberhard Hopf. The partial differential equation u_t+uu_x=μu_{xx}. *Comm. Pure Appl. Math.*, 3:201–230, 1950.

[IW89]   Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1989.

[Kac51]   M Kac. On some connections between probability theory and differential and integral equations. *Proceedings of the Second Berkeley Symposium on Mathematical Statstics and Probability*, page 189, 1951.

[Kap05]   Hilbert J. Kappen. Linear Theory for Control of Nonlinear Stochastic Systems. *Physical Review Letters*, 95(20):200201, November 2005.

[Kar80]   Ioannis Karatzas. On a stochastic representation for the principal eigenvalue of a second-order differential equation. *Stochastics*, 3(1-4):305–321, December 1980.

[Lue69]   David G Luenberger. *Optimization by vector space methods*. John Wiley & Sons Inc., New York, 1969.

[NP82]   Hung T Nguyen and Tuan D Pham. On the law of large numbers for continuous-time martingales and applications to statistics. *Stochastica*, 6(1):5–23, 1982.

[PZ03]   Andrei D Polyanin and Valentin F Zaitsev. *Handbook of exact solutions for ordinary differential equations*. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2003.

[RBW08]   Per Rutquist, Claes Breitholtz, and Torsten Wik. On the infinite time solution to state-constrained stochastic optimal control problems. *Automatica*, 44(7):1800–1805, 2008.

[Ris89]   H Risken. *The Fokker-Planck equation: Methods of Solution and Applications*, volume 18 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, second edition, 1989.

[She84]   Shuenn-Jyi Sheu. Stochastic control and principal eigenvalue. *Stochastics*, 11(3-4):191–211, 1984.

[Tay96]   Michael E Taylor. *Partial differential equations. I Basic Theory*, volume 115 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.

[Tod06]   Emanuel Todorov. Linearly-solvable Markov decision problems. In B Schölkopf, J Platt, and T Hoffman, editors, *Advances in Neural Information Processing Systems 19*, number 1, pages 1369–1376. MIT Press, Cambridge, MA, 2006.

[Var84]   S R S Varadhan. *Large deviations and applications*, volume 46 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984.

[War83]   Frank W Warner. *Foundations of differentiable manifolds and Lie groups*, volume 94 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1983.

24   JORIS BIERKENS, VLADIMIR Y. CHERNYAK, MICHAEL CHERTKOV, AND HILBERT J. KAPPEN

J. Bierkens, University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK
E-mail address: `j.bierkens@warwick.ac.uk`

V. Y. Chernyak, Wayne State University, Department of Chemistry, Detroit (MI), USA
E-mail address: `chernyak@chem.wayne.edu`

M. Chertkov, Los Alamos National Laboratory, Center for Nonlinear Studies, Los Alamos (NM), USA
E-mail address: `chertkov@lanl.gov`

H. J. Kappen, Radboud University, Faculty of Science, Nijmegen, The Netherlands
E-mail address: `b.kappen@science.ru.nl`