CONTRIBUTIONS TO THE THEORY OF GAUSSIAN MEASURES AND PROCESSES WITH APPLICATIONS
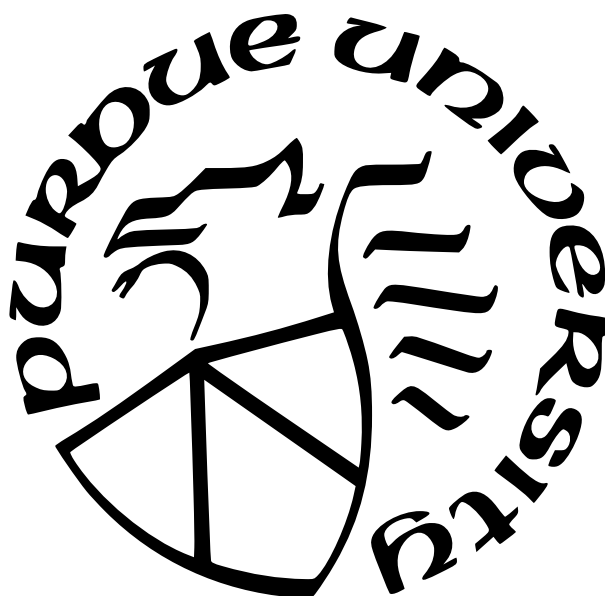
by

Zachary Selk

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Mathematics

West Lafayette, Indiana

May 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. David McReynolds, Chair

College of Science

Dr. Harsha Honnappa

School of Industrial Engineering

Dr. Nung Kwan Yip

College of Science

Dr. Kiseop Lee

College of Science

Approved by:

Dr. Plamen Stefanov

This thesis is dedicated to my parents Irene and Travis Selk.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

ABSTRACT

This thesis studies infinite dimensional Gaussian measures on Banach spaces. Let μ_0 be a centered Gaussian measure on Banach space $\mathcal{B}$, and μ^* is a measure equivalent to μ_0. We are interested in approximating, in sense of relative entropy (or KL divergence) the quantity $\frac{d\mu^z}{d\mu^*}$ where μ^z is a mean shift measure of μ_0 by an element z in the so-called "Cameron-Martin" space $\mathcal{H}_{\mu_0}$. That is, we want to find the information projection

$$\inf_{z \in \mathcal{H}_{\mu_0}} D_{KL}(\mu^z || \mu_0) = \inf_{z \in \mathcal{H}_{\mu_0}} E_{\mu^z} \left(\log \left(\frac{d\mu^z}{d\mu^*} \right) \right).$$

We relate this information projection to a mode computation, to an "open loop" control problem, and to a variational formulation leading to an Euler-Lagrange equation. Furthermore, we use this relationship to establish a kind of Feynman-Kac theorem for systems of ordinary differential equations. We demonstrate that the solution to a system of second order linear ordinary differential equations is the mode of a diffusion, analogous to the result of Feynman-Kac for parabolic partial differential equations.

1. INTRODUCTION

Infinite dimensional Gaussian measure theory uses tools from functional analysis, measure theory and probability theory and has found use in all areas of pure and applied mathematics, including quantum field theory via Feynman path integration, statistical mechanics, financial mathematics and biology to name a few. The generalization from finite dimensional measures to infinite dimensional measures poses many technical challenges; including analyzing various statistics like the mean, covariance and more recently, the mode. Similarly to the finite dimensional case, the mode of an infinite dimensional measure is an optimization problem, with all of the challenges that infinite dimensional optimization poses. The means of measures on Gaussian Banach spaces provide much information and have been used to prove interesting equivalences. For example, Feynman's originally ill-posed theory of infinite dimensional path integration which underpins much of quantum field theory was put on rigorous ground via the Feynman-Kac theorem introduced in [2] - which says that the solution to a parabolic PDE is the mean of an infinite dimensional measure on the classical Wiener space. This equivalence not only resolved one of the most fundamental problems in mathematical quantum field theory, but the equivalence spurred much generalization including to stochastic PDE in e.g. [3]–[5]. It has also seen utility in physics, see e.g. [6]–[10]. It has been used also in interacting particle systems, see e.g. [11]–[14], and in finance, see e.g. [15]. One very useful implication of the classical Feynman-Kac formula is that it leads to Monte Carlo methods for efficient numerical estimation of the solution to parabolic PDEs. As with the mean, the mode is an essential statistic.

In this thesis, we first prove a "Portmanteau" theorem that relates information projections, modes defined through Onsager-Machlup formalism, "open loop" or state independent KL weighted control and an Euler-Lagrange equation. Then we use a modified version of this theorem to include "conditional" modes to prove a Feynman-Kac type theorem for systems of ordinary differential equations.

1.1 Preliminaries

In this section we collect several classical results that are frequently referenced throughout this thesis. First, we discuss results for Gaussian measure theory on Banach spaces. We refer the reader to [16], [17] for more information on Gaussian measure theory.

We will work on a separable Banach space $(\mathcal{B}, \|\cdot\|)$, where we denote elements of $\mathcal{B}$ as $\omega \in \mathcal{B}$. We let $\mathcal{B}^*$ denote the space of continuous linear functionals $\ell : \mathcal{B} \to \mathbb{R}$ (with respect to the norm on $\mathcal{B}$).

Definition 1.1.1. *[17, Definition 3.2] Let μ be a Borel probability measure on $\mathcal{B}$.*

*(i) The **pushforward measure** $\ell^*(\mu)$ for $\ell \in \mathcal{B}^*$ is defined by $\ell^*(\mu)(A) := \mu(\ell^{-1}(A))$ for all Borel $A \subset \mathbb{R}$.*

*(ii) μ is **Gaussian** if, for all $\ell \in \mathcal{B}^*$, $\ell^*(\mu)$ is a Gaussian measure on $\mathbb{R}$.*

*(iii) μ is **centered** if, for all $\ell \in \mathcal{B}^*$, $\ell^*(\mu)$ is centered.*

Formally, we consider Dirac-δ measures to be Gaussian in the limit of infinitesimally small variance. This convention includes the case where e.g. $\ell \equiv 0$.

Definition 1.1.2. *[17, Equation (3.2)] Let μ be a centered Gaussian measure on $\mathcal{B}$. The **covariance operator** corresponding to μ is $C_\mu : \mathcal{B}^* \times \mathcal{B}^* \to \mathbb{R}$ defined by*

$$C_\mu(\ell, \ell') := \int_{\mathcal{B}} \ell(x)\ell'(x)\mu(dx),$$

for all $\ell, \ell' \in \mathcal{B}^*$.

When $\mathcal{B} = \mathbb{R}^d$, we may identify $(\mathbb{R}^d)^*$ with $\mathbb{R}^d$ so that the covariance operator from Definition 1.1.2 is just the usual covariance matrix.

Next we define the Cameron-Martin space which plays a central role in this thesis. In particular, in our upcoming information projection problem we will optimize over all shift measures corresponding to the Cameron-Martin space.

Definition 1.1.3. *[17, Definition 3.24] Let μ be a centered Gaussian measure on $\mathcal{B}$. The **Cameron-Martin space** $\mathcal{H}_\mu$ is the completion of the subspace*

$$\mathring{\mathcal{H}}_\mu := \{h \in \mathcal{B} : \exists h^* \in \mathcal{B}^* \text{ with } C_\mu(h^*, \ell) = \ell(h), \, \forall \ell \in \mathcal{B}^*\}$$

with respect to the norm $\| \cdot \|_\mu$ corresponding to μ defined by $\|h\|_\mu := \langle h, h \rangle_\mu := C_\mu(h^*, h^*)$ for all $h \in \mathcal{B}$.

Remark 1.1.1. *There might be multiple h^* associated to h in the sense of Definition 1.1.3. However, the norm $\| \cdot \|_\mu$ is independent of the choice of h^*. Furthermore, there is a canonical representation for h^* (see e.g. [17, Proposition 3.31]). In the rest of our work, we just assume h^* is this canonical representation.*

One of the main achievements of Gaussian measure theory is the Cameron-Martin theorem. For $h \in \mathcal{B}$, the translation map $T_h : \mathcal{B} \to \mathcal{B}$ is defined by $T_h(x) := x + h$ for all $x \in \mathcal{B}$. Let μ_0 be a reference Gaussian measure on $\mathcal{B}$. Corresponding to the translation map T_h, we define the pushforward measure $\mu := T_h^*(\mu_0)$ by $\mu(A) = T_h^*(\mu_0)(A) = \mu_0(T_h^{-1}(A))$ for all Borel A.

Theorem 1.1.1. *[17, Theorem 3.41] Let μ_0 be a centered Gaussian measure on $\mathcal{B}$. The pushforward measure $\mu := T_h^*(\mu_0)$ is absolutely continuous with respect to μ_0 if and only if $h \in \mathcal{H}_{\mu_0}$. Furthermore, the Radon-Nikodym derivative of μ with respect to μ_0 is*

$$\frac{d\mu}{d\mu_0} = \exp\left(h^*(x) - \frac{1}{2}\|h\|_\mu^2\right).$$

We now recall the Kullback-Leibler (KL) divergence, which is central to our information projection problem (it is the objective of this problem). Continue to let μ_0 be a reference measure on $\mathcal{B}$ and let μ be another measure on $\mathcal{B}$ such that the Radon-Nikodym derivative $\frac{d\mu}{d\mu_0}$ exists. Then, the KL-divergence of μ with respect to μ_0 is defined by:

$$D_{KL}(\mu\|\mu_0) := E_\mu\left[\log\left(\frac{d\mu}{d\mu_0}\right)\right].$$

We recall three key properties of the KL-divergence: (i) $D_{KL}(\mu\|\mu_0) \geq 0$; (ii) $D_{KL}(\mu\|\mu_0) = 0$ if and only if $\mu = \mu_0$; and (iii) $\mu \mapsto D_{KL}(\mu\|\mu_0)$ is convex. The next result shows how to compute the KL-divergence between two shift measures drawn from the Cameron-Martin space $\mathcal{H}_{\mu_0}$.

Theorem 1.1.2. *[18, Lemma 3.20] Let μ_0 be a centered Gaussian measure on $\mathcal{B}$ with Cameron-Martin space $\mathcal{H}_{\mu_0}$, and let $\mu_1 = T_{h_1}^*(\mu_0)$ and $\mu_2 = T_{h_2}^*(\mu_0)$ for some $h_1, h_2 \in \mathcal{H}_{\mu_0}$. Then, the Radon-Nikodym derivative $\frac{d\mu_1}{d\mu_2}$ exists and*

$$D_{KL}(\mu_1\|\mu_2) = \frac{1}{2}\|h_1 - h_2\|_{\mu_0}^2.$$

Next we define classical Wiener space associated to a standard Brownian motion. This is our first specific example of a Banach space with a Gaussian measure, and later in the thesis we obtain some more detailed results for this special setting.

Definition 1.1.4. *Let $\mathcal{B} = \mathcal{C}_0[0,T]$ be the set of continuous functions $f : [0,T] \to \mathbb{R}$ equipped with the supremum norm $\|\cdot\|_\infty$, such that $f(0) = 0$. Let μ_0 be the Borel measure on $\mathcal{C}_0[0,T]$ associated to the standard Brownian motion $B(t)$. Then $(\mathcal{C}_0[0,T], \mu_0)$ is **classical Wiener space**.*

Proposition 1.1.1. *[17, Exercise 3.27] Let μ_0 be the Borel measure on $\mathcal{C}_0[0,T]$ associated to a standard Brownian motion $B(t)$.*

(i) The measure μ_0 is a centered Gaussian measure with covariance operator $C_{\mu_0}(\delta_s, \delta_t) = \min\{s, t\}$.

(ii) The Cameron-Martin space associated to μ_0 is the Sobolev space

$$W_0^{1,2} := \left\{ F : [0,T] \to \mathbb{R} : f(0) = 0 \text{ and } \exists f \in L^2[0,T] \text{ so that } F(t) = \int_0^t f(s)ds \right\}.$$

Furthermore, the Cameron-Martin norm $\|\cdot\|_{\mu_0}$ is the Sobolev norm

$$\|F\|_{\mu_0} = \int_0^T f^2(s)ds.$$

Now we collect some results from stochastic analysis and Malliavin calculus. We refer the reader to [19], [20] for more details.

Theorem 1.1.3. *[21, Lemma 5.76] Let $(\mathcal{C}_0[0,T], \mu_0)$ be classical Wiener space, and assume:*

(a) F is a progressively measurable process with respect to the filtration generated by the standard Brownian motion $B(t)$;

(b) The sample paths are almost surely in the Cameron-Martin space $W_0^{1,2}$;

(c) Novikov's condition,

$$E_{\mu_0}\left[\exp\left(\int_0^T f^2(s)ds\right)\right] < \infty,$$

where $f(s) = F'(s)$, holds.

Then, the process $B(t) - F(t)$ is a standard Brownian motion under μ with density

$$\frac{d\mu}{d\mu_0} = \exp\left(\int_0^T f(s)dB(s) - \frac{1}{2}\int_0^T f^2(s)ds\right).$$

Theorem 1.1.4. *[21, Theorem 5.72] Let $(\mathcal{C}_0[0,T], \mu_0)$ be classical Wiener space, and suppose $\mu \sim \mu_0$. Then, there exists a progressively measurable process $F(t)$ with sample paths almost surely in the Cameron-Martin space $W_0^{1,2}$ so that the process $B(t) - F(t)$ is a standard Brownian motion under μ. Furthermore, the density is given by*

$$\frac{d\mu}{d\mu_0} = \exp\left(\int_0^T f(s)dB(s) - \frac{1}{2}\int_0^T f^2(s)ds\right),$$

where $f(s) = F'(s)$.

Remark 1.1.2. *Theorem 1.1.4 is a partial converse to Theorem 1.1.3. We had to assume Novikov's condition holds to prove the "forward" direction in Theorem 1.1.3. There are weaker sufficient conditions, such as Kazamaki's condition (see e.g. [22, P. 331]). However in the "reverse" direction in Theorem 1.1.4, the conclusion only establishes that the sample paths lie almost surely in the Cameron-Martin space $W_0^{1,2}$. This condition is weaker than both Novikov's condition and Kazamaki's condition.*

Remark 1.1.3. *By the Doob-Dynkin Lemma, the progressively measurable process f that appears in Theorem 1.1.3 and Theorem 1.1.4 has implicit dependence on the underlying Brownian motion, i.e., $f(s) = f(s, B(s))$ for all $s \in [0, T]$. We usually suppress this dependence for cleaner notation except where it is needed explicitly.*

In light of Theorem 1.1.4 we make the following definition.

Definition 1.1.5. *Let $\mu \sim \mu_0$ be a Borel measure on classical Wiener space $(C_0[0,T], \mu_0)$. We say that the progressively measurable process $F(t)$ **corresponds to** the measure μ if*

$$\tilde{B}(t) := B(t) - F(t)$$

is a Brownian motion under μ.

Theorem 1.1.5. *[21, Theorem 5.55] Let $(\mathcal{C}_0[0,T], \mu_0)$ be classical Wiener space, and let $C : \mathcal{C}_0[0,T] \to \mathbb{R}$ be a functional such that $E_{\mu_0}[C^2] < \infty$. Then, there exists a progressively measurable process $f(t)$ such that*

$$C = E_{\mu_0}[C] + \int_0^T f(t)dB(t).$$

Theorem 1.1.5 is an existence result. A natural follow-up question is how to compute the process $f(t)$ that appears in the statement of Theorem 1.1.5. The Clark-Ocone theorem gives a computational version of Theorem 1.1.5, stated in terms of the Malliavin derivative of C. We refer to [19] for more information on Malliavin calculus.

Theorem 1.1.6. *[19, Proposition 1.3.14] Let $C : \mathcal{C}_0[0,T] \to \mathbb{R}$ be in the Sobolev-Watanabe space $\mathbb{D}^{1,2}$ (see [19, page 27] for the definition), then*

$$C = E_{\mu_0}[C] + \int_0^T E_{\mu_0}[D_t C \mid \mathcal{F}_t]dB(t),$$

where D_t is the Malliavin derivative and $\mathcal{F}_t$ is the filtration generated by Brownian motion.

2. PORTMANTEAU THEOREM

2.1 Introduction

This chapter studies constrained information projections on Banach spaces with respect to a Gaussian reference measure, following closely the paper [1]. Specifically, our interest lies in characterizing projections of the reference measure onto sets of measures corresponding to changes in the mean (or *shift measures*). Information projection onto shift measures emerges in a number of domains of applied probability including stochastic optimal control, approximate inference, and large deviations analysis. Let $\mathcal{B}$ be a separable Banach space, μ_0 a centered Borel Gaussian measure on $\mathcal{B}$, and $\mu^* \sim \mu_0$ another Borel measure where $\sim$ denotes equivalence of measures. The symbol $\mathcal{H}_{\mu_0} \hookrightarrow \mathcal{B}$ represents the Cameron-Martin space associated to μ_0. By the Cameron-Martin Theorem [17, Theorem 3.41], the shift measure $T_h^*(\mu_0)(\cdot) := \mu_0(\cdot - h)$ for $h \in \mathcal{B}$ is absolutely continuous with respect to μ_0 if and only if $h \in \mathcal{H}_{\mu_0}$. Let $\mathcal{P}$ denote the set of all shift measures which are absolutely continuous with respect to μ_0. We are interested in computing the information projection of μ^* onto $\mathcal{P}$, $\inf_{\mu \in \mathcal{P}} D_{KL}(\mu || \mu^*)$, where $D_{KL}(\mu || \mu^*)$ is the KL-divergence of μ with respect to μ^*.

The main result of this chapter characterizing the information projection is consolidated in the following *portmanteau* theorem. For this result, we let $\mathrm{val}(\cdot)$ and $\mathrm{sol}(\cdot)$ denote the optimal value and the set of optimal solutions of an optimization problem, respectively.

Theorem 2.1.1. *[A Portmanteau Theorem] Let $\mathcal{B}$ be a separable Banach space. Let μ_0 be a Gaussian measure on $\mathcal{B}$ (see Definition 1.1.1) with Cameron-Martin space $\mathcal{H}_{\mu_0}$ (see Definition 1.1.3). Let $C : \mathcal{B} \to \mathbb{R}$ be a functional satisfying Hypothesis 2.2.2. Furthermore, assume that the functional $\Phi : \mathcal{B} \to \mathbb{R}$ defined by $\Phi(z) = E_{\mu_0}[C(\omega + z)]$ exists and satisfies Hypothesis 2.2.1. Define the probability measure μ^* on $\mathcal{B}$ with density*

$$\frac{d\mu^*}{d\mu_0} := \frac{\mathrm{e}^{-C}}{E_{\mu_0}[\mathrm{e}^{-C}]},$$

and the probability measure $\tilde{\mu}$ on $\mathcal{B}$ with density

$$\frac{d\tilde{\mu}}{d\mu_0} := \frac{\mathrm{e}^{-\Phi}}{E_{\mu_0}[\mathrm{e}^{-\Phi}]}.$$

Let $\mathcal{P}$ be the set of Gaussian shift measures which are absolutely continuous with respect to μ_0 (defined in Eq. (2.1)). Consider the following three optimization problems:

(a) (Information projection)

$$\mathbb{K} := \inf_{\mu \in \mathcal{P}} D_{KL}(\mu || \mu^*).$$

(b) (State independent KL-weighted control)

$$\mathbb{P} := \inf_{\mu \in \mathcal{P}} \left\{ E_\mu[C] + D_{KL}(\mu || \mu_0) \right\}.$$

(c) (Mode of $\tilde{\mu}$)

$$\mathbb{M} := \inf_{z \in \mathcal{H}_{\mu_0}} OM_\Phi(z).$$

Then, the optimal values val($\mathbb{K}$), val($\mathbb{P}$), and val($\mathbb{M}$) are all attained (and so all three problems have optimal solutions). By the Cameron-Martin theorem (see Theorem 1.1.1), we can identify each shift measure $\mu_z \in \mathcal{P}$ with its corresponding shift $z \in \mathcal{H}_{\mu_0}$, and so we have

$$\text{sol}(\mathbb{K}) = \text{sol}(\mathbb{P}) \equiv \text{sol}(\mathbb{M}).$$

In the above, we have used equivalence "$\equiv$" for shorthand to denote that $\mu_z \in \text{sol}(\mathbb{P})$ (or $\mu_z \in \text{sol}(\mathbb{K})$) if and only if its corresponding shift $z \in \text{sol}(\mathbb{M})$.

Furthermore, suppose $(\mathcal{B}, \mu_0)$ is the classical Wiener space (see Definition 1.1.4) and C satisfies Hypothesis 2.2.3. Let $L(t, z, \dot{z})$ be the Lagrangian (defined in Eq. (2.16)), and suppose $L(t, z, \dot{z})$ satisfies Hypothesis 2.4.1. Consider the following calculus of variations problem:

(d) (Calculus of variations)

$$\mathbb{L} : \inf_{q \in W_0^{1,2}[0,T]} \left\{ \mathcal{L}(q, \dot{q}; f) := \frac{1}{2} \int_0^T L(t, q(t), \dot{q}(t); f) dt \ \ s.t. \ \ q(0) = 0 \right\}.$$

14

Then, the optimal value $\mathrm{val}(\mathbb{L})$ *is attained (so Problem* $\mathbb{L}$ *has an optimal solution). We continue to identify each shift measure* $\mu_z \in \mathcal{P}$ *with its corresponding shift* $z \in \mathcal{H}_{\mu_0}$ *by the Cameron-Martin theorem (see Theorem 1.1.1), and so we have*

$$\mathrm{sol}(\mathbb{K}) = \mathrm{sol}(\mathbb{P}) \equiv \mathrm{sol}(\mathbb{M}) = \mathrm{sol}(\mathbb{L}).$$

Again, as above we use $\equiv$ *to denote that* $\mu_z \in \mathrm{sol}(\mathbb{P})$ *(or* $\mu_z \in \mathrm{sol}(\mathbb{K})$*) if and only if its corresponding shift* $z \in \mathrm{sol}(\mathbb{M})$ *(or* $z \in \mathrm{sol}(\mathbb{L})$*).*

In summary, the principal technical contributions of the present chapter are:

(i) For Gaussian reference measures over an arbitrary Banach space, we show that the optimal Gaussian shift measure of the (constrained) information projection problem corresponds precisely to the minimizer of the Onsager-Machlup function of an associated Gibbs measure over the Banach space (that is fully characterized). As a direct consequence, the optimal open loop KL-weighted control function can be viewed as finding the most likely path of a stochastic process associated with the Gibbs measure.

(ii) We further characterize the information projection onto Gaussian shift measures with respect to a reference Wiener measure. In particular, we derive a single functional constraint to express the feasible region $\mathcal{P}$. This functional constraint reveals that the feasible region of this optimization problem is non-convex, and we also find that standard convex inner and outer relaxations only yield "trivial" approximations.

(iii) Finally, we re-formulate the information projection problem and show that the optimal solution corresponds to a calculus of variations problem. More precisely, we show that the solution is completely characterized by the solution of an Euler-Lagrange equation. This yields a tractable representation, which we illustrate through a number of examples.

Related Literature. Information projections onto shift measures appear in a number of domains of applied probability, including stochastic optimal control, large deviations theory and approximate inference. Kullback-Liebler (KL) or relative-entropy weighted control [23]–[25] studies a class of restricted control problems where the total cost of a given control policy can be expressed as the KL-divergence between the path measure of the controlled process and that of the Wiener reference measure. Our investigation has special relevance for KL-

weighted optimal control in 'open-loop' settings where the control is not state dependent, such as in the control of large ensembles of particles, such as fleets of vehicles, modeling flocking behavior [26], material science applications [27], [28] and in power systems [29]–[31] (for load management for example).

The utility of the KL-weighted control formulation comes from a variational formula proved in [23] (and in [32] in a more restricted setting) that shows that the optimal value of the KL-weighted control is precisely the logarithm of the mean exponentiated cost under the Wiener measure. However, even though the optimal value (and corresponding optimal measure) can be characterized, computing the optimal control function itself is still a nontrivial problem. For example, [23] shows that the optimal control function can be computed as the Föllmer drift [33]–[35] corresponding to the optimizer of the information projection. In most cases, the computation of the Föllmer drift is fraught; for instance, [23] uses the Clark-Occone formula of Malliavin calculus to compute the optimal control function.

Organization. In the next section we present preliminary notation, definitions, and standard results that will be used throughout the thesis. In Section 2.2, we formalize our information projection problem and show that it is equivalent to minimization of an Onsager-Machlup function. In Section 2.2.1 we introduce the constrained relative entropy-weighted optimization problem and connect it to Onsager-Machlup minimization. We then characterize the constraint that the drift is state-independent in Section 2.3. Specifically, we show that this constraint can be expressed by a single functional constraint (which is a difference-of-convex (DC) functions), and we consider some convex relaxations. In Section 2.4 we turn to the Euler-Lagrange formulation of the problem and provide a clean derivation of the optimal solution in terms of the solution to the Euler-Lagrange equation. We also offer an interpretation of the solution to this problem in terms of minimizing the Onsager-Machlup function that was introduced in Section 2.2. Finally, in Section 2.5 we compute the solution of the Euler-Lagrange equation and the associated Onsager-Machlup minimizing process for a number of example cost functions. We conclude the chapter in Section 2.6 with some further discussion.

2.2 Information Projection and Onsager-Machlup

For this section, let $\mathcal{B}$ be a separable Banach space with centered Gaussian measure μ_0. We formally define the set of Gaussian shift measures to be

$$\mathcal{P} := \{T_h^*(\mu_0) : h \in \mathcal{H}_{\mu_0}\}, \tag{2.1}$$

where the measure $T_h^*(\mu_0)$ is defined before Theorem 1.1.1. Now let μ^* be another Borel probability measure on $\mathcal{B}$ such that $\mu^* \sim \mu_0$. Recall the information projection problem

$$\mathbb{K} := \inf_{\mu \in \mathcal{P}} D_{KL}(\mu || \mu^*).$$

Problem $\mathbb{K}$ projects μ^* onto the set of Gaussian shift measures. We are interested in finding the measure μ_O that attains the optimal value val($\mathbb{K}$) of Problem $\mathbb{K}$, when it exists.

We will show that computing the optimizer of Problem $\mathbb{K}$ is equivalent to minimizing the Onsager-Machlup function corresponding to an associated stochastic process. We now introduce the definition of the Onsager-Machlup function, see [36], [37] for further details.

Definition 2.2.1. *Let μ_0 be a Gaussian measure on a separable Banach space $\mathcal{B}$ with Cameron-Martin space $\mathcal{H}_{\mu_0}$. Let $B^\delta(z) \subset \mathcal{B}$ be the open ball of radius δ around z. Let μ be another measure that is absolutely continuous with respect to μ_0. If the limit*

$$\lim_{\delta \to 0} \frac{\mu(B^\delta(z_2))}{\mu(B^\delta(z_1))} = \exp\left(OM(z_1) - OM(z_2)\right)$$

*exists for all $z_1, z_2 \in \mathcal{H}_{\mu_0}$, then OM is called the **Onsager-Machlup** function for μ.*

The Onsager-Machlup function can be viewed as the Lagrangian for the "most likely" path of the associated process. In Theorem 2.2.2, we specifically show that the optimizer of Problem $\mathbb{K}$ can be identified with an optimal drift/shift function that is the minimizer of an associated Onsager-Machlup function. Crucially, this identification affords a Monte Carlo sampling method (via the associated stochastic process) for solving Problem $\mathbb{K}$.

In [36, Theorem 3.2], the Onsager-Machlup function is computed for a certain class of measures. We recall this result next.

17

Hypothesis 2.2.1. *Let* $\Phi : \mathcal{B} \to \mathbb{R}$ *be a functional satisfying the following conditions:*

(i) For every $\varepsilon > 0$ there is an $M \in \mathbb{R}$ such that

$$\Phi(\omega) \geq M - \varepsilon\|\omega\|^2,$$

for all $\omega \in \mathcal{B}$.

(ii) Φ is locally bounded above, i.e., for every $r > 0$ there is a $K = K(r) > 0$ such that

$$\Phi(\omega) \leq K,$$

for all $\omega \in \mathcal{B}$ with $\|\omega\| < r$.

(iii) Φ is locally Lipschitz continuous, i.e., for every $r > 0$ there exists $L = L(r) > 0$ such that

$$|\Phi(\omega_1) - \Phi(\omega_2)| \leq L\|\omega_1 - \omega_2\|,$$

for all $\omega_1, \omega_2 \in \mathcal{B}$ with $\|\omega_1\| < r$ and $\|\omega_2\| < r$.

Remark 2.2.1. *Note that the functional Φ in Hypothesis 2.2.1 is not necessarily of the form in Theorem 2.1.1, where the latter is given by some real-valued functional C.*

Theorem 2.2.1. *[36, Theorem 3.2] Let $\Phi : \mathcal{B} \to \mathbb{R}$ satisfy Hypothesis 2.2.1. Let the measure μ have the density*

$$\frac{d\mu}{d\mu_0} = \frac{\mathrm{e}^{-\Phi}}{E_{\mu_0}[\mathrm{e}^{-\Phi}]}.$$

Then, the OM function is given by

$$OM_\Phi(z) = \begin{cases} \Phi(z) + \frac{1}{2}\|z\|_{\mu_0}^2 & \text{if } z \in \mathcal{H}_{\mu_0} \\ \infty & \text{else} \end{cases}.$$

We introduce the following optimization problem

$$\mathbb{M} := \inf_{z \in \mathcal{H}_{\mu_0}} OM_\Phi(z),$$

which aims to minimize the OM function $OM_\Phi(z)$. Under Hypothesis 2.2.1, we know that an optimal solution of Problem $\mathbb{M}$ exists by [38, Proposition 3.4]. See [37] for the corresponding result in classical Wiener space. The minimizer of the OM function $OM_\Phi(z)$ (i.e., the solution of Problem $\mathbb{M}$) is the "most likely" element of $(\mathcal{B}, \mu_0)$.

To express Problem $\mathbb{K}$ in terms of the OM function, we need a technical lemma from [23] for functionals C satisfying the following "finite entropy" hypothesis.

Hypothesis 2.2.2. *We assume that $\mu_0(C < +\infty) > 0$ and $E_{\mu_0}[\exp(-C)\,|C|] < +\infty$.*

Lemma 2.2.1. *[23, Lemma 2.4(ii)] Let μ_0 be a Gaussian measure on a separable Banach space $\mathcal{B}$, and let $C : \mathcal{B} \to \mathbb{R}$ satisfy Hypothesis 2.2.2. Define the measure μ^* with density*

$$\frac{d\mu^*}{d\mu_0} = \frac{\mathrm{e}^{-C}}{E_{\mu_0}[\mathrm{e}^{-C}]}.$$

Then, for $\mu \ll \mu_0$ satisfying $D_{KL}(\mu\|\mu_0) < +\infty$ and $E_\mu[(C)^+] < +\infty$,

$$D_{KL}(\mu\|\mu^*) = E_\mu[C] + D_{KL}(\mu\|\mu_0) - \log E_{\mu_0}[\mathrm{e}^{-C}].$$

Remark 2.2.2. *In [23], it is shown that Hypothesis 2.2.2 implies $0 < E_{\mu_0}[\exp(-C)] < +\infty$, and so the measure μ^* in Lemma 2.2.1 is well defined.*

We can now relate Problem $\mathbb{K}$ to minimization of an OM function.

Theorem 2.2.2. *Let μ_0 be a centered Gaussian measure on a separable Banach space $\mathcal{B}$. Let μ^* be another Borel probability measure and assume:*

(a) $\mu^* \sim \mu_0$.

(b) μ^* has density
$$\frac{d\mu^*}{d\mu_0} = \frac{\mathrm{e}^{-C}}{E_{\mu_0}[\mathrm{e}^{-C}]},$$
for a functional $C : \mathcal{B} \to \mathbb{R}$ which satisfies Hypothesis 2.2.2.

(c) The functional $z \mapsto \Phi(z) := E_{\mu_0}[C(\omega + z)]$ satisfies Hypothesis 2.2.1, and there is an associated measure $\tilde{\mu}$ on $\mathcal{B}$ with density

$$\frac{d\tilde{\mu}}{d\mu_0} = \frac{\mathrm{e}^{-\Phi}}{E_{\mu_0}[\mathrm{e}^{-\Phi}]}.$$

19

Then, Problem $\mathbb{K}$ has a solution μ_O where $\mu_O(\cdot) = \mu_0(\cdot - z^O)$ and z^O is the minimizer of the OM function for $\tilde{\mu}$ (which always exists by [36, Prop 3.4]).

Proof. By Lemma 2.2.1, we have that

$$\inf_{\mu \in \mathcal{P}} D_{KL}(\mu||\mu^*) = \inf_{\mu \in \mathcal{P}} \{E_\mu[C] + D_{KL}(\mu||\mu_0)\} - \log E_{\mu_0}[\mathrm{e}^{-C}].$$

Using Theorem 1.1.2 and the definition of Φ for each $\mu \in \mathcal{P}$ with corresponding shift $z_\mu \in \mathcal{H}_{\mu_0}$, we have that

$$\begin{aligned}
D_{KL}(\mu||\mu^*) &= E_{\mu_0}[C(\omega + z_\mu)] + \frac{1}{2}\|z_\mu\|_{\mu_0}^2 - \log E_{\mu_0}[\mathrm{e}^{-C}] \\
&= \Phi(z_\mu) + \frac{1}{2}\|z_\mu\|_{\mu_0}^2 - \log E_{\mu_0}[\mathrm{e}^{-C}] \\
&= OM_\Phi(z_\mu) - \log E_{\mu_0}[\mathrm{e}^{-C}].
\end{aligned}$$

By [36, Prop 3.4], the minimizer of Problem $\mathbb{M}$, denoted $z^O \in \mathcal{H}_{\mu_0}$, exists. Then, we have the chain of relations

$$\begin{aligned}
OM_\Phi(z^O) - \log E_{\mu_0}[\mathrm{e}^{-C}] &= D_{KL}(\mu_0(\cdot - z^O)||\mu^*) \\
&\geq \inf_{\mu \in \mathcal{P}} D_{KL}(\mu||\mu^*) \\
&= \inf_{z \in \mathcal{H}_{\mu_0}} OM_\Phi(z) - \log E_{\mu_0}[\mathrm{e}^{-C}] \\
&= OM_\Phi(z^O) - \log E_{\mu_0}[\mathrm{e}^{-C}].
\end{aligned}$$

It follows that the optimal value $\mathrm{sol}(\mathbb{K})$ exists and also that $\mathrm{sol}(\mathbb{K}) = \mu_0(\cdot - z^O)$. $\qquad\square$

Remark 2.2.3. *Observe that z^O is interpreted as the most likely element of $(\mathcal{B}, \tilde{\mu})$. In other words, z^O can be viewed as the mode of $\tilde{\mu}$. Theorem 2.2.2 suggests a potential simulation-based method to compute the optimizer of the information projection Problem $\mathbb{K}$ via Monte-Carlo estimation of the mode of $\tilde{\mu}$.*

2.2.1 KL-Weighted Optimization on Wiener Space

We now apply our results to KL-weighted state independent (or "open loop") optimal control on classical Wiener space. Let $C : \mathcal{C}_0[0, T] \to \mathbb{R}$ be a cost functional satisfying Hypothesis 2.2.2 along with the following integrability condition.

Hypothesis 2.2.3. *We assume that $E_{\mu_0}[C^2] < \infty$.*

For all $\mu \sim \mu_0$, we define the functional:

$$J(\mu) := E_\mu[C] + D_{KL}(\mu || \mu_0) = E_\mu\left[C + \log\left(\frac{d\mu}{d\mu_0}\right)\right], \qquad (2.2)$$

which is the sum of the expected cost $E_\mu[C]$ with respect to μ and the KL-divergence $D_{KL}(\mu || \mu_0)$ of μ with resepct to the uncontrolled process μ_0. In particular, $D_{KL}(\mu || \mu_0)$ can be interpreted as a penalty for the "effort" of control.

Let

$$\mathcal{P}_0 := \{\text{Borel probability measures } \mu : \mu \sim \mu_0\} \qquad (2.3)$$

be the set of all measures on $\mathcal{B}$ that are equivalent to the Wiener measure μ_0. The corresponding classical unconstrained variational stochastic optimal control problem is:

$$\inf_{\mu \in \mathcal{P}_0} E_\mu\left[C + \log\left(\frac{d\mu}{d\mu_0}\right)\right], \qquad (2.4)$$

where we note the feasible region in Problem (2.4) is $\mathcal{P}_0$ instead of $\mathcal{P}$ (the set of shift measures defined in Eq. (2.1)). The optimality conditions of Problem (2.4) are characterized in the following theorem, where we denote an optimal solution of Problem (2.4) as μ^*.

Theorem 2.2.3. *[23, Lemma 2.4] Suppose Hypothesis 2.2.2 holds. Then, the measure with density*

$$\frac{d\mu^*}{d\mu_0} := \frac{\exp(-C)}{E_{\mu_0}[\exp(-C)]} \qquad (2.5)$$

exists and $\mu^* \in \mathcal{P}_0$. In addition, if μ is a Borel probability measure so that $E_\mu[(C)^+] < +\infty$ and $D_{KL}(\mu||\mu^*) < +\infty$, then we have:

$$J(\mu) = D_{KL}(\mu||\mu^*) - \log E_{\mu_0}[e^{-C}].\tag{2.6}$$

Furthermore, J is strictly convex on the set

$$\tilde{\mathcal{P}} := \{Borel\ probability\ measures\ \mu : E_\mu[(C)^+] < +\infty\ and\ D_{KL}(\mu||\mu^*) < +\infty\},$$

and μ^* is the unique optimal solution with corresponding optimal value:

$$\inf_{\mu \in \tilde{\mathcal{P}}} J(\mu) = -\log E_{\mu_0}\left[e^{-C}\right].$$

By Theorem 1.1.5, if C satisfies Hypothesis 2.2.3 then there exists a progressively measurable process $f(t)$ so that

$$C = E_{\mu_0}[C] + \int_0^T f(t)dB(t),$$

where equality holds almost surely. By Theorem 2.2.3, the density for the optimal measure μ^* is

$$\frac{d\mu^*}{d\mu_0} = \exp\left(-\int_0^T f(t)dB(t) - \frac{1}{2}\int_0^T f^2(t)dt\right),$$

for this same $f(t)$. Consequently, the optimal measure μ^* corresponds to (in the sense of Definition 1.1.5) an optimal drift $F(t) = -\int_0^t f(s)ds$ by Theorem 1.1.3.

As noted before, in this thesis, we are interested in minimizing $J(\mu)$ (defined in Eq. (2.2)) subject to the additional constraint that measures correspond to (in sense of Definition 1.1.5) a state independent drift. That is, $F(t)$ should be a deterministic path that does not depend on the underlying process $B(t)$. This requirement may be equivalently viewed as constraining the covariance of the controlled process to be the same as that of Brownian motion. In this light, we may express the feasible region $\mathcal{P}$ (originally defined in Eq. (2.1)) equivalently as:

$$\mathcal{P} := \{\mu \in \mathcal{P}_0 : \mu \text{ corresponds to a state independent drift}\}.\tag{2.7}$$

The corresponding constrained optimal control problem is:

$$\mathbb{P} := \inf_{\mu \in \mathcal{P}} J(\mu) = \inf_{\mu \in \mathcal{P}} \left\{ E_\mu[C] + D_{KL}(\mu || \mu_0) \right\}.$$

In the next section we provide a more precise description of the feasible region $\mathcal{P}$. In particular, we will show that while J is a strictly convex function, the feasible region $\mathcal{P}$ is not convex, and so Problem $\mathbb{P}$ is not a convex optimization problem.

Using Lemma 2.2.1, we can reformulate this constrained optimal control problem as:

$$\mathbb{P} \equiv \inf_{\mu \in \mathcal{P}} J(\mu) = \inf_{\mu \in \mathcal{P}} D_{KL}(\mu || \mu^*) + \log E_{\mu_0}[e^{-C}], \tag{2.8}$$

which reveals the familiar Problem $\mathbb{K}$. We will use the form of Eq. (2.8) to connect with Problem $\mathbb{M}$. Theorem 2.2.2 shows that the optimal solution of Problem $\mathbb{K}$ is the information projection of the optimal measure μ^* of Problem (2.4) onto $\mathcal{P}$. By Eq. (2.8), we see that the optimal solution of Problem $\mathbb{P}$ coincides with the optimal solution of Problem $\mathbb{K}$.

This information projection corresponds to finding the most likely path of an associated process $\tilde{X}$ (defined through its Onsager-Machlup function). This is philosophically similar to the equivalence of *maximum a posteriori* (MAP) inference and information projection in Euclidean spaces. For more information on this relation, see [39]. We also observe that the information projection in Problem $\mathbb{K}$ is intimately connected with MAP estimation and identifying the "mode" of the Gaussian measure on a general Banach space [36].

We now explicitly reformulate Problem $\mathbb{P}$ as an optimization problem on path space, and compute the associated process $\tilde{X}$. Consider the process $\tilde{X}(t)$ whose Girsanov density is given by

$$\frac{d\tilde{\mu}}{d\mu_0}(\omega) = \frac{1}{Z_{\tilde{\mu}}} \exp\left(-E_{\mu_0}[C(B + \omega)]\right), \tag{2.9}$$

where $Z_{\tilde{\mu}}$ is a normalizing constant. By Theorem 2.2.1, the Onsager-Machlup function associated to $\tilde{\mu}$ is

$$OM(\varphi(t)) = \begin{cases} E_{\mu_0}[C(B + \varphi)] + \frac{1}{2} \int_0^T (\dot{\varphi}(t))^2 dt & \text{if } \varphi \in W_0^{1,2}, \\ \infty & \text{if } \varphi \notin W_0^{1,2}. \end{cases} \tag{2.10}$$

This observation leads directly to the following optimization problem on path space:

$$\inf_{\varphi \in W_0^{1,2}} E_{\mu_0}[C(B + \varphi)] + \frac{1}{2} \int_0^T (\dot{\varphi}(t))^2 dt. \tag{2.11}$$

Next we use Girsanov to characterize the process $\tilde{X}(t)$. We will use $\big|_{y=x}$ to denote evaluation of y at the point x. That is, for a function g we define

$$g(y)\bigg|_{y=x} := g(x).$$

Proposition 2.2.1. *Let $C = C_0 - \int_0^T f(t, B(t)) dB(t)$ be the Itô representation of a functional $C : C_0[0, T] \to \mathbb{R}$ satisfying Hypotheses 2.2.2 and 2.2.3, where C_0 is a constant. Assume that Φ satisfies Hypothesis 2.2.1, where $\Phi(z) := E_{\mu_0}[C(\omega + z)]$. Then, the process $F(t)$ corresponding to the measure $\tilde{\mu}$ defined in Eq. (2.9) is:*

$$F(t) = \int_0^t E_{\mu_0}[f(s, B(s) + x)]\bigg|_{x=B(s)} ds. \tag{2.12}$$

Proof. Let $\varphi \in C_0[0, T]$, then

$$E_{\mu_0}[C(B + \varphi)] = C_0 - E_{\mu_0}\left[\int_0^T f(t, B(t) + \varphi(t)) dB(t) + \int_0^T f(t, B(t) + \varphi(t)) d\varphi(t)\right]$$

$$= C_0 - \int_0^T E_{\mu_0}[f(t, B(t) + \varphi(t))] d\varphi(t).$$

Therefore, the density of $\tilde{\mu}$ is

$$\frac{d\tilde{\mu}}{d\mu_0} = \frac{1}{Z_{\tilde{\mu}}} \exp\left(\int_0^T E_{\mu_0}[f(t, B(t) + x)]\bigg|_{x=B(t)} dB(t)\right), \tag{2.13}$$

where we absorbed the constant C_0 into the constant $Z_{\tilde{\mu}}$. The conclusion then follows by Theorem 1.1.3. $\square$

Remark 2.2.4. *Observe that the equivalent path space problem in Eq.* (2.11) *formally parallels the variational representation for the classical unconstrained KL-weighted control Problem* (2.4) *from [23], [32], where it is shown that*

$$- \log E_{\mu_0}[\mathrm{e}^{-C}] = \inf_{f \in A} E_{\mu_0} \left[C(B + f) + \frac{1}{2} \int_0^T f^2(t) dt \right],$$

where A is the set of all progressively measurable processes with respect to the natural filtration of the Wiener process B.

Remark 2.2.5. *Interestingly enough, our identification of the optimizer of Problem $\mathbb{P}$ with the most likely path of $\tilde{X}$ as defined in Eq.* (2.12) *(in the sense of minimizing the Onsager-Machlup function) is closely related to [40, Theorem 5]. The latter result identifies the optimal trajectory of a (deterministic) optimal control problem with the most likely path (in the Onsager-Machlup sense) of a related optimally controlled stochastic process.*

Specifically, [40] considers cost functionals of the form $C(B+f) = \int_0^T g(B(t)+f(t), f(t))dt + G(B(T))$, with cost rate $g(x,u) = \ell(x,u) + \frac{1}{2}|u|^2$. This total cost model corresponds precisely with the KL-weighted optimal control setting under consideration in this section. Our focus, however, is entirely on state independent, open loop control where the optimal control process/drift coincides with the most likely path of $\tilde{X}$.

The results in this section require the verification of Hypotheses 2.2.1 and 2.2.2, which are non-trivial to satisfy even on classical Wiener space. Consequently, in the next sections, we delve further into the Wiener space setting and develop alternative characterizations of the optimal solution of the information projection problem.

2.3 Characterization of the Feasible Region in Wiener Space

We can develop a finer characterization of Problem $\mathbb{K}$ when $(\mathcal{B}, \mu_0)$ is classical Wiener space (see Definition 1.1.4), which we present in this section. In particular, we construct a penalty function corresponding to state independent drifts. This will let us encode the feasible region $\mathcal{P}$ through a single functional constraint on the space of measures on continuous paths in $\mathcal{C}_0[0, T]$. In particular, we show that this constraint function is a "difference of

convex" (DC) functions that is zero if the measure corresponds to a state independent drift, and that is positive otherwise.

Define the function $D : \mathcal{P}_0 \to [0, \infty)$ via

$$D(\mu) := D_{KL}(\mu || \mu_0) - \frac{1}{2} \int_0^T (\partial_s E_\mu[B(s)])^2 \, ds. \tag{2.14}$$

We interpret D as a penalty function which measures violation of the requirement that the drift be state independent, and in particular we will show that $D(\mu) = 0$ if and only if $\mu \in \mathcal{P}$. The following lemma confirms that D correctly checks if μ corresponds to a state independent drift. The proof uses the idea that the variance of the drift is zero only when it is state independent.

Proposition 2.3.1. *For all $\mu \in \mathcal{P}_0$, $D(\mu) \geq 0$. Additionally, $D(\mu) = 0$ if and only if μ corresponds (in the sense of Definition 1.1.5) to a state independent drift.*

Proof. Fix a shift measure $\mu \in \mathcal{P}_0$. By Theorem 1.1.4, there exists a progressively measurable process F whose sample paths lie almost surely in the Sobolev space $W_0^{1,2}$ such that

$$\log \left(\frac{d\mu}{d\mu_0} \right) = \int_0^T f(s) \, dB(s) - \frac{1}{2} \int_0^T f^2(s) ds,$$

where $f(s) = F'(s)$. In addition, we have $B(t) = F(t) + \tilde{B}(t)$, where $\tilde{B}$ is a Brownian motion under μ. We may then compute the KL-divergence

$$\begin{aligned}
D_{KL}(\mu || \mu_0) &= E_\mu \left[\int_0^T f(s) \, dB(s) - \frac{1}{2} \int_0^T f^2(s) ds \right] \\
&= E_\mu \left[\int_0^T f(s) d(\tilde{B}(s) + F(s)) - \frac{1}{2} \int_0^T f^2(s) ds \right] \\
&= E_\mu \left[\int_0^T f(s) d\tilde{B}(s) + \frac{1}{2} \int_0^T f^2(s) ds \right].
\end{aligned}$$

An Itô integral has mean zero, and so

$$D_{KL}(\mu || \mu_0) = \frac{1}{2} E_\mu \left[\int_0^T f^2(s) \, ds \right]. \tag{2.15}$$

Substituting Eq. (2.15) into Eq. (2.14), we arrive at

$$D(\mu) = \frac{1}{2} E_\mu \left[\int_0^T f^2(s) \, ds \right] - \frac{1}{2} \int_0^t (\partial_s E_\mu[F(s)])^2 \, ds.$$

To write the second integral, we used the fact that $B(t) = F(t) + \tilde{B}(t)$, where $\tilde{B}$ is a Brownian motion under μ. By applying Fubini's theorem and differentiating under the integral sign, we obtain

$$D(\mu) = \frac{1}{2} \int_0^T \left(E_\mu[f^2(s)] - (E_\mu[f(s)])^2 \right) ds = \frac{1}{2} \int_0^T \mathrm{Var}_\mu(f(s)) ds.$$

It follows that $D(\mu) = 0$ if and only if $\mathrm{Var}_\mu(f(s)) = 0$ for all $s \in [0, T]$, i.e., F is deterministic.

$\square$

Remark 2.3.1. *Constructing the penalty function D is nontrivial. In general, given an arbitrary shift measure $\mu \in \mathcal{P}$, where $\mathcal{P}$ is defined in (2.7), it is difficult to construct the drift $F_\mu(t)$ corresponding to μ. It is even difficult to distinguish between $F(t)$ and $-F(t)$ for deterministic drifts $F \in W_0^{1,2}$. To see this, note that the Girsanov density for $F(t)$ is*

$$\frac{d\mu^{(F)}}{d\mu_0} = \exp \left(\int_0^T f(t) dB(t) - \frac{1}{2} \int_0^T f^2(t) dt \right),$$

where $f(t) = F'(t)$, while the Girsanov density for $-F(t)$ is

$$\frac{d\mu^{(-F)}}{d\mu_0} = \exp \left(\int_0^T -f(t) dB(t) - \frac{1}{2} \int_0^T f^2(t) dt \right).$$

As f is deterministic and therefore $\int_0^T f(t) dB(t) \overset{d}{=} - \int_0^T f(t) dB(t)$ in distribution, we must have that

$$\frac{d\mu^{(-F)}}{d\mu_0} \overset{d}{=} \frac{d\mu^{(F)}}{d\mu_0}.$$

It is thus difficult to "decouple" $F(t) + B(t)$ through a polarization identity.

Based on Proposition 2.3.1, we can reformulate Problem $\mathbb{K}$ as:

$$\mathbb{K} \equiv \inf_{\mu \in \mathcal{P}_0} \left\{ D_{KL}(\mu || \mu^*) : D(\mu) \leq 0 \right\},$$

where the constraint $\mu \in \mathcal{P}$ is now encoded by the single functional constraint $D(\mu) \leq 0$ and the implicit constraint $\mu \in \mathcal{P}_0$.

Problem $\mathbb{K}$ is a non-convex optimization problem because the constraint function $D(\mu)$ is non-convex. Specifically, $D(\mu)$ is a "difference of convex functions", i.e., the KL-divergence $D_{KL}(\mu||\mu_0)$ (which is convex) and the function $\frac{1}{2}\int_0^T (\partial_s E_\mu[B(s)])^2 ds$ (which is convex because the function $\partial_s E_\mu[B(s)]$ is linear in μ, and the square of a linear function is convex). So, Problem $\mathbb{K}$ has a single DC constraint (see [41], [42] for discussions on this class of optimization problems). In this situation, it is natural to consider both inner and outer convex approximations of Problem $\mathbb{K}$. However, we will now see that the feasible region $\mathcal{P}$ is pathological from the perspective of convex approximation.

First we consider an inner approximation. As the following proposition shows, the strict convex combination of measures corresponding to deterministic drifts does not correspond to a deterministic drift.

Proposition 2.3.2. *Let $\mu_1 \sim \mu_0$ and $\mu_2 \sim \mu_0$ be two measures corresponding to drifts $F_1(t)$ and $F_2(t)$ (possibly state dependent) in the sense of Definition 1.1.5. Then, for $\lambda \in (0,1)$, the convex combination $\mu := \lambda\mu_1 + (1-\lambda)\mu_2$ corresponds to a drift of*

$$F(t) := A\, F_1(t) + (1 - A)F_2(t),$$

where A is a Bernoulli random variable independent of $B(t)$ with parameter λ.

Proof. Let $B(t)$ be a Brownian motion under μ_0. Then, for each time $t \in [0,T]$ and $z \in \mathbb{R}$ we have

$$\mu\left(\{\omega \in \Omega : B(t,\omega) \leq z\}\right) = \lambda\mu_1\left(\{\omega \in \Omega : B(t,\omega) \leq z\}\right)$$
$$+ (1-\lambda)\mu_2\left(\{\omega \in \Omega : B(t,\omega) \leq z\}\right)$$
$$= \lambda\mu_1\left(\{\omega \in \Omega : F_1(t,\omega) + B^1(t,\omega) \leq z\}\right)$$
$$+ (1-\lambda)\mu_2\left(\{\omega \in \Omega : F_2(t,\omega) + B^2(t,\omega) \leq z\}\right),$$

where B^1 is a Brownian motion under μ_1 and B^2 is a Brownian motion under μ_2. Now let μ_F be the measure that corresponds to the drift of $F(t) = AF_1(t) + (1-A)F_2(t)$. By the law of total probability, for $t \in [0,T]$ and $z \in \mathbb{R}$ we have that

$$
\begin{aligned}
\mu_F\left(\{\omega \in \Omega : B(t,\omega) \leq z\}\right) &= \lambda\,\mu_F\left(\{\omega \in \Omega : B(t,\omega) \leq z\} \mid A = 1\right) \\
&\quad + (1-\lambda)\mu_F\left(\{\omega \in \Omega : B(t,\omega) \leq z\} \mid A = 0\right) \\
&= \lambda\,\mu_1\left(\{\omega \in \Omega : F_1(t,\omega) + B^1(t,\omega) \leq z\}\right) \\
&\quad + (1-\lambda)\mu_2\left(\{\omega \in \Omega : F_2(t,\omega) + B^2(t,\omega) \leq z\}\right).
\end{aligned}
$$

Then, for all $t \in [0,T]$ and $z \in \mathbb{R}$, we have the equality

$$
\mu\left(\{\omega \in \Omega : B(t,\omega) \leq z\}\right) = \mu_F\left(\{\omega \in \Omega : B(t,\omega) \leq z\}\right).
$$

$\square$

Remark 2.3.2. *Proposition 2.3.2 implies that the set of measures $\mathcal{P}$ is precisely the set of extreme points of the set $\mathcal{P}_0$. If $\mu = \lambda\mu_1 + (1-\lambda)\mu_2$ for some $\lambda \in (0,1)$ and $\mu_1, \mu_2 \in \mathcal{P}_0$, then the drift associated to μ is necessarily state dependent (and thus random). Consequently, any inner convex approximation can only consist of singletons, yielding only a trivial inner approximation to Problem $\mathbb{K}$.*

Now we consider outer approximation of $\mathcal{P}$. An obvious convex relaxation of $\mathcal{P}$ is to consider the convex hull of $\mathcal{P}$, defined as:

$$
\operatorname{conv}\mathcal{P} := \left\{ \sum_{i=1}^{n} \lambda_i \mu_i : \sum_{i=1}^{n} \lambda_i = 1,\ \lambda \geq 0,\ \{\mu_i\}_{i=1}^{n} \subset \mathcal{P},\ n \in \mathbb{N} \right\}.
$$

As a consequence of Proposition 2.3.2, we have the following characterization of $\operatorname{conv}\mathcal{P}$.

Corollary 2.3.1. *conv$\mathcal{P}$ is the set of all measures corresponding to drifts of finite* random *combinations of deterministic drifts. That is, $\mu \in conv\mathcal{P}$ if and only if its corresponding drift $F(t)$ is of the form*

$$F(t) = \begin{cases} F_1(t) & \text{with probability } p_1 \\ \qquad \dots \\ F_n(t) & \text{with probability } p_n \end{cases}$$

for some $n \in \mathbb{N}$, where F_i are deterministic drifts and $p_1 + \cdots + p_n = 1$.

It follows that the convex hull conv$\mathcal{P}$ is too large to optimize over effectively (especially since we are talking about measures on path space). In light of Proposition 2.3.2 and Corollary 2.3.1, we see that the obvious inner and outer convex approximations of Problem $\mathbb{K}$ do not yield useful approximations for computing the optimal drift (or the corresponding measure).

2.4 Euler-Lagrange Equation for Wiener Space

In response to the difficulties with the non-convex formulation of Problem $\mathbb{K}$ in Wiener space discussed in the last section, we give another reformulation of the problem in this section. In particular, we show that the solution of a certain calculus of variations problem (when it exists) corresponds to an optimal solution of Problem $\mathbb{K}$. In many cases this method is more tractable, and, as we demonstrate in several of our examples, this calculus of variations problem is often a convex optimization problem (in contrast to the formulation of Problem $\mathbb{K}$ just discussed in Section 2.3, which is intrinsically non-convex).

The objective of Problem $\mathbb{K}$ is to minimize $D_{KL}(\mu||\mu^*)$ over $\mu \in \mathcal{P}$ with respect to some μ^*. This measure μ^* satisfies $\mu^* \sim \mu_0$, so by Theorem 1.1.4 there is a progressively measurable f such that:

$$\frac{d\mu^*}{d\mu_0} = \exp\left(\int_0^T f(s, B(s))dB(s) - \frac{1}{2}\int_0^T f^2(s, B(s))ds\right).$$

For this f, we define the following Lagrangian for all $q \in W_0^{1,2}[0,T]$:

$$L(t, q(t), \dot{q}(t); f) := E_{\mu_0}\left[(\dot{q}(t) - f(t, B(t) + q(t)))^2\right]. \tag{2.16}$$

The form of this Lagrangian is not a coincidence. In our upcoming Theorem 2.4.2, we will show that it is a natural consequence of trying to minimize the KL-divergence $D_{KL}(\mu||\mu^*)$.

With this Lagrangian now in hand, we pose the corresponding calculus of variations problem:

$$\inf_{q \in W_0^{1,2}[0,T]} \left\{ \mathcal{L}(q, \dot{q}; f) := \frac{1}{2}\int_0^T L(t, q(t), \dot{q}(t); f)dt \text{ s.t. } q(0) = 0 \right\}. \tag{2.17}$$

In Problem (2.17), we note the initial condition $q(0) = 0$ at time $t = 0$, but we do not enforce any terminal condition at time $t = T$. Much of the classical theory for the calculus of variations requires explicit initial and terminal conditions (see e.g. [43]). So, we can simply consider the alternative calculus of variations problem:

$$\inf_{q \in W_0^{1,2}[0,T]} \left\{ \mathcal{L}(q, \dot{q}; f) := \frac{1}{2}\int_0^T L(t, q(t), \dot{q}(t); f)dt \text{ s.t. } q(0) = 0, \, q(T) = a \right\}, \tag{2.18}$$

with dummy terminal condition $q(T) = a$ for any $a \in \mathbb{R}$. Then, we can solve Problem (2.18) for these specific initial and terminal conditions, and later optimize over $a \in \mathbb{R}$. We will see in our examples in Section 2.5 that this issue does not present any practical difficulty.

We first discuss existence of solutions to Problem (2.17), and we refer the reader to [43] for the details of these existence results. We first require the following coercivity condition:

Hypothesis 2.4.1. *There exist $\alpha > 0$ and $\beta \geq 0$ such that $L(t, x, \dot{x}) \geq \alpha|\dot{x}|^2 - \beta$ for all $t \in [0, T]$, $x \in \mathbb{R}$, and $\dot{x} \in \mathbb{R}$.*

For the next existence result, we recall that the Lagrangian $L(t, q, \dot{q}; f)$ is automatically bounded below by zero by construction (since it is the expectation of a square), and so the optimal value of Problem (2.17) is also bounded below by zero.

Theorem 2.4.1. *[43, Theorem 4.25] Suppose $L(t, q, \dot{q}; f)$ satisfies Hypothesis 2.4.1, and that $\dot{q} \to L(t, q, \dot{q}; f)$ is convex. Then, Problem (2.17) has a solution in $W_0^{1,2}[0,T]$.*

The Euler-Lagrange equation corresponding to Problem (2.17) is:

$$L_q(t, q, \dot{q}; f) - \frac{d}{dt} L_{\dot{q}}(t, q, \dot{q}; f) = 0, \ \forall t \in [0, T]. \tag{2.19}$$

Eq. (2.19) is a common *necessary* condition for the optimal solutions of Problem (2.17). The next theorem connects the solutions of Problem (2.17) (that can be characterized by Eq. (2.19)) with the optimal solutions of Problem $\mathbb{K}$.

Theorem 2.4.2. *Let $\mu^* \sim \mu_0$ be a probability measure with density*

$$\frac{d\mu^*}{d\mu_0} = \exp\left(\int_0^T f(s, B(s))dB(s) - \frac{1}{2}\int_0^T f^2(s, B(s))ds\right).$$

Let $q \in W_0^{1,2}[0,T]$ be a solution, if it exists, to Eq. (2.19). Then, the measure defined by

$$\frac{d\mu}{d\mu_0} = \exp\left(\int_0^T \dot{q}(s)dB(s) - \frac{1}{2}\int_0^T (\dot{q}(s))^2 ds\right) \tag{2.20}$$

is an optimal solution of Problem $\mathbb{K}$.

Proof. Let

$$\frac{d\mu}{d\mu_0} = \exp\left[\int_0^T h(s)dB(s) - \frac{1}{2}\int_0^T h^2(s)ds\right]$$

for some state independent $h(s) \in W_0^{1,2}[0,T]$. We can simplify the ratio

$$\frac{d\mu}{d\mu^*} = \exp\left(\int_0^T [h(s) - f(s, B(s))]dB(s) - \frac{1}{2}\int_0^T [h^2(s) - f^2(s, B(s))]ds\right)$$

$$= \exp\left(\int_0^T [h(s) - f(s, \tilde{B}(s) + H(s))]d(\tilde{B}(s) + H(s)) - \frac{1}{2}\int_0^T [h^2(s) - f^2(s, \tilde{B}(s) + H(s))]ds\right),$$

where $\tilde{B}$ is a Brownian motion under μ and $H' = h$ with $H(0) = 0$.

Because H is differentiable, we may write that

$$\frac{d\mu}{d\mu^*} = \exp\left(\int_0^T [h(s) - f(s, \tilde{B}(s) + H(s))]d\tilde{B}(s)\right.$$

$$\left. + \frac{1}{2}\int_0^T [h^2(s) - 2h(s)f(s, \tilde{B}(s) + H(s)) + f^2(s, \tilde{B}(s) + H(s))]ds\right).$$

32

Factoring and then computing the KL-divergence as in the proof of Proposition 2.3.1 shows that

$$D_{KL}(\mu||\mu^*) = E_\mu \left[\log \left(\frac{d\mu}{d\mu^*} \right) \right] = \frac{1}{2} E_\mu \left[\int_0^T [h(s) - f(s, \tilde{B}(s) + H(s))]^2 ds \right]. \qquad (2.21)$$

Minimizing Eq. (2.21) is equivalent to the Euler-Lagrange Eq. (2.19). $\qquad \square$

To conclude this discussion, we identify some sufficient conditions for a solution of Eq. (2.19) to be an optimal solution of Problem (2.17). This depends on the following joint convexity condition.

Hypothesis 2.4.2. *The Lagrangian is jointly convex in $(q, \dot{q}) \to L(t, q, \dot{q}; f)$ for all $t \in [0, T]$.*

Theorem 2.4.3. *[43, Theorem 4.32] Suppose $L(t, q, \dot{q}; f)$ satisfies Hypothesis 2.4.1 and Hypothesis 2.4.2. If $q \in W_0^{1,2}[0, T]$ is a weak solution of Eq. (2.19), then q is an optimal solution of Problem (2.17).*

We use Theorem 2.4.3 in our examples to justify solving Problem (2.17) by solving the ODE in Eq. (2.19).

2.5 Examples

In this section, we solve several specific instances of Problem $\mathbb{P}$ where $(\mathcal{B}, \mu_0)$ is classical Wiener space (see Definition 1.1.4).

2.5.1 Case of $C = \int_0^T f(t, B(t)) dB(t)$

We start in this subsection with the case where the Itô representation $C = \int_0^T f(t, B(t)) dB(t)$ of our cost functional is already known. We may assume $E_{\mu_0}[C] = 0$ W.L.O.G. since constants do not change the optimal solution of the minimization problem.

Example 2.5.1. *Let $f(s, B(s)) = -B(s)$ so we have*

$$C = \int_0^T -B(s) dB(s).$$

First, we note that C can easily be shown to satisfy Hypotheses 2.2.2 and 2.2.3. To check that C satisfies Hypothesis 2.2.1, first we write

$$C = -\frac{1}{2}B^2(T) + \frac{1}{2}T.$$

Then, we compute Φ as

$$\Phi(z) = \frac{1}{2}E_{\mu_0}[(B(T) + z(T))^2 - T] = \frac{1}{2}z^2(T).$$

To check Hypothesis 2.2.1 Part (i), we just note that for $\varepsilon > 0$ and $z \in C_0[0, T]$, we have

$$\Phi(z) \geq 0 \geq 0 - \varepsilon\|z\|_\infty^2.$$

Thus, Part (i) is satisfied with $M = 0$. To check Hypothesis 2.2.1 Part (ii), let $r > 0$ and let $K(r) = r^2$. Then, if $\|z\|_\infty < r$ we have

$$\Phi(z) = \frac{1}{2}z^2(T) \leq \frac{1}{2}r^2 \leq K.$$

To check Hypothesis 2.2.1 Part (iii), let $r > 0$ and $L(r) = r$. For $\|z_1\|_\infty < r$ and $\|z_2\|_\infty < r$, we have

$$
\begin{aligned}
|\Phi(z_1) - \Phi(z_2)| &= \frac{1}{2}|z_1^2(T) - z_2^2(T)| \\
&= \frac{1}{2}|z_1(T) - z_2(T)||z_1(T) + z_2(T)| \\
&\leq \frac{1}{2}2r\|z_1 - z_2\|_\infty \\
&= L\|z_1 - z_2\|_\infty.
\end{aligned}
$$

Since we have verified Hypothesis 2.2.1, we know that Problem $\mathbb{M}$ has a solution. The Lagrangian L corresponding to this problem is:

$$
\begin{aligned}
L(t, q, \dot{q}) &= \frac{1}{2} E_{\mu_0} \left[(\dot{q}(t) - (B(t) + q(t)))^2 \right] \\
&= \frac{1}{2} E_{\mu_0} \left[\dot{q}^2(t) - 2\dot{q}(t)(B(t) + q(t)) + (B(t) + q(t))^2 \right] \\
&= \frac{1}{2} \left[\dot{q}^2(t) - 2q(t)\dot{q}(t) + t + q^2(t) \right] \\
&= \frac{1}{2} \left[(\dot{q}(t) - q(t))^2 \right] + \frac{1}{2} t.
\end{aligned}
$$

We see that $(q, \dot{q}) \to L(t, q, \dot{q})$ is jointly convex for all $t \in [0, T]$, and it is bounded below by zero. The functional $\mathcal{L}$ is minimized at $\dot{q}(t) = q(t)$ which yields the solution $q(t) = Ce^t$. We additionally require $q(0) = 0$, and so we set $C = 0$ to find the solution is $q(t) \equiv 0$.

To confirm the optimality of this solution using the Euler-Lagrange equation, we see the derivatives of L are:

$$
L_q(t, q, \dot{q}) = -\dot{q}(t) + q(t),
$$

and

$$
L_{\dot{q}}(t, q, \dot{q}) = \dot{q}(t) - q(t).
$$

The Euler-Lagrange equation is then

$$
-\dot{q}(t) + q(t) - \ddot{q}(t) + \dot{q}(t) = 0,
$$

which has solution $q(t) = C_1 e^t + C_2 e^{-t}$. We require $q(0) = 0$, and so we must have $C_1 = -C_2$. The minimizer of L is therefore $q(t) = C(e^t - e^{-t})$. As stated in Section 2.4, in order to get the constant C we must solve

$$
\inf_C \int_0^T L(t, q, \dot{q}) dt = \inf_C \int_0^T \left[2C^2 e^{-2t} + \frac{1}{2} t \right] dt,
$$

which is minimized when $C = 0$. Furthermore, we can compute:

$$
E_{\mu_0}[f(s, B(s) + x)] \Big|_{x = B(s)} = B(s),
$$

35

to see that the associated process is:

$$\tilde{X}(t) = \int_0^t B(s)ds + B(t).$$

Example 2.5.2. *Suppose that μ^* (the optimal unconstrained measure) corresponds to a deterministic drift $F(t)$ where $F'(s) = f(s)$ for a deterministic function f. The Lagrangian for this problem is:*

$$L(t, q, \dot{q}) = \frac{1}{2} E_{\mu_0} \left[\dot{q}^2(t) - 2\dot{q}(t)f(t) + f^2(t) \right]$$
$$= \frac{1}{2} \left[\dot{q}^2(t) - 2\dot{q}(t)f(t) + f^2(t) \right].$$

We see that $(q, \dot{q}) \to L(t, q, \dot{q})$ is jointly convex for all $t \in [0, T]$, and the derivatives of L are:

$$L_q = 0,$$

and

$$L_{\dot{q}} = \dot{q}(t) - f(t).$$

The Euler-Lagrange equation is then simply

$$-\frac{d}{dt}(\dot{q}(t) - f(t)) = 0,$$

which yields the solution $q(t) = F(t) + C_1 t + C_2$. We set $C_2 = 0$ to satisfy the initial condition $q(0) = 0$.

Furthermore, we can compute

$$E_{\mu_0}[f(s, B(s) + x)]\Big|_{x=B(s)} = \tilde{f}(s),$$

to see that the associated process $\tilde{X}(t)$ is:

$$\tilde{X}(t) = \int_0^t \tilde{f}(s)ds + B(t).$$

Example 2.5.3. *Suppose that μ^* (the optimal unconstrained measure) corresponds to the drift $\int_0^T B^2(s)ds$. The Lagrangian for this problem is then:*

$$
\begin{aligned}
L(t, q, \dot{q}) &= \frac{1}{2} E_{\mu_0}[\dot{q}^2(t) - 2(B(t) + q(t))^2 \dot{q}(t) + (B(t) + q(t))^4] \\
&= \frac{1}{2}[\dot{q}^2(t) - 2(t + q^2(t))\dot{q}(t) + q^4(t) + 6tq^2(t) + 3t^2].
\end{aligned}
$$

The derivatives of L are:

$$
L_q(t, q, \dot{q}) = \frac{1}{2}[-4q(t)\dot{q}(t) + 4q^3(t) + 12tq(t)]
$$

and

$$
L_{\dot{q}}(t, q, \dot{q}) = \frac{1}{2}[2\dot{q}(t) - 2t - 2q^2(t)].
$$

The resulting Euler-Lagrange equation is:

$$
-2q(t)\dot{q}(t) + 2q^3(t) + 6tq(t) - (\ddot{q}(t) - 1 - 2q(t)\dot{q}(t)) = 0,
$$

which simplifies to the second-order ODE

$$
2q^3(t) + 6tq(t) - \ddot{q}(t) + 1 = 0.
$$

Furthermore, we can compute

$$
E_{\mu_0}[f(s, B(s) + x)]\Big|_{x=B(s)} = s + B^2(s),
$$

to see that the process $F(t)$ corresponding to $\tilde{\mu}$ is

$$
F(t) = \int_0^t [s + B^2(s)]ds.
$$

2.5.2 Case of $C = \int_0^T g(B(t))dt + G(B(T))$

In this subsection, we consider cost functionals of the form $C = \int_0^T g(B(t))dt + G(B(T))$. First, we show that such cost functionals satisfy Hypotheses 2.2.1.

Proposition 2.5.1. *Let $C : \mathcal{C}_0[0,T] \to \mathbb{R}$ be a functional of the form $C = \int_0^T g(B(t))dt + G(B(T))$, where: (a) g is L_g-Lipschitz continuous; (b) G is L_G-Lipschitz continuous; and (c) both g and G are bounded below. Then, $\Phi(z) := E_{\mu_0}[C(\omega + z)]$ satisfies Hypothesis 2.2.1.*

Proof. First we check Hypothesis 2.2.1 Part (i). As both g and G are bounded below, there is some $M \in \mathbb{R}$ such that

$$g(x) \geq M \text{ and } G(x) \geq M,$$

for all $x \in \mathbb{R}$. Then, for any $\varepsilon > 0$ and $z \in \mathcal{C}_0[0,T]$, we have that

$$\Phi(z) = E_{\mu_0}\left[\int_0^T g(B(t) + z(t))dt + G(B(T) + z(T))\right] \geq MT + T := M' \geq M' - \varepsilon\|z\|_\infty^2.$$

Therefore, Part (i) is satisfied with $M' = MT + T$. Next, we check that Hypothesis 2.2.1 Part (ii) is satisfied. First, we note that by Lipschitz continuity we have for $t \in [0,T]$ and $z \in W_0^{1,2}$ that

$$E_{\mu_0}|g(B(t) + z(t))| \leq E_{\mu_0}|g(B(t) + z(t)) - g(0)| + |g(0)|$$
$$\leq L_g E_{\mu_0}|B(t) + z(t)| + |g(0)|$$
$$\leq L_g\sqrt{\frac{2t}{\pi}} + z(t) + |g(0)|$$
$$\leq L_g\sqrt{\frac{2T}{\pi}} + \|z\|_\infty + |g(0)|.$$

An analogous chain of inequalities holds for G. Thus, we have that

$$\Phi(z) = E_{\mu_0}\left[\int_0^T g(B(t) + z(t))dt + G(B(T) + z(T))\right]$$
$$\leq TL_g\sqrt{\frac{2T}{\pi}} + |g(0)| + L_G\sqrt{\frac{2T}{\pi}} + |G(0)| + 2\|z\|_\infty.$$

38

It follows that Φ is locally bounded above. Lastly, we need to check Hypothesis 2.2.1 Part (iii). Let $z_1, z_2 \in W_0^{1,2}$, then

$$
\begin{aligned}
|\Phi(z_1) - \Phi(z_2)| &= \left| \int_0^T E_{\mu_0}(g(B(t) + z_1(t)) - g(B(t) + z_2(t))dt \right. \\
&\quad \left. + E_{\mu_0}(G(B(T) + z_1(T)) - G(B(T) + z_2(T))) \right| \\
&\leq \int_0^T L_g |z_1(t) - z_2(t)| dt + L_G |z_1(T) - z_2(T)| \\
&\leq (L_g T + L_G) \|z_1 - z_2\|_\infty.
\end{aligned}
$$

Therefore, we find that Φ is Lipschitz, which implies that it is locally Lipschitz. $\qquad \square$

To handle this class of cost functionals in our framework, we need to first compute the Itô representation for C by using the Clark-Ocone formula, and then apply Theorem 2.4.2 (see [19] for more information on Malliavin calculus).

For the following examples, we will need to compute the Malliavin derivative of C. We use the following result on the Malliavin derivative of Lebesgue integrals to do this computation.

Lemma 2.5.1. *Let X_u be a progressively measurable process in the Sobolev-Watanabe space $\mathbb{D}^{1,2}$, then*

$$
D_t \int_0^T X_u du = \int_t^T D_t(X_u) du.
$$

Proof. First note that by the dominated convergence theorem, we have that

$$
D_t \int_0^T X_u du = \int_0^T D_t(X_u) du.
$$

Then [19, Corrolary 1.2.1] implies that $D_t(X_u) = 0$ for $u \leq t$, and hence

$$
\int_0^T D_t(X_u) du = \int_t^T D_t(X_u) du.
$$

$\qquad \square$

Note that, for the indicator function $\chi_{[0,t]}(s)$, we can rewrite:

$$g(B(t)) = g\left(\int_0^T \chi_{[0,t]}(s)dB(s)\right)$$

and

$$G(B(T)) = G\left(\int_0^T dB(s)\right).$$

Then, using the above lemma along with the definition of the Malliavin derivative, we have

$$D_tC = \int_t^T g'(B(u))du + G'(B(T)). \tag{2.22}$$

Next, in order to use the Clark-Ocone formula, we need to compute the conditional expectations $E_{\mu_0}[D_tC \mid \mathcal{F}_t]$. We prove the following technical lemma for this purpose.

Lemma 2.5.2. *Let $h : \mathbb{R} \to \mathbb{R}$ be Borel measurable and integrable. Then,*

$$E_{\mu_0}[h(B(t)) \mid \mathcal{F}_s] = E_{\mu_0}[h(N_x)]\Big|_{x=B(s)},$$

where $N_x \sim \mathcal{N}(x, t-s)$.

Proof. Let X and Y be integrable random variables, and let $\mathcal{G}$ be a $\sigma-$algebra. If X is $\mathcal{G}-$measurable and Y is independent of $\mathcal{G}$, then we have:

$$E[f(X, Y)|\mathcal{G}] = E[f(x, Y)]\Big|_{x=X}.$$

Applying this observation to h with $Y = B(t) - B(s)$, $X = B(s)$, and $f(X, Y) = h(X + Y)$ gives the desired result. $\qquad\square$

Combining the previous two lemmas with the Clark-Ocone formula, we arrive at the expression

$$C = E_{\mu_0}[C] + \int_0^T f(s, B(s))dB(s),$$

where

$$f(s, B(s)) = \int_s^T E_{\mu_0}[g'(N_x)]\Big|_{x=B(s)} du + E_{\mu_0}[G'(N_x)]\Big|_{x=B(s)}. \tag{2.23}$$

Once we find the Itô representation for $C = \int_0^T g(B(t))dt + G(B(T))$, we can use Theorem 2.4.2 to solve for the optimal deterministic drift.

We now work out some specific examples for various g and G.

Example 2.5.4. *Let $C = \int_0^T B(s)ds$. Then, the function f defined in Eq. (2.23) is specifically*

$$f(s, B(s)) = \int_s^T E_{\mu_0}[1]\Big|_{x=B(s)} du$$

$$= T - s,$$

and we have:

$$\int_0^T B(s)ds = \int_0^T (T-s)dB(s).$$

Example 2.5.5. *Let $C = \int_0^T B^2(s)ds$. Then, the function f defined in Eq. (2.23) is specifically*

$$f(s, B(s)) = \int_s^T E_{\mu_0}[2(N_x)]\Big|_{x=B(s)} du$$

$$= \int_s^T 2B(s)du$$

$$= 2(T-s)B(s),$$

and we have:

$$\int_0^T B^2(s)ds = E_{\mu_0}[C] + \int_0^T 2(T-s)B(s)dB(s). \tag{2.24}$$

Example 2.5.6. *Let $C = \int_0^T B^3(s)ds$. Then, the function f defined in Eq. (2.23) is specifically*

$$f(s, B(s)) = \int_s^T E_{\mu_0}[3(N_x)^2]\Big|_{x=B(s)} du$$

$$= 3\int_s^T [(u-s) + B(s)]du$$

$$= (T-s)^3 + 3(T-s)B(s),$$

41

and we have:

$$\int_0^T B^3(s)ds = E_{\mu_0}[C] + \int_0^T [(T-s)^3 + 3(T-s)B(s)]dB(s). \qquad (2.25)$$

In line with the previous three examples, we observe that for any polynomial g, we will have a polynomial f in Eq. (2.23).

2.6 Conclusion

This thesis is motivated by the problem of doing information projection on Banach spaces with respect to a Gaussian reference measure. We have shown that, in many cases, this kind of projection can be done effectively. In particular, we show that we can often reformulate this information projection problem as a (convex) OM-functional minimization problem and a (convex) calculus of variations problem. The OM perspective suggests a simulation-based scheme for solving the original problem, and the calculus of variations perspective suggests a discretization scheme for solving the original problem (for solving the Euler-Lagrange equation which is a general second-order ODE).

An important future direction for the results of this thesis is to use the information projection of the Wiener measure onto Gaussian shift measures, and the characterization of the optimal drift, as a computational approximation to the "full" Föllmer drift. Further problems include the consideration of more general constraint sets and other types of marginal or path constraints on the information projection problem, and characterize the solutions in those settings.

3. FEYNMAN-KAC FOR ODE

3.1 Introduction

The Feynman-Kac theorem is a widely known and extensively used formula that relates the solutions of second order parabolic partial differential equations (PDEs) to the average of a transformation of a diffusion. It was introduced in [2] and has since seen tremendous utility and generalization. The Feynman-Kac formula has been extended to stochastic PDE in e.g. [3]–[5]. It has also seen utility in physics, see e.g. [6]–[10]. It has been used also in interacting particle systems, see e.g. [11]–[14]. It has been used in finance, see e.g. [15].

There are many forms of the Feynman-Kac formula, see e.g. [44] for a standard reference. In one form, the Feynman-Kac formula says the following.

Theorem 3.1.1 (Feynman-Kac). *Consider the partial differential equation*

$$\frac{\partial}{\partial t}u(t,x) + \mu(t,x)\frac{\partial}{\partial x}u(t,x) + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2}{\partial x^2}u(t,x) = V(t,x)u(t,x),$$

on the domain $[0,T] \times \mathbb{R}$, with terminal condition $u(T,x) = \psi(x)$ and μ, σ, V sufficiently regular. Then

$$u(t,x) = E_{\mu_0}[e^{-\int_t^T V(s,X(s))ds}\psi(X(T))|X(t) = x],$$

where $X(t)$ is the solution to the solution to the stochastic differential equation

$$dX(t) = \mu(t,X(t))dt + \sigma(t,X(t))dB(t),$$

with initial condition $X(t) = x$, $B(t)$ a standard Brownian motion and μ_0 is the measure associated to $B(t)$.

Linear second order ordinary differential equations and systems thereof find extensive use in engineering, physics, biology, geometry, control theory and many other areas of pure and applied mathematics. One example of a system of such equations in physics is the matrix Airy equation, which has connections to the intersection theory on the moduli space of curves, see [45]. Further examples abound in the computation of long-run average costs associated with diffusion processes, which can be expressed as solutions of second-order

43

ordinary differential equations (ODEs). In this article, we demonstrate a theorem analogous to Theorem 3.1.1 for second order ordinary differential equations. However in the ODE case, solutions are represented as the conditional "mode" of a diffusion instead of a conditional mean. By "mode", we mean the minimizer of an Onsager-Machlup function. It was first shown in [37] that the Onsager-Machlup function is a Lagrangian for the "most likely path" or "mode" of a diffusion. Our main theorem can be summarized as follows, see Theorem 3.4.1 for a full version. This chapter closely follows the paper [46].

Theorem 3.1.2 (Feynman-Kac for ODEs). *Consider the differential equation*

$$\ddot{y}(t) + f(t)\dot{y}(t) + g(t)y(t) + h(t) = 0, \tag{3.1}$$

where the unknown $y \in C^2([0,T], \mathbb{R})$, with boundary conditions $y(0) = 0$ and $y(T) = a$. Let $f(t), g(t)$ and $h(t)$ be (almost) arbitrary C^1 real valued functions. Then there exists a choice of $C^1([0,T], \mathbb{R})$ functions $A(t)$ and $D(t)$ that depend on f, g and h so that the solution to (3.1) is

$$y(t) = \mathrm{e}^{-V(t)} \operatorname{Mode}\left[X(t)\middle| X(T) = a\right], \tag{3.2}$$

where $X(t)$ solves the stochastic differential equation $dX = -A(t)X(s)dt + dB(t)$, $B(t)$ is a standard Brownian motion and additionally $V(t) = 2^{-1}\int_0^t f(s)ds$. Here, the conditional mode of a diffusion is defined as the minimizer of the Onsager-Machlup function defined in Theorem 2.1.1 over all paths z with $z(T) = a$.

The $\mathrm{e}^{-V(t)}$ term is analogous to the Feynman-Kac for PDEs case, Theorem 3.1.1. The Feynman-Kac formula can first be proven for diffusions with no factor in front of the $u(t, x)$ term. Then including the exponential weight term, one can get more general terms in front of $u(t, x)$.

The primary tool we will use is the Onsager-Machlup formalism established in [1], [36], [37]. In particular, we will extend the Portmanteau theorem in [1], which relates information projections, Kullback-Leibler weighted control, finding the most likely path of a diffusion and Euler-Lagrange equations. In [37], it was shown that the so-called "Onsager-Machlup" function is a Lagrangian for the most likely path of a diffusion. That is, the minimizer is the

44

"mode" of the path. We show that the solution to (almost) any second order linear ODE is the (potentially weighted) conditional mode a diffusion, defined as the minimizer of the Onsager-Machlup function. Additionally, we extend our results to include systems of second order linear ODEs.

In Section 3.2, we introduce this formalism and the required stochastic analysis. In Section 3.3 we extend the Portmanteau theorem of [1] for our purposes. In Section 3.4 we prove Theorem 3.1.2. Finally in Section 3.5 we give some examples.

3.2 Preliminaries

Recall that the Onsager-Machlup function is defined through the relation

$$\lim_{\delta \to 0} \frac{\mu(B^\delta(z_2))}{\mu(B^\delta(z_1))} = \exp\left(OM(z_1) - OM(z_2)\right).$$

Remark 3.2.1. *The minimizer of the Onsager-Machlup function can be seen as the "most likely" path of the diffusion, which was shown in [37]. This fact is central to the approach in this thesis. Furthermore, we define the conditional mode to minimize OM over a region. That is, for a subset $A \subset \mathcal{H}_{\mu_0}$ and a process X whose path measure has a density*

$$\frac{d\mu}{d\mu_0} = \frac{\mathrm{e}^{-\Phi}}{E_{\mu_0}[\mathrm{e}^{-\Phi}]},$$

we define

$$\mathrm{Mode}\left[X(t)\Big|A\right] := \arg\min_{z \in A} OM_\Phi(z),$$

if the infimum is indeed achieved.

3.3 A Portmanteau Theorem

We present a slightly modified Theorem 2.1.1, a portmanteau theorem that lets us convert from Euler-Lagrange equations, to information projections, to Kullback Leibler (KL) divergence weighted control, to Onsager-Machlup functions for measures on classical Wiener space.

Theorem 3.3.1. *Let $\mathcal{C}_0$ be a classical Wiener space with Cameron-Martin space $\mathcal{H}_{\mu_0}$ (see Definition 1.1.3). Let $C : \mathcal{C}_0 \to \mathbb{R}$ be a functional so that $\mu_0(C < +\infty) > 0$ and $E_{\mu_0}[\exp(-C)\,|C|] < +\infty$. Furthermore, assume that the functional $\Phi : \mathcal{C}_0 \to \mathbb{R}$ defined by $\Phi(z) = E_{\mu_0}[C(\omega + z)]$ exists and satisfies Hypothesis 2.2.1. Define the probability measure μ^* on $\mathcal{C}_0$ with density*

$$\frac{d\mu^*}{d\mu_0} := \frac{\mathrm{e}^{-C}}{E_{\mu_0}[\mathrm{e}^{-C}]},$$

and the probability measure $\tilde{\mu}$ on $\mathcal{C}_0$ with density

$$\frac{d\tilde{\mu}}{d\mu_0} := \frac{\mathrm{e}^{-\Phi}}{E_{\mu_0}[\mathrm{e}^{-\Phi}]}.$$

Let $\mathcal{P}$ be the set of Gaussian shift measures which are absolutely continuous with respect to μ_0 (defined in Eq. (2.1)). Define $\mathcal{P}_a$ to be the subset of $\mathcal{P}$ defined in (2.1) defined by

$$\mathcal{P}_a := \{T_h^*(\mu_0) : h \in \mathcal{H}_{\mu_0}, h(T) = a\}. \tag{3.3}$$

Furthermore, suppose that C is such that that $E_{\mu_0}[C^2] < \infty$ and therefore there is a progressively measurable process $\Gamma(t, B(t))$ so that $C = \int_0^T \Gamma(t, B(t)) \cdot dB(t)$ by Itô's representation theorem. Let $L(t, z, \dot{z})$ be the Lagrangian

$$L(t, q(t), \dot{q}(t); \Gamma) := E_{\mu_0}\left[\|\dot{q}(t) - \Gamma(t, B(t) + q(t))\|^2\right]. \tag{3.4}$$

Suppose that there exist $\alpha > 0$ and $\beta \geq 0$ such that $L(t, x, \dot{x}) \geq \alpha|\dot{x}|^2 - \beta$ for all $t \in [0, T]$, $x \in \mathbb{R}^n$, and $\dot{x} \in \mathbb{R}^n$. Consider the calculus of variations problem:

Consider the following four optimization problems:

(a) (Information projection)

$$\mathbb{K} := \inf_{\mu \in \mathcal{P}_a} D_{KL}(\mu \| \mu^*).$$

(b) (State independent KL-weighted control)

$$\mathbb{P} := \inf_{\mu \in \mathcal{P}_a} \{E_\mu[C] + D_{KL}(\mu \| \mu_0)\}.$$

(c) (Conditional mode of $\tilde{\mu}$)

$$\mathbb{M} := \inf_{z \in \mathcal{H}_{\mu_0}, z(T)=a} OM_\Phi(z).$$

(d) (Calculus of variations)

$$\mathbb{L} : \inf_{z \in W_0^{1,2}[0,T]} \left\{ \mathcal{L}(z, \dot{z}; \Gamma) := \frac{1}{2} \int_0^T L(t, z(t), \dot{z}(t); \Gamma) dt \ \ s.t. \ \ z(0) = 0, z(T) = a \right\}.$$

Then, the optimal values val($\mathbb{K}$), val($\mathbb{P}$), val($\mathbb{M}$) *and* val($\mathbb{L}$) *are all attained (and so all three problems have optimal solutions). By the Cameron-Martin theorem (see Theorem 1.1.1), we can identify each shift measure $\mu_z \in \mathcal{P}$ with its corresponding shift $z \in \mathcal{H}_{\mu_0}$, and so we have*

$$\mathrm{sol}(\mathbb{K}) = \mathrm{sol}(\mathbb{P}) \equiv \mathrm{sol}(\mathbb{M}) = \mathrm{sol}(\mathbb{L}).$$

In the above, we have used equivalence "$\equiv$" for shorthand to denote that $\mu_z \in \mathrm{sol}(\mathbb{P})$ (or $\mu_z \in \mathrm{sol}(\mathbb{K})$) if and only if its corresponding shift $z \in \mathrm{sol}(\mathbb{M})$. (or $z \in \mathrm{sol}(\mathbb{L})$).

Proof. **Equivalence of (a) and (b)** Under assumptions the $\mu_0(C < +\infty) > 0$ and $E_{\mu_0}[\exp(-C)|C|] < +\infty$ the measure μ^* is well defined, see Lemma 2.4 in [47]. Then we can note that for any μ equivalent to μ^* we have that

$$\begin{aligned}
D_{KL}(\mu||\mu^*) &= E_\mu \left[\log \left(\frac{d\mu}{d\mu^*} \right) \right] \\
&= E_\mu \left[\log \left(\frac{d\mu}{d\mu_0} \right) \right] - E_\mu \left[\log \left(\frac{d\mu^*}{d\mu_0} \right) \right] \\
&= E_\mu[C] + D_{KL}(\mu||\mu_0) - \log E_{\mu_0}[e^{-C}].
\end{aligned}$$

Taking an infimum of both sides shows that problem (a) is equivalent to problem (b).

Equivalence of (a) and (d) We just have to check that for all $\mu \in \mathcal{P}$ where $\mathcal{P}$ is defined in (2.1), with shift $q \in \mathcal{H}_{\mu_0}$ we have that

$$D_{KL}(\mu||\mu^*) = \frac{1}{2} \int_0^T L(t, q(t), \dot{q}(t); \Gamma) dt.$$

47

To this aim, we note that if $\mu \in \mathcal{P}$ where $\mathcal{P}$ is defined in (2.1) with corresponding shift q that

$$\frac{d\mu}{d\mu^*} = \exp\left(\int_0^T (\dot{q} - \Gamma(t, B(t))) \cdot dB(t) - \frac{1}{2}\int_0^T \|\dot{q}\|^2 - \|\Gamma(t, B(t))\|^2 ds\right).$$

By Girsanov's theorem, Theorem 1.1.3, we have that $B(t) = \tilde{B}(t) + \dot{q}(t)$ where $\tilde{B}(t)$ is a Brownian motion under μ. Therefore the KL divergence is thus

$$D_{KL}(\mu||\mu^*) = E_\mu\left[\int_0^T (\dot{q}(t) - \Gamma(t, \tilde{B}(t) + q(t))) \cdot d(\tilde{B}(t) + q(t)) \right.$$
$$\left. - \frac{1}{2}\int_0^T (\|\dot{q}\|^2 - \|\Gamma(t, \tilde{B}(t) + q(t))\|^2)dt\right].$$

Using the mean zero property of Itô integration yields that

$$D_{KL}(\mu||\mu^*) = E_\mu\left[\int_0^T (\|\dot{q}(t)\|^2 - \Gamma(t, \tilde{B}(t) + q(t))) \cdot \dot{q}(t))dt \right.$$
$$\left. - \frac{1}{2}\int_0^T (\|\dot{q}\|^2 - \|\Gamma(t, \tilde{B}(t) + q(t))\|^2)dt\right].$$

Therefore using Fubini's theorem and a polarization identity we conclude that

$$D_{KL}(\mu||\mu^*) = \frac{1}{2}\int_0^T E_{\mu_0}\left[\|\dot{q}(t) - \Gamma(t, B(t) + q(t))\|^2\right] dt.$$

The assumptions on the Lagrangian L imply that there is a unique solution to problem (d) and hence there is a solution to problem (a).

Equivalence of (a) and (c) Using Theorem 1.1.2 and the definition of Φ for each $\mu \in \mathcal{P}$ with corresponding shift $z_\mu \in \mathcal{H}_{\mu_0}$, we have that

$$D_{KL}(\mu||\mu^*) = E_\mu[C] + D_{KL}(\mu||\mu_0) - \log E_{\mu_0}[e^{-C}]$$
$$= E_{\mu_0}[C(\omega + z_\mu)] + \frac{1}{2}\|z_\mu\|_{\mu_0}^2 - \log E_{\mu_0}[e^{-C}]$$
$$= \Phi(z_\mu) + \frac{1}{2}\|z_\mu\|_{\mu_0}^2 - \log E_{\mu_0}[e^{-C}]$$
$$= OM_\Phi(z_\mu) - \log E_{\mu_0}[e^{-C}].$$

Using the equivalence between (a) and (d), along with the assumptions on the Lagrangian L concludes.

$\square$

Remark 3.3.1. *Theorem 2.1.1 is based on Theorem 2.1.1. However, there are three key differences. First, parts (a), (b), and (c) are stated in terms of a general Gaussian measure in Theorem 2.1.1 whereas in Theorem 3.3.1 it is stated for a n-dimensional Brownian motion. Second, part (d) in Theorem 2.1.1 is only stated in terms of a 1 dimensional Brownian motion whereas in Theorem 3.3.1 it is for n dimensions. Third, in Theorem 2.1.1, everything is stated in terms of unconditional mode instead of conditional mode from Theorem 3.3.1.*

3.4 Proof of Feynman-Kac for ODEs

In this section, we restate Theorem 3.1.2 in more detail and offer a proof. For clarity, we first prove the result for a single ODE. We then extend to the case of a system of ODEs. The proof for the system is similar and included for completeness.

3.4.1 Feynman-Kac for single ODE

In this subsection we prove our Feynman-Kac result for a single linear second order ODE. First we prove the case in which $f(t) = 0$ and thus $V(t) = 0$ and we are working with no exponential weight. We then use this to construct the solution for general f. First we have the following lemma.

Lemma 3.4.1. *If* $\Gamma(t, x) = A(t)x + D(t)$ *for* $C^1([0, T], \mathbb{R}^{n \times n})$ *matrix* A *and* $C^1([0, T], \mathbb{R}^n)$ *vector* D, *then* $\mu^* = \tilde{\mu}$.

Proof. We just have to show that $C = \Phi$. Recall that $C = \int_0^T \Gamma(t, B(t)) \cdot dB(t)$. To the stated aim, we write that

$$
\begin{aligned}
\Phi(z) &= E_{\mu_0}[C(\omega + z)] \\
&= E_{\mu_0}\left[\int_0^T [A(t)(B(t) + z(t)) + D(t)] \cdot d(B(t) + z(t))\right] \\
&= \int_0^T [A(t)z(t) + D(t)] \cdot dz(t) \\
&= C(z).
\end{aligned}
$$

$\square$

Proposition 3.4.1. *Consider the differential equation*

$$
\ddot{y}(t) + g(t)y(t) + h(t) = 0, \tag{3.5}
$$

where the unknown $y \in C^2([0, T], \mathbb{R})$, with boundary conditions $y(0) = 0$ and $y(T) = a$. Let $g(t)$ and $h(t)$ be $C^1([0, T], \mathbb{R})$ functions so that the differential equations

$$
A^2(t) + \dot{A}(t) = -g(t) \tag{3.6}
$$

and

$$
\dot{D}(t) + D(t)A(t) = -h(t)
$$

have solutions, $A(t), D(t)$. Let X be the diffusion corresponding to μ^* given in Theorem 2.1.1 with $\Gamma(t, x) = A(t)x + D(t)$. Then the solution to (3.5) is

$$
y(t) = \text{Mode}\left[X(t)\middle|X(T) = a\right], \tag{3.7}
$$

Here, the conditional mode of a diffusion is defined as the minimizer of the Onsager-Machlup function defined in Theorem 2.1.1 over all paths z with $z(T) = a$, as in Remark 3.2.1.

Proof. Let $\Gamma(t,x) = A(t)x + D(t)$ for the Lagrangian defined in Theorem 2.1.1. The Lagrangian is thus

$$L(t, q, \dot{q}) = E_{\mu_0}\left[(\dot{q}(t) - A(t)(B(t) + q(t)) - D(t))^2\right]. \tag{3.8}$$

First, we note that all of the hypotheses of Theorem 2.1.1 hold as A, D are both C^1 functions by assumption. Computing the derivatives of L yields that

$$L_q(t, q, \dot{q}) = -2A(t)\left(\dot{q}(t) - A(t)q(t) - D(t)\right)$$

and

$$L_{\dot{q}}(t, q, \dot{q}) = 2\left(\dot{q}(t) - A(t)q(t) - D(t)\right).$$

The Euler Lagrange equation for L is then

$$-A(t)\left(\dot{q}(t) - A(t)q(t) - D(t)\right) - \ddot{q}(t) + \dot{A}(t)q(t) + A(t)\dot{q}(t) + \dot{D}(t) = 0.$$

We may simplify this as

$$-\ddot{q}(t) + q(t)\left(A^2(t) + \dot{A}(t)\right) + \left(\dot{D}(t) + A(t)D(t)\right) = 0. \tag{3.9}$$

By the assumption of A and D, we have that the Euler Lagrange equation for the Lagrangian L is equation (3.5). Therefore by Theorem 2.1.1 we have that the solution to equation (3.5) is the conditional mode of $\tilde{\mu}$. By Lemma 3.4.1 we have that $\mu^* = \tilde{\mu}$. Therefore the solution to equation (3.5) is the conditional mode of μ^*. $\qquad\square$

Remark 3.4.1. *The equation*

$$A^2(t) + \dot{A}(t) = -g(t) \tag{3.10}$$

is an example of a Riccati equation, whose general form is

$$\dot{A}(t) = \gamma_0(t) + \gamma_1(t)A(t) + \gamma_2(t)A^2(t),$$

where γ_i are continuous functions and $\gamma_0, \gamma_2 \neq 0$. See [48] for conditions on solving Riccati equations explicitly. To this aim, they convert the Riccati equation to a second order linear differential equation

$$\ddot{z}(t) + (-\gamma_1(t) + \dot{\gamma}_2(t)/\gamma_2(t))\dot{z}(t) + \gamma_0(t)\gamma_2(t)z(t) = 0,$$

where $A(t) = -\dot{z}(t)/(\gamma_2(t)z(t))$. In the case of Theorem 3.4.1, the Riccati equation (3.10) turns into the homogeneous equation

$$\ddot{z}(t) + g(t)z(t) = 0,$$

with $A(t) = -\dot{z}(t)/z(t)$. Note however that this is not quite the homogeneous version of (3.5) because in (3.5) we insist that $y(0) = 0$.

Explicitly giving C for an equation (3.5) comes down to the solvability of the equation (3.10). We give an example of when (3.10) is explicitly solvable as a corollary.

Corollary 3.4.1. *Consider the differential equation*

$$\ddot{y}(t) + gy(t) + h(t) = 0, \tag{3.11}$$

where the unknown $y \in C^2([0,T],\mathbb{R})$, with boundary conditions $y(0) = 0$ and $y(T) = a$. Let $h(t)$ be $C^1([0,T],\mathbb{R})$ and let $g \leq 0$ be a constant. Then the solution to (3.11) is

$$y(t) = \mathrm{Mode}\left[X(t)\middle|X(T) = a\right], \tag{3.12}$$

where the process $F(t)$ corresponding to the measure $\tilde{\mu}$ (and hence μ^*) is the process

$$F(t) = \int_0^t \left(\sqrt{-g}B(u) - \mathrm{e}^{-u\sqrt{-g}}\int_1^u \mathrm{e}^{s\sqrt{-g}}h(s)ds\right)du.$$

Remark 3.4.2. *One might ask if one can find a $\Gamma(t,x)$ that yields an Euler-Lagrange equation of the form (3.1) with $f \neq 0$. The answer is no, and the $\Gamma(t,x)$ used in Proposition 3.4.1 is of maximal generality, as the below lemma will show.*

Lemma 3.4.2. *Let $\Gamma(t,x)$ as in Portmanteau theorem be such that $\frac{\partial}{\partial x}\Gamma(t,x)$ depends non-trivially on x. Then the resulting Euler Lagrange equation has a nonlinearity in q.*

Proof. We write down the Lagrangian as

$$L(t,q,\dot{q}) = E_{\mu_0}\left[(\dot{q}(t) - \Gamma(t,B(t)+q(t)))^2\right].$$

Taking derivatives yields

$$L_q(t,q,\dot{q}) = 2E_{\mu_0}\left[(\dot{q}(t) - \Gamma(t,B(t)+q(t)))(-\Gamma_x(t,B(t)+q(t)))\right],$$

and

$$L_{\dot{q}}(t,q,\dot{q}) = 2E_{\mu_0}\left[(\dot{q}(t) - \Gamma(t,B(t)+q(t)))\right].$$

The Euler-Lagrange equation for L is thus

$$-\dot{q}(t)E_{\mu_0}[\Gamma_x(t,B(t)+q(t))] + E_{\mu_0}[\Gamma(t,B(t)+q(t))\Gamma_x(t,B(t)+q(t))]$$
$$-\ddot{q}(t) + \frac{d}{dt}E_{\mu_0}[\Gamma(t,B(t)+q(t))] = 0.$$

Because $\Gamma_x(t,x)$ depends nontrivially on x, the term

$$E_{\mu_0}[\Gamma(t,B(t)+q(t))\Gamma_x(t,B(t)+q(t))]$$

depends nonlinearly on q with no terms involving $\dot{q}$ or $\ddot{q}$. The three other terms will all involve a $\dot{q}$ or a $\ddot{q}$, so there is no chance for cancellation. $\qquad\square$

Corollary 3.4.2. *If an ODE*

$$\ddot{y}(t) + f(t)\dot{y}(t) + g(t)y(t) + h(t) = 0,$$

where $f,g,h \in C^1(0,T)$, comes from the Euler-Lagrange equation in Theorem 2.1.1, then $\Gamma(t,x) = A(t)x + C(t)$, for some $A,C \in C^1(0,T)$. Thus $f = 0$ in Proposition 3.4.1 is of maximal generality.

We finally prove our main theorem for general f by applying a simple transformation to solutions given in Proposition 3.4.1. This is the full version of Theorem 3.1.2.

Theorem 3.4.1 (Feynman-Kac for ODEs)**.** *Consider the differential equation*

$$\ddot{y}(t) + f(t)\dot{y}(t) + g(t)y(t) + h(t) = 0, \tag{3.13}$$

where the unknown $y \in C^2([0, T], \mathbb{R})$, with boundary conditions $y(0) = 0$ and $y(T) = a$. Let $f(t), g(t)$ and $h(t)$ be C^1 real valued functions. Let $V(t) = 2^{-1} \int_0^t f(s)ds$. Define the functions

$$\hat{g}(t) = -\ddot{V}(t) + (\dot{V}(t))^2 - f(t)\dot{V}(t) + g(t)$$

and

$$\hat{h}(t) = e^{V(t)}h(t).$$

Suppose that the differential equations

$$A^2(t) + \dot{A}(t) = -\hat{g}(t)$$

and

$$\dot{D}(t) + D(t)A(t) = -\hat{h}(t)$$

have solutions, $A(t), D(t)$, implying that the solution $\hat{y}(t) = \mathrm{Mode}\,[X \mid X(T) = a]$ given in Proposition 3.4.1 to the equation

$$\hat{y}''(t) + \hat{g}(t)\hat{y}(t) + \hat{h}(t) = 0$$

exists. Then the solution to (3.13) is

$$y(t) = e^{-V(t)}\hat{y}(t) = e^{-V(t)} \mathrm{Mode}\,[X \mid X(T) = a]. \tag{3.14}$$

Proof. Differentiating y once yields that

$$\dot{y}(t) = e^{-V(t)}[\hat{y}'(t) - \dot{V}(t)\hat{y}(t)].$$

Differentiating again gives

$$\ddot{y}(t) = e^{-V(t)}[\hat{y}''(t) - \ddot{V}(t)\hat{y}(t) - \dot{V}(t)\hat{y}'(t) - \dot{V}(t)\hat{y}'(t) + (\dot{V}(t))^2\hat{y}(t)].$$

Checking that y satisfies equation (3.13) concludes that

$$\begin{aligned}
\ddot{y}(t) + f(t)\dot{y}(t) + g(t)y(t) + h(t) &= e^{-V(t)}[\hat{y}''(t) + \hat{g}(t)\hat{y}(t)] + h(t) \\
&= e^{-V(t)}[-e^{V(t)}h(t)] + h(t) \\
&= 0.
\end{aligned}$$

$\square$

3.4.2 Feynman-Kac for Systems of Equations

In this subsection, we extend Theorem 3.4.1 to the case of systems of linear second order ODEs. The proofs are similar to the proofs in the previous section, but we include them for completeness.

Proposition 3.4.2. *Consider the system of differential equations*

$$\ddot{y}(t) + g(t)y(t) + h(t) = 0, \tag{3.15}$$

where the unknown $y \in C^2([0,T], \mathbb{R}^n)$, with boundary conditions $y(0) = 0$ and $y(T) = a$. Let $g(t)$ be a $C^1([0,T], \mathbb{R}^{n \times n})$ matrix and let $h(t)$ be a $C^1([0,T], \mathbb{R}^n)$ vector so that the differential equations

$$A^T(t)A(t) + \dot{A}(t) = -g(t)$$

and

$$\dot{D}(t) + A^T(t)D(t) = -h(t)$$

have solutions, $A(t), D(t)$. Let X be the diffusion corresponding to μ given in Theorem 2.1.1 with $\Gamma(t, x) = A(t)x + D(t)$. Then the solution to (3.15) is

$$y(t) = \text{Mode}\left[X(t)\Big|X(T) = a\right], \tag{3.16}$$

Here, the conditional mode of a diffusion is defined as the minimizer of the Onsager-Machlup function defined in Theorem 2.1.1 over all paths z with $z(T) = a$, as in Remark 3.2.1.

Proof. All the assumptions for Theorem 2.1.1 are satisfied so we may write the Lagrangian for Γ as

$$L(t, q, \dot{q}) = E_{\mu_0}\left[\|\dot{q}(t) - A(t)(B(t) + q(t)) - D(t)\|^2\right]. \tag{3.17}$$

Differentiating under the integral sign yields that

$$
\begin{aligned}
L_q(t, q, \dot{q}) &= -E_{\mu_0}\left[2(\dot{q}(t) - A(t)(B(t) + q(t)) - D(t))^T A(t)\right] \\
&= -2(\dot{q}(t) - A(t)q(t) - D(t))^T A(t) \\
&= -2(\dot{q}^T(t)A(t) - q^T(t)A^T(t)A(t) - D^T(t)A(t))
\end{aligned}
$$

and

$$
\begin{aligned}
L_{\dot{q}}(t, q, \dot{q}) &= E_{\mu_0}\left[2(\dot{q}(t) - A(t)(B(t) + q(t)) - D(t))^T\right] \\
&= 2(\dot{q}(t) - A(t)q(t) - D(t))^T \\
&= 2(\dot{q}^T(t) - q^T(t)A^T(t) - D^T(t)).
\end{aligned}
$$

The Euler-Lagrange equation is then

$$-\dot{q}^T(t)A(t) + q^T(t)A^T(t)A(t) + D^T(t)A(t) - \ddot{q}^T(t) + q^T\dot{A}^T(t) + \dot{q}^T(t)A^T(t) + \dot{D}^T(t) = 0.$$

Collecting terms yields that

$$-\ddot{q}^T(t) + q^T(t)(A^T(t)A(t) + \dot{A}(t)) + D^T(t)A(t) + \dot{D}^T(t) = 0. \tag{3.18}$$

56

Taking the transpose of equation (3.18) yields that

$$-\ddot{q}(t) + (A^T(t)A(t) + \dot{A}^T(t))\dot{q}(t) + A^T(t)D(t) + \dot{D}(t) = 0,$$

which is equation (3.15). $\square$

Theorem 3.4.2 (Feynman-Kac for system of ODEs)**.** *Consider the system of differential equations*

$$\ddot{y}(t) + f(t)\dot{y}(t) + g(t)y(t) + h(t) = 0, \tag{3.19}$$

where the unknown $y \in C^2([0,T], \mathbb{R}^n)$, with boundary conditions $y(0) = 0$ and $y(T) = a$. Let $f(t), g(t)$ be $C^1([0,T], \mathbb{R}^{n \times n})$ matrices and let $h(t)$ be a $C^1([0,T], \mathbb{R}^n)$ vector. Let $V(t) = 2^{-1} \int_0^t f(s)ds$. Suppose that the matrix exponentials $\mathrm{e}^{V(t)}$ and $\mathrm{e}^{-V(t)}$ exist and $\mathrm{e}^{-V(t)}$ commutes with both $f(t)$ and $g(t)$. Define the functions

$$\hat{g}(t) = -\ddot{V}(t) + (\dot{V}(t))^2 - f(t)\dot{V}(t) + g(t)$$

and

$$\hat{h}(t) = \mathrm{e}^{V(t)}h(t),$$

where $\mathrm{e}^{V(t)}$ denotes the standard exponential of a matrix. Suppose that the differential equations

$$A^T(t)A(t) + \dot{A}(t) = -\hat{g}(t) \tag{3.20}$$

and

$$\dot{D}(t) + A^T(t)D(t) = -\hat{h}(t)$$

have solutions, $A(t), D(t)$, implying that the solution $\hat{y}(t) = \mathrm{Mode}\,[X \mid X(T) = a]$ given in Proposition 3.4.2 to the equation

$$\hat{y}''(t) + \hat{g}(t)\hat{y}(t) + \hat{h}(t) = 0$$

exists. Then the solution to (3.13) *is*

$$y(t) = \mathrm{e}^{-V(t)}\hat{y}(t) = \mathrm{e}^{-V(t)}\operatorname{Mode}\left[X \mid X(T) = a\right]. \tag{3.21}$$

Proof. Differentiating y once yields that

$$\dot{y}(t) = \mathrm{e}^{-V(t)}[\hat{y}'(t) - \dot{V}(t)\hat{y}(t)].$$

Differentiating again gives

$$\ddot{y}(t) = \mathrm{e}^{-V(t)}[\hat{y}''(t) - \ddot{V}(t)\hat{y}(t) - \dot{V}(t)\hat{y}'(t) - \dot{V}(t)\hat{y}'(t) + (\dot{V}(t))^2\hat{y}(t)].$$

Checking that y satisfies equation (3.13) using the commutivity assumption concludes that

$$\begin{aligned}
\ddot{y}(t) + f(t)\dot{y}(t) + g(t)y(t) + h(t) &= \mathrm{e}^{-V(t)}[\hat{y}''(t) + \hat{g}(t)\hat{y}'(t)] + h(t) \\
&= \mathrm{e}^{-V(t)}[-\mathrm{e}^{V(t)}h(t)] + h(t) \\
&= 0.
\end{aligned}$$

$\square$

3.5   Examples

We now identify the process X for a number of example ODEs.

Example 3.5.1. *Consider the ODE*

$$\ddot{y}(t) - y(t) = 0.$$

As there is no $\dot{y}$ term, we may use Proposition 3.4.1 directly. In this case, $A(t)$ solves the first order differential equation

$$A^2(t) + \dot{A}(t) = 1.$$

Thus we may let $A = \pm 1$. As there is no h term, we may simply let $D = 0$. Therefore choosing $\Gamma(t, x) = x$ and thus

$$C = \int_0^T B(t)dB(t).$$

Thus the solution is the conditional mode of the process $X(t)$ with density

$$\frac{d\mu^*}{d\mu_0} = \frac{e^{-C}}{E_{\mu_0}[e^{-C}]}.$$

Example 3.5.2. *Consider the ODE*

$$\ddot{y}(t) + \dot{y}(t) - y(t) = 0.$$

As $f \neq 0$ we must use the full Theorem 3.4.1. We note that $\hat{g}(t) = -1$ and $\hat{h}(t) = -1$. So therefore X is the same as the previous example and we only need to include the exponential weight to get that

$$y(t) = e^{-t/2} \operatorname{Mode}[X(t) \mid X(T) = a]$$

Example 3.5.3. *Consider the ODE*

$$\ddot{y}(t) - y(t) + t = 0.$$

Then we can let $A(t) = 1$ and $D(t) = 1 - t$. Therefore $\Gamma(t, x) = x + 1 - t$. Therefore

$$C = \int_0^T [B(t) + 1 - t]dB(t).$$

Thus the solution is the conditional mode of the process $X(t)$ with density

$$\frac{d\mu^*}{d\mu_0} = \frac{e^{-C}}{E_{\mu_0}[e^{-C}]}.$$

3.6   Conclusions

We presented a new Feynman-Kac type formula for systems of linear second order ODEs, by demonstrating that the solution of the ODE is the *mode* of a specific diffusion process

that is determined by the coefficients of the ODE. To the best of our knowledge, this characterization of the solution of a given linear ODE in terms of a path functional of a stochastic process is novel and has not been identified in the literature before. This result opens the door to new ways of solving second order linear ODEs, potentially leveraging stochastic simulation, just as the Feynman-Kac theorem has led to the development of Monte Carlo methods for solving linear parabolic PDEs. We gave a few examples of the utility of our result for numerically solving ODEs. The primary bottleneck to numerically implementing this method is the calculation of A which, as noted before, is the solution to a Riccati equation which is known to be difficult to solve explicitly. There is work, of course, on numerical solutions of Riccati equations, see e.g. [49].

However, once A is identified, numerically computing the solution to the second order ODE can be done reasonably efficiently. Part (d) of our Portmanteau in Theorem 2.1.1 provides one possible way of computing the mode. Alternatively, the part (a) of the theorem suggests that the mode could be computed by solving the constrained information projection. One approach to doing this would be to use the Iterative Proportional Fitting Procedure (IPFP) or its discrete counterpart the Sinkhorn algorithm [50], for instance. There are also direct methods for computing the mode; see [51] for examples using a trapezoidal discretization.

To conclude, we pose two open questions:

Question 3.6.1. *Is there an efficient way of numerically estimating the solution without estimating A first?*

Question 3.6.2. *Can we extend our results to nonlinear ODE? That is, given a general equation*

$$\ddot{y}(t) + F(t, y, \dot{y}) = 0,$$

is there a condition on F so that we can guarantee to find a corresponding Γ?

REFERENCES

[1] Z. Selk, W. Haskell, and H. Honnappa, "Information projection on banach spaces with applications to state independent kl-weighted optimal control," *Appl. Math. Optim.*, 2021.

[2] M. Kac, "On distributions of certain Wiener functionals," *Trans. Amer. Math. Soc.*, vol. 65, pp. 1–13, 1949, ISSN: 0002-9947. DOI: 10.2307/1990512.

[3] L. Bertini and N. Cancrini, "The stochastic heat equation: Feynman-Kac formula and intermittence," *J. Statist. Phys.*, vol. 78, no. 5-6, pp. 1377–1401, 1995, ISSN: 0022-4715. DOI: 10.1007/BF02180136.

[4] Y. Hu, F. Lu, and D. Nualart, "Feynman-Kac formula for the heat equation driven by fractional noise with Hurst parameter $H < 1/2$," *Ann. Probab.*, vol. 40, no. 3, pp. 1041–1068, 2012, ISSN: 0091-1798. DOI: 10.1214/11-AOP649.

[5] Y. Hu, D. Nualart, and J. Song, "Feynman-Kac formula for heat equation driven by fractional white noise," *Ann. Probab.*, vol. 39, no. 1, pp. 291–326, 2011, ISSN: 0091-1798. DOI: 10.1214/10-AOP547.

[6] L. Accardi, "On the quantum Feynman-Kac formula," *Rend. Sem. Mat. Fis. Milano*, vol. 48, 135–180 (1980), 1978, ISSN: 0370-7377. DOI: 10.1007/BF02925568.

[7] K. L. Chung and K. M. Rao, "Feynman-Kac functional and the Schrödinger equation," in *Seminar on Stochastic Processes, 1981 (Evanston, Ill., 1981)*, ser. Progr. Prob. Statist. Vol. 1, Birkhäuser, Boston, Mass., 1981, pp. 1–29.

[8] B. Güneysu, M. Keller, and M. Schmidt, "A Feynman-Kac-Itô formula for magnetic Schrödinger operators on graphs," *Probab. Theory Related Fields*, vol. 165, no. 1-2, pp. 365–399, 2016, ISSN: 0178-8051. DOI: 10.1007/s00440-015-0633-9.

[9] J. Lőrinczi, F. Hiroshima, and V. Betz, *Feynman-Kac-type theorems and Gibbs measures on path space*, ser. De Gruyter Studies in Mathematics. Walter de Gruyter & Co., Berlin, 2011, vol. 34, pp. xii+505, With applications to rigorous quantum field theory, ISBN: 978-3-11-020148-2. DOI: 10.1515/9783110203738.

[10] Z. X. Zhao, "Green function for Schrödinger operator and conditioned Feynman-Kac gauge," *J. Math. Anal. Appl.*, vol. 116, no. 2, pp. 309–334, 1986, ISSN: 0022-247X. DOI: 10.1016/S0022-247X(86)80001-4.

[11] F. Cérou, P. Del Moral, and A. Guyader, "A nonasymptotic theorem for unnormalized Feynman-Kac particle models," *Ann. Inst. Henri Poincaré Probab. Stat.*, vol. 47, no. 3, pp. 629–649, 2011, ISSN: 0246-0203. DOI: 10.1214/10-AIHP358.

[12] P. Del Moral and L. Miclo, "Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering," in *Séminaire de Probabilités, XXXIV*, ser. Lecture Notes in Math. Vol. 1729, Springer, Berlin, 2000, pp. 1–145. DOI: 10.1007/BFb0103798.

[13] P. Del Moral, *Feynman-Kac formulae*, ser. Probability and its Applications (New York). Springer-Verlag, New York, 2004, pp. xviii+555, Genealogical and interacting particle systems with applications, ISBN: 0-387-20268-4. DOI: 10.1007/978-1-4684-9393-1.

[14] P. Del Moral, A. Doucet, and S. S. Singh, "A backward particle interpretation of Feynman-Kac formulae," *M2AN Math. Model. Numer. Anal.*, vol. 44, no. 5, pp. 947–975, 2010, ISSN: 0764-583X. DOI: 10.1051/m2an/2010048.

[15] S. Janson and J. Tysk, "Feynman-Kac formulas for Black-Scholes-type operators," *Bull. London Math. Soc.*, vol. 38, no. 2, pp. 269–282, 2006, ISSN: 0024-6093. DOI: 10.1112/S 0024609306018194.

[16] V. I. Bogachev, *Gaussian measures*, ser. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 1998, vol. 62, pp. xii+433, ISBN: 0-8218-1054-5. DOI: 10.1090/surv/062.

[17] M. Hairer, "An introduction to stochastic PDEs," *arXiv preprint arXiv:0907.4178*, 2009.

[18] Y. Dan, "Bayesian inference for Gaussian models: Inverse problems and evolution equations," 2020.

[19] D. Nualart, *The Malliavin calculus and related topics*, Second, ser. Probability and its Applications (New York). Springer-Verlag, Berlin, 2006, pp. xiv+382, ISBN: 978-3-540-28328-7; 3-540-28328-5.

[20] P. Protter, *Stochastic integration and differential equations*, ser. Applications of Mathematics (New York). Springer-Verlag, Berlin, 1990, vol. 21, pp. x+302, A new approach, ISBN: 3-540-50996-8. DOI: 10.1007/978-3-662-02619-9.

[21] F. Baudoin, *Diffusion processes and stochastic calculus*, ser. EMS Textbooks in Mathematics. European Mathematical Society (EMS), Zürich, 2014, pp. xii+276, ISBN: 978-3-03719-133-0. DOI: 10.4171/133.

[22] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, Third, ser. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1999, vol. 293, pp. xiv+602, ISBN: 3-540-64325-7. DOI: 10.1007/978-3-662-06400-9.

[23] J. Bierkens and H. J. Kappen, "Explicit solution of relative entropy weighted control," *Systems & Control Letters*, vol. 72, pp. 36–43, 2014.

[24] E. Todorov, "Linearly-solvable Markov decision problems," in *Advances in Neural Information Processing Systems*, 2007, pp. 1369–1376.

[25] E. Todorov, "Efficient computation of optimal actions," *Proceedings of the National Academy of Sciences*, vol. 106, no. 28, pp. 11 478–11 483, 2009.

[26] R. Brockett, "Notes on the control of the liouville equation," in *Control of partial differential equations*, Springer, 2012, pp. 101–129.

[27] S. Whitelam and I. Tamblyn, "Learning to grow: Control of material self-assembly using evolutionary reinforcement learning," *Physical Review E*, vol. 101, no. 5, p. 052 604, 2020.

[28] X. Tang, J. Zhang, M. A. Bevan, and M. A. Grover, "A comparison of open-loop and closed-loop strategies in colloidal self-assembly," *Journal of Process Control*, vol. 60, pp. 141–151, 2017.

[29] M. Chertkov and V. Chernyak, "Ensemble of thermostatically controlled loads: Statistical physics approach," *Scientific reports*, vol. 7, no. 1, pp. 1–9, 2017.

[30] M. Chertkov, V. Y. Chernyak, and D. Deka, "Ensemble control of cycling energy loads: Markov decision approach," in *Energy Markets and Responsive Grids*, Springer, 2018, pp. 363–382.

[31] D. Métivier and M. Chertkov, "Mean-field control for efficient mixing of energy loads," *Physical Review E*, vol. 101, no. 2, p. 022 115, 2020.

[32] M. Boué, P. Dupuis, *et al.*, "A variational representation for certain functionals of Brownian motion," *The Annals of Probability*, vol. 26, no. 4, pp. 1641–1659, 1998.

[33] H. Föllmer, "An entropy approach to the time reversal of diffusion processes," in *Stochastic Differential Systems Filtering and Control*, Springer, 1985, pp. 156–163.

[34] H. Föllmer, "Time reversal on Wiener space," in *Stochastic Processes—Mathematics and Physics*, Springer, 1986, pp. 119–129.

[35] J. Lehec, "Representation formula for the entropy and functional inequalities," in *Annales de l'IHP Probabilités et statistiques*, vol. 49, 2013, pp. 885–899.

[36] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss, "MAP estimators and their consistency in Bayesian nonparametric inverse problems," *Inverse Problems*, vol. 29, no. 9, pp. 095017, 27, 2013, ISSN: 0266-5611. DOI: 10.1088/0266-5611/29/9/095017.

[37] D. Dürr and A. Bach, "The Onsager-Machlup function as Lagrangian for the most probable path of a diffusion process," *Communications in Mathematical Physics*, vol. 60, no. 2, pp. 153–170, 1978.

[38] M. Dashti, K. J. Law, A. M. Stuart, and J. Voss, "MAP estimators and their consistency in Bayesian nonparametric inverse problems," *Inverse Problems*, vol. 29, no. 9, p. 095 017, 2013.

[39] J. M. Joyce, "Kullback-leibler divergence," in *International Encyclopedia of Statistical Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 720–722, ISBN: 978-3-642-04898-2. DOI: 10.1007/978-3-642-04898-2_327.

[40] E. Todorov, "Finding the most likely trajectories of optimally-controlled stochastic systems," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 4728–4734, 2011.

[41] P. D. Tao and L. T. H. An, "Convex analysis approach to DC programming: Theory, algorithms and applications," *Acta mathematica vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.

[42] P. D. Tao *et al.*, "The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems," *Annals of Operations Research*, vol. 133, no. 1-4, pp. 23–46, 2005.

[43] J. Calder, "The calculus of variations," 2020.

[44] B. Øksendal, *Stochastic differential equations*, Sixth, ser. Universitext. Springer-Verlag, Berlin, 2003, pp. xxiv+360, An introduction with applications, ISBN: 3-540-04758-1. DOI: 10.1007/978-3-642-14394-6.

[45] M. Kontsevich, "Intersection theory on the moduli space of curves and the matrix Airy function," *Comm. Math. Phys.*, vol. 147, no. 1, pp. 1–23, 1992, ISSN: 0010-3616.

[46] Z. Selk and H. Honnappa, "A feynman-kac type theorem for odes: Solutions of second order odes as modes of diffusions," *arXiv preprint arXix:2106.08525*, 2021.

[47] J. Bierkens and H. J. Kappen, "Explicit solution of relative entropy weighted control," *Systems Control Lett.*, vol. 72, pp. 36–43, 2014, ISSN: 0167-6911. DOI: 10.1016/j.sysconle.2014.08.001.

[48] A. Al Bastami, M. R. Belić, and N. Z. Petrović, "Special solutions of the Riccati equation with applications to the Gross-Pitaevskii nonlinear PDE," *Electron. J. Differential Equations*, No. 66, 10, 2010.

[49] J. Biazar and M. Didgar, "Numerical solution of Riccati equations by the Adomian and asymptotic decomposition methods over extended domains," *Int. J. Differ. Equ.*, Art. ID 580741, 7, 2015, ISSN: 1687-9643. DOI: 10.1155/2015/580741.

[50] S. Di Marino and A. Gerolin, "An optimal transport approach for the schrödinger bridge problem and convergence of sinkhorn algorithm," *Journal of Scientific Computing*, vol. 85, no. 2, pp. 1–28, 2020.

[51] D. A. Dutra, B. O. S. Teixeira, and L. A. Aguirre, "Maximum *a posteriori* state path estimation: Discretization limits and their interpretation," *Automatica J. IFAC*, vol. 50, no. 5, pp. 1360–1368, 2014, ISSN: 0005-1098. DOI: 10.1016/j.automatica.2014.03.003.