

On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint

Yongxin Chen, Tryphon Georgiou and Michele Pavon

Abstract

We take a new look at the relation between the optimal transport problem and the Schrödinger bridge problem from the stochastic control perspective. We show that the connections are richer and deeper than described in existing literature. In particular: a) We give an elementary derivation of the Benamou-Brenier fluid dynamics version of the optimal transport problem; b) We provide a new fluid dynamics version of the Schrödinger bridge problem; c) We observe that the latter provides an important connection with optimal transport without zero noise limits; d) We propose and solve a fluid dynamic version of optimal transport with *prior*; e) We can then view optimal transport with prior as the zero noise limit of Schrödinger bridges when the prior is any Markovian evolution. In particular, we work out the Gaussian case. A numerical example of the latter convergence involving Brownian particles is also provided.

Index Terms

Optimal transport problem, Schrödinger bridge, stochastic control, zero noise limit.

I. INTRODUCTION

We discuss two problems of very different beginning. Optimal mass transport (OMT) originates in the work of Gaspar Monge in 1781 [34] and seeks a transport plan that corresponds in an optimal way two distributions of equal total mass. The cost penalizes the distance that mass is transported to ensure exact

Y. Chen and T. Georgiou are with the Department of Electrical and Computer Engineering, University of Minnesota, 200 Union street S.E. Minneapolis, Minnesota MN 55455, U.S.A. chen2468@umn.edu, tryphon@umn.edu

Michele Pavon is with the Dipartimento di Matematica, Università di Padova, via Trieste 63, 35121 Padova, Italy pavon@math.unipd.it

correspondence. Likewise, data for Erwin Schrödinger's 1931/32 bridge problem [44], [45] are again two distributions of equal total mass, in fact, probability distributions. Here however, these represent densities of diffusive particles at two points in time and the problem seeks the most likely path that establishes a correspondence between the two. A rich relationship between the two problems emerges in the case where the transport cost is quadratic in the distance, and in fact, the problem of OMT emerges as the limit of Schrödinger bridges as the diffusivity tends to zero. The parallel treatment of both problems highlights the time-symmetry of both problems and points of contact between stochastic optimal control and information theoretic concepts.

Historically, the modern formulation of optimal mass transport is due to Leonid Kantorovich [27] and the subject has been the focus of renewed and increased interest because of its relevance in a wide range of fields including economics, physics, engineering, and probability [43], [47], [48]. In fact, Kantorovich's contributions and their impact to resource allocation was recognized with the Nobel Prize in Economics in 1975 while in the past twenty years contributions by Ambrosio, Benamou, Brenier, McCann, Cullen, Gangbo, Kinderlehrer, Lott, Otto, Rachev, Rüschemdorf, Tannenbaum, Villani, and many others have launched a new fast developing phase, see e.g., [19], [38], [3], [1], [47], [48], [37]. On the other hand, the Schrödinger bridge problem [44], [45], has been the subject of strong but intermittent interest by mostly probabilists, physicists, and quantum theorists. Early important contributions were due to Fortet, Beurling, Jamison and Föllmer [18], [4], [24], [16], see [49] for a survey. Renewed interest was sparked in the past twenty years after a close relationship to stochastic control was recognized [10], [11], [42] and a similarly fast developing phase is underway, see the semi-expository paper [29] and [33], [41], [30], [20] for other recent contributions.

Besides the intrinsic importance of optimal mass transport to the geometry of spaces and the multitude of applications, a significant impetus for some recent work has been the need for effective computation [3], [2] which is often challenging. Likewise, excepting special cases [13], [14], the computation of the optimal stochastic control for the Schrödinger bridge problem is challenging, as it amounts to two partial differential equations nonlinearly coupled through their boundary values [49]. Only very recently implementable forms have become available for corresponding linear stochastic systems [5], [6], [8] and for versions of the problem involving Markov chains and Kraus maps of statistical quantum mechanics [20]; see also [9] which deals with the Schrödinger bridge problem with finite or infinite horizon for a system of nonlinear stochastic oscillators.

The aim of the present paper is to elucidate some of the connections between optimal mass transport and Schrödinger bridges thereby extending both theories. We follow in the footsteps of Léonard [30], [29], who investigated their relation, and of Mikami and Thieullen [31], [32], [33] who employed stochastic control and Schrödinger bridges to solve the optimal transport problem. This paper may then be seen to complement the results in these papers by providing a unifying view of the relationship between these two problems via optimal control. In particular, we give an elementary derivation of the Benamou-Brenier fluid dynamics version of the Monge-Kantorovich problem. We also provide a time-symmetric fluid dynamic version of the Schrödinger bridge problem different from [29, Section 4]; it underscores that an important connection with optimal transport exists even without zero noise limits. We then formulate and solve a fluid dynamic version of the optimal transport problem with *prior*. This allows us to study zero noise limits of Schrödinger bridges when the prior is any Markovian evolution. In particular, employing our results of [6], we study the case when the prior evolution is a Gauss-Markov process.

The outline of the paper is as follows: In Section II, we derive the Benamou-Brenier version of the OMT problem. In Section III, we provide some background on the classical Schrödinger bridge problem. Section IV is devoted to characterizing the optimal forward and backward drift in the bridge problem. In Section V, we give a control time-symmetric formulation of the Schrödinger bridge problem. This leads, in the following Section VI, to a new fluid dynamic formulation of the bridge problem. Section VII is dedicated to the optimal mass transfer problem with prior. In Section VIII, we investigate the zero noise limit when the prior is Gaussian. The paper concludes with two examples. In Section IX, we discuss the zero noise limit when the prior is Wiener measure and the goal is shifting the mean of a normal distribution. Finally, in Section X, we provide a numerical two-dimensional example of overdamped Brownian particles. In the zero noise limit, we obtain the solution of the corresponding OMT problem with prior.

II. OPTIMAL MASS TRANSPORT AS A STOCHASTIC CONTROL PROBLEM

A. The Monge-Kantorovich problem

Given two distributions μ, ν on \mathbb{R}^n having equal total mass, the original formulation due to G. Monge sought to identify a transport (measurable) map T from $\mathbb{R}^n \rightarrow \mathbb{R}^n$ so that the push-forward $T\#\mu$ is equal to ν , in the sense that $\nu(\cdot) = \mu(T^{-1}(\cdot))$, while the cost of transportation $\int c(x, T(x))\mu(dx)$ is

minimal. Here, $c(x, y)$ represents the transference cost from point x to point y and for the purposes of the present it will be $c(x, y) = \frac{1}{2}\|x - y\|^2$.

The dependence of the cost of transportation on T is highly nonlinear which complicated early analyses of the problem. Thus, it was not until Kantorovich's relaxed formulation in 1942 that the Monge's problem received a definitive solution. In this, instead of the transport map one seeks a joint distribution $\Pi(\mu, \nu)$ on the product space $\mathbb{R}^n \times \mathbb{R}^n$, referred to as a "coupling" between μ and ν , so that the marginals along the two coordinate directions coincide with μ and ν respectively. Thence, one seeks to determine

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{2} \|x - y\|^2 d\pi(x, y) \quad (1)$$

In case an optimal transport map exists, the optimal coupling has support on the graph of this map, see [47]. Herein, we consider this relaxed Kantorovich formulation. We wish first to give next an elementary derivation of the fact that Problem 1 can be turned into a stochastic control problem as stated in [33, formula (1.6)] and then, to derive an alternative "fluid-dynamic" formulation due to Benamou-Brenier. We strive for clarity rather than generality. In particular, we (tacitly) assume throughout the paper that μ does not give mass to sets of dimension $\leq n - 1$. Then, by Brenier's theorem [47], there exists a unique optimal transport plan (Kantorovich) induced by a map (Monge) which is the gradient of a convex function.

B. A stochastic control formulation

As customary, let us start by observing that

$$\frac{1}{2} \|x - y\|^2 = \inf_{x \in \mathcal{X}_{xy}} \int_0^1 \frac{1}{2} \|\dot{x}\|^2 dt \quad (2)$$

where \mathcal{X}_{xy} is the family of $C^1([0, 1]; \mathbb{R}^n)$ paths with $x(0) = x$ and $x(1) = y$. Let

$$x^*(t) = (1 - t)x + ty$$

be the solution of (2), namely the straight line joining x and y . Since $x^*(t)$ is a Euclidean geodesic, any probabilistic average of the lengths of C^1 trajectories starting at x at time 0 and ending in y at time 1 gives necessarily a higher value. Thus, the probability measure on $C^1([0, 1]; \mathbb{R}^n)$ concentrated on the path $\{x^*(t); 0 \leq t \leq 1\}$ solves the following problem

$$\inf_{P_{xy} \in \mathbb{D}^1(\delta_x, \delta_y)} \mathbb{E}_{P_{xy}} \left\{ \int_0^1 \frac{1}{2} \|\dot{x}\|^2 dt \right\}, \quad (3)$$

where $\mathbb{D}^1(\delta_x, \delta_y)$ are the probability measures on $C^1([0, 1]; \mathbb{R}^n)$ whose initial and final one-time marginals are Dirac's deltas concentrated at x and y , respectively. Since (3) provides us with yet another representation for $\frac{1}{2}\|x - y\|^2$, in view of (1), we also get that

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{2} \|x - y\|^2 d\pi(x, y) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \inf_{P_{xy} \in \mathbb{D}^1(\delta_x, \delta_y)} \mathbb{E}_{P_{xy}} \left\{ \int_0^1 \frac{1}{2} \|\dot{x}\|^2 dt \right\} d\pi(x, y) \quad (4)$$

Now observe that if $P_{xy} \in \mathbb{D}^1(\delta_x, \delta_y)$ and $\pi \in \Pi(\mu, \nu)$ then

$$P = \int_{\mathbb{R}^n \times \mathbb{R}^n} P_{xy} d\pi(x, y)$$

is a probability measure in $\mathbb{D}^1(\mu, \nu)$, namely a measure on $C^1([0, 1]; \mathbb{R}^n)$ whose one-time marginal at t_0 and t_1 are specified to be μ and ν , respectively. Conversely, the disintegration of any measure $P \in \mathbb{D}^1(\mu, \nu)$ with respect to the initial and final positions yields $P_{xy} \in \mathbb{D}^1(\delta_x, \delta_y)$ and $\pi \in \Pi(\mu, \nu)$.

Thus we get that the original optimal transport problem is equivalent to

$$\inf_{P \in \mathbb{D}^1(\mu, \nu)} \mathbb{E}_P \left\{ \int_0^1 \frac{1}{2} \|\dot{x}\|^2 dt \right\}. \quad (5)$$

So far, we have followed [29, pp. 2-3]. Instead of the ‘‘particle’’ picture, we can also consider the hydrodynamic version of (2), namely the optimal control problem

$$\frac{1}{2} \|x - y\|^2 = \inf_{v \in \mathcal{V}_y} \int_0^1 \frac{1}{2} \|v(x^v(t), t)\|^2 dt \quad (6)$$

$$\dot{x}^v(t) = v(x^v(t), t), \quad x(0) = x,$$

where the admissible feedback control laws $v(\cdot, \cdot)$ in \mathcal{V}_y are continuous and such that $x^v(1) = y$.

Following the same steps as before, we get that the optimal transport problem is equivalent to the following stochastic control problem with atypical boundary constraints

$$\inf_{v \in \mathcal{V}} \mathbb{E} \left\{ \int_0^1 \frac{1}{2} \|v(x^v(t), t)\|^2 dt \right\} \quad (7a)$$

$$x^v(t) = v(x^v(t), t), \quad \text{a.s.}, \quad x(0) \sim \mu, \quad x(1) \sim \nu. \quad (7b)$$

Finally suppose $d\mu(x) = \rho_0(x)dx$, $d\nu(y) = \rho_1(y)dy$ and $x^v(t) \sim \rho(t, x)dx$. Then, necessarily, ρ satisfies (weakly) the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0 \quad (8)$$

expressing the conservation of probability mass. Moreover,

$$\mathbb{E} \left\{ \int_0^1 \frac{1}{2} \|v(x^v(t), t)\|^2 dt \right\} = \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2} \|v(x, t)\|^2 \rho(t, x) dt dx.$$

Hence (7) turns into the celebrated ‘‘fluid-dynamic’’ version of the optimal transport problem due to Benamou and Brenier [3]:

$$\inf_{(\rho, v)} \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2} \|v(x, t)\|^2 \rho(t, x) dt dx, \quad (9a)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \quad (9b)$$

$$\rho(0, x) = \rho_0(x), \quad \rho(1, y) = \rho_1(y). \quad (9c)$$

The variational analysis for (7) or, equivalently, for (9) can be carried out in many different ways. For instance, let $\mathcal{P}_{\rho_0\rho_1}$ be the family of flows of probability densities $\rho = \{\rho(\cdot, t); 0 \leq t \leq 1\}$ satisfying (9c) and let \mathcal{V} be the family of continuous feedback control laws $v(\cdot, \cdot)$. Consider the unconstrained minimization of the Lagrangian over $\mathcal{P}_{\rho_0\rho_1} \times \mathcal{V}$

$$\mathcal{L}(\rho, v) = \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2} \|v(x, t)\|^2 \rho(t, x) + \lambda(x, t) \left(\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) \right) \right] dt dx, \quad (10)$$

where λ is a C^1 Lagrange multiplier. Integrating by parts, assuming that limits for $x \rightarrow \infty$ are zero, we get

$$\int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2} \|v(x, t)\|^2 + \left(-\frac{\partial \lambda}{\partial t} - \nabla \lambda \cdot v \right) \right] \rho(x, t) dt dx + \int_{\mathbb{R}^n} [\lambda(x, 1)\rho_1(x) - \lambda(x, 0)\rho_0(x)] dx. \quad (11)$$

The last integral is constant over $\mathcal{P}_{\rho_0\rho_1}$ and can therefore be discarded. We are left to minimize

$$\int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2} \|v(x, t)\|^2 + \left(-\frac{\partial \lambda}{\partial t} - \nabla \lambda \cdot v \right) \right] \rho(x, t) dt dx \quad (12)$$

over $\mathcal{P}_{\rho_0\rho_1} \times \mathcal{V}$. We consider doing this in two stages, starting from minimization with respect to v for a fixed flow of probability densities $\rho = \{\rho(\cdot, t); 0 \leq t \leq 1\}$ in $\mathcal{P}_{\rho_0\rho_1}$. Pointwise minimization of the integrand at each time $t \in [0, 1]$ gives that

$$v_\rho^*(x, t) = \nabla \lambda(x, t) \quad (13)$$

which is continuous. Plugging this form of the optimal control into (12), we get

$$J(\rho) = - \int_{\mathbb{R}^n} \int_0^1 \left[\frac{\partial \lambda}{\partial t} + \frac{1}{2} \|\nabla \lambda\|^2 \right] \rho(x, t) dt dx. \quad (14)$$

In view of this, if λ satisfies the Hamilton-Jacobi equation

$$\frac{\partial \lambda}{\partial t} + \frac{1}{2} \|\nabla \lambda\|^2 = 0, \quad (15)$$

then $J(\rho)$ is identically zero over $\mathcal{P}_{\rho_0 \rho_1}$ and any $\rho \in \mathcal{P}_{\rho_0 \rho_1}$ minimizes the Lagrangian (10) together with the feedback control (13). We have therefore established the following [3]:

Proposition 2.1: Let $\rho^*(x, t)$ with $t \in [0, 1]$ and $x \in \mathbb{R}^n$, satisfy

$$\frac{\partial \rho^*}{\partial t} + \nabla \cdot (\nabla \psi \rho^*) = 0, \quad \rho^*(x, 0) = \rho_0(x), \quad (16)$$

where ψ is a solution of the Hamilton-Jacobi equation

$$\frac{\partial \psi}{\partial t} + \frac{1}{2} \|\nabla \psi\|^2 = 0 \quad (17)$$

for some boundary condition $\psi(x, 1) = \psi_1(x)$. If $\rho^*(x, 1) = \rho_1(x)$, then the pair (ρ^*, v^*) with $v^*(x, t) = \nabla \psi(x, t)$ is a solution of (9).

The stochastic nature of the Benamou-Brenier formulation (9) stems from the fact that initial and final densities are specified. Accordingly, the above requires solving a two-point boundary value problem and the resulting control dictates the local velocity field. In general, one cannot expect to have a classical solution of (17) and has to be content with a viscosity solution [17]. See [46] for a recent contribution in the case when only samples of ρ_0 and ρ_1 are known.

III. BACKGROUND ON SCHRÖDINGER BRIDGES

A. Finite energy diffusions

We follow [24], [16], [49]. Let $\Omega := C([0, 1], \mathbb{R}^n)$ denote the family of n -dimensional continuous functions, let W_x denote Wiener measure on Ω starting at x at $t = 0$, and let

$$W := \int W_x dx$$

be stationary Wiener measure. Let \mathbb{D} be the family of distributions on Ω that are equivalent to W . By Girsanov's theorem, under $Q \in \mathbb{D}$, the coordinate process $x(t, \omega) = \omega(t)$ admits the representations

$$dx(t) = \beta_+^Q dt + dw_+(t), \quad \beta_+^Q \text{ is } \mathcal{F}_t^+ \text{ - adapted}, \quad (18)$$

$$dx(t) = \beta_-^Q dt + dw_-(t), \quad \beta_-^Q \text{ is } \mathcal{F}_t^- \text{ - adapted}, \quad (19)$$

where \mathcal{F}_t^+ and \mathcal{F}_t^- are σ -algebras of events observable up to time t and from time t on, respectively, and w_-, w_+ are standard n -dimensional Wiener processes, [15]. Moreover,

$$Q \left[\int_0^1 \|\beta_+^Q\|^2 dt < \infty \right] = Q \left[\int_0^1 \|\beta_-^Q\|^2 dt < \infty \right] = 1.$$

For $Q, P \in \mathbb{D}$, we define the *relative entropy* $H(Q, P)$ of Q with respect to P as

$$H(Q, P) = E_Q \left[\ln \frac{dQ}{dP} \right].$$

It then follows from Girsanov's theorem that

$$H(Q, P) = H(q_0, p_0) + E_Q \left[\int_0^1 \frac{1}{2} \|\beta_+^Q - \beta_+^P\|^2 dt \right] \quad (20a)$$

$$= H(q_1, p_1) + E_Q \left[\int_0^1 \frac{1}{2} \|\beta_-^Q - \beta_-^P\|^2 dt \right]. \quad (20b)$$

Here q_0, q_1 are the marginal densities of Q at 0 and 1, respectively. Similarly, p_0, p_1 are the marginal densities of P . Then, β_+^Q and β_-^Q are the forward and the backward drifts of Q , respectively, and similarly for P . The sketch of the proof goes as follows:

$$\begin{aligned} \frac{dQ}{dW} &= q_0(x(0)) \exp \left[\int_0^1 \beta_+^Q dx - \int_0^1 \frac{1}{2} \|\beta_+^Q\|^2 dt \right], \quad Q \text{ a.s.}, \\ \frac{dW}{dP} &= \frac{1}{p_0(x(0))} \exp \left[- \int_0^1 \beta_+^P dx + \int_0^1 \frac{1}{2} \|\beta_+^P\|^2 dt \right], \quad P \text{ a.s.} \Rightarrow Q \text{ a.s..} \end{aligned} \quad (21)$$

Hence,

$$\begin{aligned} \ln \frac{dQ}{dP} &= \ln \frac{q_0(x(0))}{p_0(x(0))} + \int_0^1 (\beta_+^Q - \beta_+^P) dx + \int_0^1 \frac{1}{2} (\|\beta_+^P\|^2 - \|\beta_+^Q\|^2) dt, \quad Q \text{ a.s.}, \\ &= \ln \frac{q_0(x(0))}{p_0(x(0))} + \int_0^1 (\beta_+^Q - \beta_+^P) dW_t^+ + \int_0^1 \frac{1}{2} \|\beta_+^P - \beta_+^Q\|^2 dt, \quad Q \text{ a.s..} \end{aligned} \quad (22)$$

Taking E_Q on both sides, one gets (20a) provided the stochastic integral has zero expectation. In general,

$$\int_0^1 (\beta_+^Q - \beta_+^P) dw_+(t)$$

is only a local, \mathcal{F}_t^+ - martingale. In order to claim that it has zero expectation one needs to "localize" [28, p.36]. Similarly, one can show (20b).

B. The Schrödinger bridge problem

Now let ρ_0 and ρ_1 be two everywhere positive probability densities. Let $\mathbb{D}(\rho_0, \rho_1)$ denote the set of distributions in \mathbb{D} having the prescribed marginal densities at 0 and 1. Given $P \in \mathbb{D}$, we consider the following problem:

$$\text{Minimize } H(Q, P) \text{ over } \mathbb{D}(\rho_0, \rho_1). \quad (23)$$

If there is at least one Q in $\mathbb{D}(\rho_0, \rho_1)$ such that $H(Q, P) < \infty$, there exists a unique minimizer Q^* in $\mathbb{D}(\rho_0, \rho_1)$ called *the Schrödinger bridge* from ρ_0 to ρ_1 over P . Indeed, let

$$P_x^y = P[\cdot \mid x(0) = x, x(1) = y], \quad Q_x^y = Q[\cdot \mid x(0) = x, x(1) = y]$$

be the disintegrations of P and Q with respect to the initial and final positions. Let also

$$\mu^P = P[(x(0), x(1)) \in (\cdot)], \quad \mu^Q = Q[(x(0), x(1)) \in (\cdot)]$$

be the joint initial-final time distributions under P and Q , respectively. Then, we have

$$P = \int P_x^y(\cdot) \mu^P(dx, dy), \quad Q = \int Q_x^y(\cdot) \mu^Q(dx, dy).$$

By the multiplication formula,

$$\frac{dQ}{dP} = \frac{d\mu^Q}{d\mu^P}(x(0), x(1)) \frac{dQ_{x(0)}^{x(1)}}{dP_{x(0)}^{x(1)}}, \quad Q \text{ a.s.},$$

we get

$$\begin{aligned} H(Q, P) &= E_Q \left[\ln \frac{dQ}{dP} \right] = E_Q \left[\ln \frac{d\mu^Q}{d\mu^P}(x(0), x(1)) \right] + E_Q \left[\ln \frac{dQ_{x(0)}^{x(1)}}{dP_{x(0)}^{x(1)}}(x) \right] = \\ &= \int \left(\ln \frac{d\mu^Q}{d\mu^P} \right) d\mu^Q + \int \int \left(\ln \frac{dQ_x^y}{dP_x^y} \right) dQ_x^y \mu^Q(dx, dy). \end{aligned} \quad (24)$$

This is the sum of two nonnegative quantities. The second becomes zero if and only if

$$Q_x^y = P_x^y, \quad \mu^Q \text{ a.s.}$$

Thus, as already observed by Schrödinger, the problem reduces to minimizing

$$\int \left(\ln \frac{d\mu^Q}{d\mu^P} \right) d\mu^Q \quad (25)$$

subject to the (linear) constraints

$$\mu^Q(dx \times \mathbb{R}^n) = \rho_0(x)dx, \quad \mu^Q(\mathbb{R}^n \times dy) = \rho_1(y)dy. \quad (26)$$

By the results of Beurlin-Jamison-Föllmer, this problem has a unique solution μ^* and

$$Q^* = \int P_x^y(\cdot) \mu^*(dx, dy)$$

solves (23).

IV. A STOCHASTIC CONTROL FORMULATION

Consider now the case where (the coordinate process under) P is a Markovian diffusion with forward drift field $b_+^P(x, t)$ and transition density $p(\sigma, x, \tau, y)$. The one-time density $\rho(x, t)$ of P is a weak solution of the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (b_+^P \rho) - \frac{1}{2} \Delta \rho = 0. \quad (27)$$

Moreover, forward and backward drifts are related through Nelson's relation [36]

$$b_-^P(x, t) = b_+^P(x, t) - \nabla \ln \rho(x, t). \quad (28)$$

Then Q^* is also Markovian with forward drift field

$$b_+^{Q^*}(x, t) = b_+^P(x, t) + \nabla \ln \varphi(x, t), \quad (29)$$

where the (everywhere positive) function φ solves together with another function $\hat{\varphi}$ the system

$$\varphi(t, x) = \int p(t, x, 1, y) \varphi(1, y) dy, \quad (30)$$

$$\hat{\varphi}(t, x) = \int p(0, y, t, x) \hat{\varphi}(0, y) dy. \quad (31)$$

with boundary conditions

$$\varphi(x, 0) \hat{\varphi}(x, 0) = \rho_0(x), \quad \varphi(x, 1) \hat{\varphi}(x, 1) = \rho_1(x).$$

Moreover, the one-time density $\tilde{\rho}$ of Q^* satisfies the factorization

$$\tilde{\rho}(x, t) = \varphi(x, t) \hat{\varphi}(x, t), \forall t \in [0, 1]. \quad (32)$$

Let us give an elementary derivation of (29). Let $\varphi(x, t)$ be any positive, space-time harmonic function, namely φ satisfies on $\mathbb{R}^n \times [0, 1]$

$$\frac{\partial \varphi}{\partial t} + b_+^P \cdot \nabla \varphi + \frac{1}{2} \Delta \varphi = 0. \quad (33)$$

It follows that $\ln \varphi$ satisfies

$$\frac{\partial \ln \varphi}{\partial t} + b_+^P \cdot \nabla \ln \varphi + \frac{1}{2} \Delta \ln \varphi = -\frac{1}{2} \|\nabla \ln \varphi\|^2. \quad (34)$$

Observe now that, in view of (20a), problem (23) is equivalent to minimizing over $\mathbb{D}(\rho_0, \rho_1)$ the functional

$$I(Q) = E_Q \left[\int_0^1 \frac{1}{2} \|\beta_+^Q - b_+^P(x(t), t)\|^2 dt - \ln \varphi(x(1), 1) + \ln \varphi(x(0), 0) \right]. \quad (35)$$

This follows from the fact that $H(Q, P)$ and (35) differ by a quantity which is constant over $\mathbb{D}(\rho_0, \rho_1)$.

Observe now that, under Q , by Ito's rule,

$$d \ln \varphi(x(t), t) = \left[\frac{\partial \ln \varphi}{\partial t}(x(t), t) + \beta_+^Q \cdot \nabla \ln \varphi(x(t), t) + \frac{1}{2} \Delta \ln \varphi(x(t), t) \right] dt + \nabla \ln \varphi(x(t), t) dw_+(t). \quad (36)$$

Using this and (34) in (35), we now get

$$\begin{aligned} I(Q) &= E_Q \left[\int_0^1 \frac{1}{2} \|\beta_+^Q - b_+^P(x(t), t)\|^2 dt - \ln \varphi(x(1), 1) + \ln \varphi(x(0), 0) \right] \\ &= E_Q \left[\int_0^1 \left(\frac{1}{2} \|\beta_+^Q - b_+^P(x(t), t)\|^2 - \left[\frac{\partial \ln \varphi}{\partial t} + \beta_+^Q \cdot \nabla \ln \varphi + \frac{1}{2} \Delta \ln \varphi \right] (x(t), t) \right) dt \right. \\ &\quad \left. - \int_0^1 \nabla \ln \varphi(x(t), t) dw_+(t) \right] \\ &= E_Q \left[\int_0^1 \left(\frac{1}{2} \|\beta_+^Q - b_+^P(x(t), t)\|^2 - \left(\beta_+^Q - b_+^P(x(t), t) \right) \cdot \nabla \ln \varphi(x(t), t) + \frac{1}{2} \|\nabla \ln \varphi(x(t), t)\|^2 \right) dt \right] \\ &= E_Q \left[\int_0^1 \frac{1}{2} \|\beta_+^Q - b_+^P(x(t), t) - \nabla \ln \varphi(x(t), t)\|^2 dt \right], \end{aligned} \quad (37)$$

where again we assumed that the stochastic integral has zero expectation. Then the form (29) of the forward drift of Q^* follows. Define now

$$\hat{\varphi}(x, t) = \frac{\tilde{\rho}(x, t)}{\varphi(x, t)}.$$

Then a direct calculation using (33), and the Fokker-Planck equation satisfied by $\tilde{\rho}$

$$\frac{\partial \tilde{\rho}}{\partial t} + \nabla \cdot ((b_+^P + \nabla \ln \varphi) \tilde{\rho}) - \frac{1}{2} \Delta \tilde{\rho} = 0, \quad (38)$$

yields

$$\frac{\partial \hat{\varphi}}{\partial t} + \nabla \cdot (b_+^P \hat{\varphi}) - \frac{1}{2} \Delta \hat{\varphi} = 0. \quad (39)$$

Thus, $\hat{\varphi}$ is co-harmonic, namely it satisfies the original Fokker-Planck equation (27) just like $\rho(x, t)$, the one-time density of the ‘‘prior’’ P .

Suppose we start instead with $\psi(x, t)$, a positive, reverse-time space-time harmonic function, namely ψ satisfies on $\mathbb{R}^n \times [0, 1]$

$$\frac{\partial \psi}{\partial t} + b_-^P \cdot \nabla \psi - \frac{1}{2} \Delta \psi = 0, \quad (40)$$

where $b_-^P(x) = b_+^P(x) - \nabla \ln \rho(x, t)$ is the backward drift of P . Then $\ln \psi$ satisfies

$$\frac{\partial \ln \psi}{\partial t} + b_-^P \cdot \nabla \ln \psi - \frac{1}{2} \Delta \ln \psi = \frac{1}{2} \|\nabla \ln \psi\|^2. \quad (41)$$

Consider now the functional

$$\bar{I}(Q) = E_Q \left[\int_0^1 \frac{1}{2} \|\beta_-^Q - b_-^P(x(t), t)\|^2 dt + \ln \psi(x(1), 1) - \ln \psi(x(0), 0) \right]. \quad (42)$$

Again, minimizing $\bar{I}(Q)$ over $\mathbb{D}(\rho_0, \rho_1)$ is equivalent to (23). By Ito’s rule, under Q , we have

$$d \ln \psi(x(t), t) = \left[\frac{\partial \ln \psi}{\partial t} + \beta_-^Q \cdot \nabla \ln \psi + \frac{1}{2} \Delta \ln \psi \right] (x(t), t) dt + \nabla \ln \psi(x(t), t) dw_-(t). \quad (43)$$

Using this differential in (42), we now get

$$\begin{aligned} \bar{I}(Q) &= E_Q \left[\int_0^1 \frac{1}{2} \|\beta_-^Q - b_-^P(x(t), t)\|^2 dt + \ln \psi(x(1), 1) - \ln \psi(x(0), 0) \right] \\ &= E_Q \left[\int_0^1 \left(\frac{1}{2} \|\beta_-^Q - b_-^P(x(t), t)\|^2 + \left[\frac{\partial \ln \psi}{\partial t} + \beta_-^Q \cdot \nabla \ln \psi - \frac{1}{2} \Delta \ln \psi \right] (x(t), t) dt \right. \right. \\ &\quad \left. \left. + \int_0^1 \nabla \ln \psi(x(t), t) dw_-(t) \right) \right] \\ &= E_Q \left[\int_0^1 \left(\frac{1}{2} \|\beta_-^Q - b_-^P(x(t), t)\|^2 + (\beta_-^Q - b_-^P(x(t), t)) \cdot \nabla \ln \psi(x(t), t) + \frac{1}{2} \|\nabla \ln \psi(x(t), t)\|^2 \right) dt \right] \\ &= E_Q \left[\int_0^1 \frac{1}{2} \|\beta_-^Q - b_-^P(x(t), t) + \nabla \ln \psi(x(t), t)\|^2 dt \right]. \end{aligned}$$

We then get

$$b_-^{Q*}(x, t) = b_-^P(x, t) - \nabla \ln \psi(x, t). \quad (44)$$

Thus, the solution Q^* is, in the language of Doob, an h -path process both in the forward and in the backward direction of time. We now identify ψ . By (32), we have

$$\begin{aligned}\tilde{\rho}(x, t) &= \varphi(x, t)\hat{\varphi}(x, t) = \varphi(x, t)\frac{\hat{\varphi}(x, t)}{\rho(x, t)}\rho(x, t) \\ &= \varphi(x, t)\psi(x, t)\rho(x, t), \quad \psi(x, t) = \frac{\hat{\varphi}(x, t)}{\rho(x, t)}, \quad \forall t \in [0, 1].\end{aligned}\quad (45)$$

Indeed, ψ , being the ratio of two solutions of the original Fokker-Planck (39), is reverse-time space-time harmonic, it namely satisfies (40) [39]. This agrees with the following calculation using (44), (28), (45) and (29)

$$\begin{aligned}b_-^{Q^*}(x, t) &= b_-^P(x, t) - \nabla \ln \psi(x, t) \\ &= b_+^P(x, t) - \nabla \ln \rho(x, t) - \nabla \ln \psi(x, t) \\ &= b_+^P(x, t) - \nabla \ln \hat{\varphi}(x, t) = b_+^P(x, t) \pm \nabla \ln \varphi(x, t) - \nabla \ln \hat{\varphi}(x, t) \\ &= b_+^{Q^*}(x, t) - \nabla \ln \tilde{\rho}(x, t),\end{aligned}\quad (46)$$

which is simply Nelson's duality relation for the drifts of Q^* . Formula (45) should be compared to [29, Theorem 3.4].

Finally, there are also conditional versions of these variational problems which are closer to standard stochastic control problems. Consider, for instance, minimizing the functional

$$\begin{aligned}J(u) &= E_{tx} \left[\int_t^1 \frac{1}{2} \|u(t)\|^2 dt - \ln \varphi_1(x(1)) \right], \\ dx(t) &= [b_+^P(x(t), t) + u(x(t), t)] dt + dw_+(t), \quad x(t) = x \text{ a.s.}\end{aligned}\quad (47)$$

over feedback controls u such that the differential equation has a weak solution. If $\varphi(x, t)$ solves (33) with terminal condition $\varphi_1(x)$, then, the same argument as before shows that $u^*(x, t) = \nabla \ln \varphi(x, t)$ is optimal and that $S(x, t) = -\ln \varphi(x, t) = \inf_u J(u)$ is the *value function* of the control problem. By (34), the Hamilton-Jacobi-Bellman equation has the form

$$\frac{\partial S}{\partial t} + \inf_u \left[(b_+^P + u) \cdot \nabla S + \frac{1}{2} \|u\|^2 \right] + \frac{1}{2} \Delta S = 0, \quad S(x, 1) = -\ln \varphi_1(x).$$

V. A TIME-SYMMETRIC FORMULATION

Inspired by a paper by Nagasawa [35], we proceed to derive a control time-symmetric formulation of the bridge problem. For any $Q \in \mathbb{D}$, define the *current* and *osmotic* drifts

$$v^Q = \frac{\beta_+^Q + \beta_-^Q}{2}, \quad u^Q = \frac{\beta_+^Q - \beta_-^Q}{2}.$$

Then

$$\beta_+^Q = v^Q + u^Q, \quad \beta_-^Q = v^Q - u^Q.$$

Observe that

$$\begin{aligned} H(Q, P) &= \frac{1}{2}H(q(0), p(0)) + \frac{1}{2}H(q(1), p(1)) \\ &\quad + E_Q \left[\int_0^1 \frac{1}{4} \|\beta_+^Q - \beta_+^P\|^2 + \frac{1}{4} \|\beta_-^Q - \beta_-^P\|^2 dt \right] \\ &= \frac{1}{2}H(q(0), p(0)) + \frac{1}{2}H(q(1), p(1)) \\ &\quad + E_Q \left[\int_0^1 \frac{1}{2} \|v^Q - v^P\|^2 + \frac{1}{2} \|u^Q - u^P\|^2 dt \right]. \end{aligned} \quad (48)$$

Since $H(q_0, p_0)$ and $H(q_1, p_1)$ are constant over $\mathbb{D}(\rho_0, \rho_1)$, it follows that the Schrödinger bridge Q^* minimizes the sum of the two incremental kinetic energies. Finally, we consider minimizing over $\mathbb{D}(\rho_0, \rho_1)$ the functional

$$I_s(Q) = \frac{1}{2} [I(Q) + \bar{I}(Q)].$$

By the previous calculation, this is equivalent to minimizing over $\mathbb{D}(\rho_0, \rho_1)$ the functional

$$E_Q \left[\int_0^1 \left(\frac{1}{2} \|v^Q - v^P\|^2 + \frac{1}{2} \|u^Q - u^P\|^2 \right) dt - \frac{1}{2} \ln \frac{\varphi}{\psi}(x(1), 1) + \frac{1}{2} \ln \frac{\varphi}{\psi}(x(0), 0) \right]. \quad (49)$$

The following current and osmotic drifts make the functional equal to zero and are therefore optimal

$$v^{Q^*}(x, t) = v^P(x, t) + \frac{1}{2} \nabla \ln \frac{\varphi}{\psi}(x, t), \quad (50)$$

$$u^{Q^*}(x, t) = u^P(x, t) + \frac{1}{2} \nabla \ln (\varphi\psi)(x, t), \quad (51)$$

which agree with (29) and (44). A variational analysis with the two controls v and u can be developed along the lines of [40, Sections III-IV].

VI. A FLUID DYNAMIC FORMULATION OF THE SCHRÖDINGER BRIDGE PROBLEM

Let us go back to the symmetric representation (48). In the case where the prior measure is $P = W$ stationary Wiener measure, we have $v^W = u^W = 0$ ¹. It basically corresponds to the situation where

¹See [29, pp. 7-8] for a justification of employing unbounded path measures in relative entropy problems.

there is no prior information. Considering that the boundary relative entropies are constant, we get that the problem is equivalent to minimizing

$$\mathbb{E} \left\{ \int_0^1 \left[\frac{1}{2} \|v\|^2 + \frac{1}{2} \|u\|^2 \right] dt \right\}$$

over $\mathbb{D}(\rho_0, \rho_1)$. Let us restrict our search to Markovian processes and recall Nelson's duality formula relating the two drifts

$$u(x, t) = \frac{1}{2} \nabla \ln \rho(x, t). \quad (52)$$

where u is the osmotic drift field, and the current drift field

$$v(x, t) = \frac{b_+(x, t) + b_-(x, t)}{2}.$$

Then,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0. \quad (53)$$

Thus, we get that the problem is equivalent to minimizing

$$\inf_{(\rho, v)} \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2} \|v(x, t)\|^2 + \frac{1}{8} \|\nabla \ln \rho(x, t)\|^2 \right] \rho(t, x) dt dx, \quad (54)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \quad (55)$$

$$\rho(0, x) = \rho_0(x), \quad \rho(1, y) = \rho_1(y), \quad (56)$$

which should be compared to (9a)-(9b)-(9c). We notice, in particular, that the two functionals differ by a term which is a multiple of the integral over time of the Fisher information functional

$$\int_{\mathbb{R}^n} \|\nabla \ln \rho(x, t)\|^2 \rho(t, x) dx.$$

This unveils a relation between the two problems without zero noise limits [31], [30].

Finally, we mention that a fluid dynamic problem concerning swarms of particles diffusing anisotropically with losses has been proposed and studied in [7]. It may or may not have a probabilistic counterpart as a Schrödinger bridge problem.

VII. OPTIMAL TRANSPORT WITH A ‘‘PRIOR’’

Considering the relation we have seen between the fluid dynamic versions of the optimal transport problem and the Schrödinger bridge problem, one may wonder whether there exists a formulation of the former which allows for an ‘‘a priori’’ evolution like in the latter. Relative entropy on path space does not work for zero-noise random evolutions as they are singular. Indeed, let P_ϵ and Q_ϵ be the measures on $C([0, 1], \mathbb{R}^n)$ equivalent to stationary Wiener measure W with forward differentials

$$\begin{aligned} dx(t) &= \beta_+^{P_\epsilon} dt + \sqrt{\epsilon} dw_+(t), \\ dx(t) &= \beta_+^{Q_\epsilon} dt + \sqrt{\epsilon} dw_+(t). \end{aligned} \tag{57}$$

Then, one can argue along the same lines as in Section III that

$$H(Q_\epsilon, P_\epsilon) = H(q_0, p_0) + \mathbb{E}_{Q_\epsilon} \left[\int_0^1 \frac{1}{2\epsilon} \|\beta_+^{Q_\epsilon} - \beta_+^{P_\epsilon}\|^2 dt \right].$$

For $\epsilon \searrow 0$, the relative entropy becomes infinite unless $Q_\epsilon = P_\epsilon$ ². We need therefore to take a different route, namely start with the following fluid dynamic control problem. Suppose we have two probability densities ρ_0 and ρ_1 and a flow of probability densities $\{\rho(x, t); 0 \leq t \leq 1\}$ satisfying

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \tag{58}$$

for some continuous vector field $v(\cdot, \cdot)$. We take (58) as our ‘‘prior’’ evolution and formulate the following problem. Let

$$\inf_{(\tilde{\rho}, \tilde{v})} \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2} \|\tilde{v}(x, t) - v(x, t)\|^2 \tilde{\rho}(t, x) dt dx, \tag{59a}$$

$$\frac{\partial \tilde{\rho}}{\partial t} + \nabla \cdot (\tilde{v}\tilde{\rho}) = 0, \tag{59b}$$

$$\tilde{\rho}(0, x) = \rho_0(x), \quad \tilde{\rho}(1, y) = \rho_1(y). \tag{59c}$$

Clearly, if the prior flow satisfies $\rho(x, 0) = \rho_0(x)$ and $\rho(x, 1) = \rho_1(x)$, then it solves the problem and $\tilde{v}^* = v$. Moreover, the standard optimal transport problem is recovered when $v \equiv 0$, namely the prior evolution is constant in time.

Let us try to provide further motivation to study problem (59). Consider the situation where a previous optimal transport problem (9) has been solved with boundary marginals $\bar{\rho}_0$ and $\bar{\rho}_1$ leading to

² This calculation indicates that there may be a limit as $\epsilon \searrow 0$ of $\inf\{\epsilon H(Q_\epsilon, P_\epsilon)\}$ and, hopefully, in suitable sense, of the minimizers. This is indeed the case, see [31], [30], [29] for a precise statement of limiting results.

the optimal velocity field $v(x, t)$. Here say $\bar{\rho}_0$ represent resources being produced to satisfy the demand $\bar{\rho}_1$. Suppose now new information becomes available showing that the actual resources available are distributed according to ρ_0 and the actual demand is distributed according to ρ_1 . As we had already set up a transportation plan according to velocity field v , we seek to solve a new transport problem where the new evolution is close to the one we would have employing the previous velocity field. This is represented in problem (59).

Remark 7.1: The particle version of (59) takes the form of a more familiar OMT problem, namely, in the notation of Section II,

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y) d\pi(x, y), \quad (60)$$

where

$$c(x, y) = \inf_{x \in \mathcal{X}_{xy}} \int_0^1 L(t, x(t), \dot{x}(t)) dt, \quad L(t, x, \dot{x}) = \|\dot{x} - v(x, t)\|^2. \quad (61)$$

The explicit calculation of the function $c(x, y)$ when $v \neq 0$ is nontrivial. Moreover, the zero noise limit results of [30, Section 3], based on a Large Deviations Principle [12], although very general in other ways, seem to cover here only the case where $c(x, y) = c(x - y)$ strictly convex originating from a Lagrangian $L(t, x, \dot{x}) = c(\dot{x})$. Finally, we feel that our formulation is a most natural one in which to study zero noise limits of Schroedinger bridges with a general Markovian prior evolution. In the next section, we discuss this problem in the Gaussian case. The proof of the convergence of the path-space measures of the minimisers can be done along the lines of [30] where Γ -convergence of the bridge minimum problems to the OMT problem is established. This, under suitable assumptions, guarantees convergence of the minimizers.

The variational analysis for (59) can be carried out as in Section II. Let $\mathcal{P}_{\rho_0 \rho_1}$ be again the family of flows of probability densities $\rho = \{\rho(\cdot, t); 0 \leq t \leq 1\}$ satisfying (59c). Let \mathcal{V} be the family of continuous feedback control laws $\tilde{v}(\cdot, \cdot)$. Consider the unconstrained minimization of the Lagrangian over $\mathcal{P}_{\rho_0 \rho_1} \times \mathcal{V}$

$$\mathcal{L}(\tilde{\rho}, v) = \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2} \|\tilde{v}(x, t) - v(x, t)\|^2 \tilde{\rho}(t, x) + \lambda(x, t) \left(\frac{\partial \tilde{\rho}}{\partial t} + \nabla \cdot (\tilde{v} \tilde{\rho}) \right) \right] dt dx, \quad (62)$$

where again λ is a C^1 Lagrange multiplier. After integration by parts, assuming that limits for $x \rightarrow \infty$ are zero, and observing that the boundary values are constant over $\mathcal{P}_{\rho_0 \rho_1}$, we get the problem

$$\inf_{(\tilde{\rho}, \tilde{v}) \in \mathcal{P}_{\rho_0 \rho_1} \times \mathcal{V}} \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2} \|\tilde{v}(x, t) - v(x, t)\|^2 + \left(-\frac{\partial \lambda}{\partial t} - \nabla \lambda \cdot \tilde{v} \right) \right] \tilde{\rho}(x, t) dt dx \quad (63)$$

Pointwise minimization with respect to \tilde{v} for each fixed flow of probability densities $\tilde{\rho}$ gives

$$v_\rho^*(x, t) = v(x, t) + \nabla\lambda(x, t). \quad (64)$$

Plugging this form of the optimal control into (63), we get the functional of $\tilde{\rho} \in \mathcal{P}_{\rho_0, \rho_1}$

$$J(\tilde{\rho}) = - \int_{\mathbb{R}^n} \int_0^1 \left[\frac{\partial\lambda}{\partial t} + v \cdot \nabla\lambda + \frac{1}{2} \|\nabla\lambda\|^2 \right] \tilde{\rho}(x, t) dt dx. \quad (65)$$

We then have the following result:

Proposition 7.2: If $\tilde{\rho}^*$ satisfying

$$\frac{\partial\tilde{\rho}^*}{\partial t} + \nabla \cdot [(v + \nabla\psi)\tilde{\rho}^*] = 0, \quad \tilde{\rho}^*(x, 0) = \rho_0(x), \quad (66)$$

where ψ is a solution of the Hamilton-Jacobi equation

$$\frac{\partial\psi}{\partial t} + v \cdot \nabla\psi + \frac{1}{2} \|\nabla\psi\|^2 = 0, \quad (67)$$

is such that $\tilde{\rho}^*(x, 1) = \rho_1(x)$, then the pair $(\tilde{\rho}^*(x, t), v^*(x, t) = v(x, t) + \nabla\psi(x, t))$ is a solution of the problem (59).

If $v(x, t) = \alpha(t)x$, and both ρ_0 and ρ_1 are Gaussian, then the optimal evolution is given by a linear equation and is therefore given by a Gaussian process as we will study next.

VIII. GAUSSIAN CASE

In this section, we consider the correspondence between Schrödinger bridges and optimal mass transport for the special case where the underlying dynamics are linear and the marginals are normal distributions. To this end, consider the reference evolution

$$dx(t) = A(t)x(t)dt + \sqrt{\epsilon}dw(t) \quad (68)$$

and the two marginals

$$\rho_0(x) = (2\pi|\Sigma_0|)^{-n/2} \exp \left[-\frac{1}{2}(x - m_0)' \Sigma_0^{-1}(x - m_0) \right], \quad (69a)$$

$$\rho_1(x) = (2\pi|\Sigma_1|)^{-n/2} \exp \left[-\frac{1}{2}(x - m_1)' \Sigma_1^{-1}(x - m_1) \right], \quad (69b)$$

where prime denotes transposition. In our previous work [6], we derived a ‘‘closed form’’ expression for the corresponding Schrödinger bridge for the case when $m_0 = m_1 = 0$, namely,

$$dx(t) = (A(t) - \Pi_\epsilon(t))x(t)dt + \sqrt{\epsilon}dw(t) \quad (70)$$

with $\Pi_\epsilon(t)$ satisfying the matrix Riccati equation

$$\dot{\Pi}_\epsilon(t) + A(t)' \Pi_\epsilon(t) + \Pi_\epsilon(t) A(t) - \Pi_\epsilon(t)^2 = 0 \quad (71)$$

and the boundary condition

$$\Pi_\epsilon(0) = \Sigma_0^{-1/2} \left[\frac{\epsilon}{2} I + \Sigma_0^{1/2} \Phi'_{10} M_{10}^{-1} \Phi_{10} \Sigma_0^{1/2} - \left(\frac{\epsilon^2}{4} I + \Sigma_0^{1/2} \Phi'_{10} M_{10}^{-1} \Sigma_1 M_{10}^{-1} \Phi_{10} \Sigma_0^{1/2} \right)^{1/2} \right] \Sigma_0^{-1/2}.$$

Here $\Phi_{10} := \Phi(1, 0)$ is the state transition matrix from 0 to 1 and

$$M_{10} := M(1, 0) = \int_0^1 \Phi(1, t) \Phi(1, t)' dt$$

is the controllability gramian. This can be easily adjusted for the case when $m_0 \neq 0$ or $m_1 \neq 0$ by adding an extra deterministic drift term to account for the change in the mean as follows:

$$dx(t) = (A(t) - \Pi_\epsilon(t))x(t)dt + m(t)dt + \sqrt{\epsilon}dw(t) \quad (72)$$

where

$$m(t) = \hat{\Phi}(1, t)' \hat{M}(1, 0)^{-1} (m_1 - \hat{\Phi}(1, 0)m_0) \quad (73)$$

with $\hat{\Phi}(t, s)$, $\hat{M}(t, s)$ satisfying

$$\frac{\partial \hat{\Phi}(t, s)}{\partial t} = (A(t) - \Pi_\epsilon(t)) \hat{\Phi}(t, s), \quad \hat{\Phi}(t, t) = I$$

and

$$\hat{M}(t, s) = \int_s^t \hat{\Phi}(t, \tau) \hat{\Phi}(t, \tau)' d\tau.$$

We now consider ‘‘slowing down’’ the reference evolution by letting ϵ go to 0. In the case where $A(t) \equiv 0$, the Schrödinger bridge solution process converges to the solution of optimal mass transport problem [31], [29]. In general, when $A(t) \neq 0$, by taking $\epsilon = 0$ we obtain

$$\Pi_0(0) = \Sigma_0^{-1/2} \left[\Sigma_0^{1/2} \Phi'_{10} M_{10}^{-1} \Phi_{10} \Sigma_0^{1/2} - \left(\Sigma_0^{1/2} \Phi'_{10} M_{10}^{-1} \Sigma_1 M_{10}^{-1} \Phi_{10} \Sigma_0^{1/2} \right)^{1/2} \right] \Sigma_0^{-1/2}, \quad (74)$$

and a limiting process

$$dx(t) = (A(t) - \Pi_0(t))x(t)dt + m(t)dt, \quad x(0) \sim (m_0, \Sigma_0) \quad (75)$$

with $\Pi_0(t)$, $m(t)$ satisfying (71), (73) and (74). In fact $\Pi_0(t)$ has explicit expression

$$\begin{aligned} \Pi_0(t) = & -M(t, 0)^{-1} - M(t, 0)^{-1} \Phi(t, 0) \left[\Phi'_{10} M_{10}^{-1} \Phi_{10} \right. \\ & \left. - \Sigma_0^{-1/2} \left(\Sigma_0^{1/2} \Phi'_{10} M_{10}^{-1} \Sigma_1 M_{10}^{-1} \Phi_{10} \Sigma_0^{1/2} \right)^{1/2} \Sigma_0^{-1/2} - \Sigma_0^{-1/2} \right]^{-1} \Phi(t, 0)' M(t, 0)^{-1} \end{aligned} \quad (76)$$

It turns out that process (75) yields an optimal solution to the transport problem (59) as stated next.

Theorem 8.1: Let $\tilde{\rho}(\cdot, t)$ be the probability density of $x(t)$ in (75), and $\tilde{v}(x, t) = (A(t) - \Pi_0(t))x + m(t)$. Then the pair $(\tilde{\rho}, \tilde{v})$ is a solution of the problem (59) with prior velocity field $v(x, t) = A(t)x$.

Proof: To show that the pair $(\tilde{\rho}, \tilde{v})$ is a solution, we need to prove i) $\tilde{\rho}$ satisfies the boundary condition $\tilde{\rho}(x, 1) = \rho_1(x)$ and ii) $\tilde{v}(x, t) - v(x, t) = \nabla\psi(x, t)$ for some ψ with ψ satisfying the Hamilton-Jacobi equation (67). Here $v(x, t) = A(t)x$ is the drift of the prior process.

We first show that $\tilde{\rho}$ satisfies the boundary condition $\tilde{\rho}(x, 1) = \rho_1(x)$. Since the process (75) is a linear diffusion with gaussian initial condition, $x(t)$ is a gaussian random vector for all $t \in [0, 1]$. Let

$$\tilde{\rho}(x, t) = (2\pi|\Sigma(t)|)^{-n/2} \exp \left[-\frac{1}{2}(x - n(t))'\Sigma(t)^{-1}(x - n(t)) \right].$$

Then obviously the mean value $n(t)$ is

$$n(t) = \hat{\Phi}(t, 0)m_0 + \int_0^t \hat{\Phi}(t, \tau)m(\tau)d\tau.$$

We claim that the covariance $\Sigma(t)$ has the explicit expression

$$\begin{aligned} \Sigma(t) &= M(t, 0)\Phi(0, t)'\Sigma_0^{-1/2} \left[-\Sigma_0^{1/2}\Phi'_{10}M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2} + (\Sigma_0^{1/2}\Phi'_{10}M_{10}^{-1}\Sigma_1M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2} \right. \\ &\quad \left. + \Sigma_0^{1/2}\Phi(t, 0)'M(t, 0)^{-1}\Phi(t, 0)\Sigma_0^{1/2} \right]^2 \Sigma_0^{-1/2}\Phi(0, t)M(t, 0), \end{aligned} \quad (77)$$

for $t \in (0, 1]$. This expression is consistent with the initial condition Σ_0 since

$$\lim_{t \searrow 0} \Sigma(t) = \Sigma_0.$$

To see that $\Sigma(t)$ is the covariance matrix of $x(t)$, we only need to show that $\Sigma(t)$ satisfies the differential equation

$$\dot{\Sigma}(t) = (A(t) - \Pi_0(t))\Sigma(t) + \Sigma(t)(A(t) - \Pi_0(t))'.$$

This can be verified directly from the expression (77) and (76) after some straightforward but lengthy computations. Now observing that

$$\begin{aligned} n(1) &= \hat{\Phi}(1, 0)m_0 + \int_0^1 \hat{\Phi}(1, \tau)m(\tau)d\tau \\ &= \hat{\Phi}(1, 0)m_0 + \int_0^1 \hat{\Phi}(1, \tau)\hat{\Phi}(1, \tau)'d\tau\hat{M}(1, 0)^{-1}(m_1 - \hat{\Phi}(1, 0)m_0) = m_1 \end{aligned}$$

and

$$\Sigma(1) = M(1, 0)\Phi(0, 1)'\Sigma_0^{-1/2} \left[(\Sigma_0^{1/2}\Phi'_{10}M_{10}^{-1}\Sigma_1M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2} \right]^2 \Sigma_0^{-1/2}\Phi(0, 1)M(1, 0) = \Sigma_1,$$

we conclude that $\tilde{\rho}$ satisfies the boundary condition $\tilde{\rho}(x, 1) = \rho_1(x)$.

We next show ii). Let

$$\psi(x, t) = -\frac{1}{2}x'\Pi_0(t)x + m(t)'x + c(t)$$

with

$$c(t) = -\frac{1}{2}\int_0^t m(\tau)'m(\tau)d\tau,$$

then

$$\begin{aligned} \frac{\partial\psi}{\partial t} + v \cdot \nabla\psi + \frac{1}{2}\|\nabla\psi\|^2 &= -\frac{1}{2}x'\dot{\Pi}_0(t)x + \dot{m}(t)'x + \dot{c}(t) + x'A(t)'(m(t) - \Pi_0(t)x) + \frac{1}{2}\|m(t)' - x'\Pi_0(t)\|^2 \\ &= \frac{1}{2}x'(A(t)'\Pi_0 + \Pi_0A(t) - \Pi_0(t)^2)x - m(t)'(A(t) - \Pi_0(t))x + \dot{c}(t) \\ &\quad + x'A(t)'(m(t) - \Pi_0(t)x) + \frac{1}{2}(m(t)' - x'\Pi_0(t))(m(t) - \Pi_0(t)x) \\ &= \dot{c}(t) + \frac{1}{2}m(t)'m(t) = 0. \end{aligned}$$

This completes the proof. ■

IX. EXAMPLE: SHIFTING THE MEAN OF NORMAL DISTRIBUTIONS

For illustration purposes, we consider the Schrödinger bridge problem on the time interval $[0, 1]$ and $x \in \mathbb{R}$ when the “prior” is σW_t and the two marginals are

$$\rho_0(x) = (2\pi)^{-1/2} \exp\left[-\frac{x^2}{2}\right], \quad \rho_1(x) = (2\pi)^{-1/2} \exp\left[-\frac{(x-1)^2}{2}\right]. \quad (78)$$

By the general theory, we know that the bridge has forward differential

$$dx(t) = \frac{\partial}{\partial x} \ln \varphi(x(t), t) dt + \sigma dw(t)$$

where φ solves together with $\hat{\varphi}$ the Schrödinger system

$$\begin{cases} \frac{\partial\varphi}{\partial t} + \frac{\sigma^2}{2}\Delta\varphi = 0, & \varphi(x, 0)\hat{\varphi}(x, 0) = \rho_0(x) \\ \frac{\partial\hat{\varphi}}{\partial t} - \frac{\sigma^2}{2}\Delta\hat{\varphi} = 0, & \varphi(x, 1)\hat{\varphi}(x, 1) = \rho_1(x). \end{cases}, \quad (79)$$

It can be seen that

$$dx(t) = \left[\frac{\sigma^2}{\sigma^2 t + c} x(t) + \frac{c}{\sigma^2 t + c} \right] dt + \sigma dw(t),$$

with

$$c = -\frac{\sigma^2}{\sigma^2/2 + 1 - \sqrt{1 + \sigma^4/4}}.$$

It follows that $m_t = \mathbb{E}\{x(t)\}$ satisfies

$$\dot{m}_t = \frac{\sigma^2}{\sigma^2 t + c} m_t + \frac{c}{\sigma^2 t + c}, \quad m(0) = 0, \quad m(1) = 1.$$

We get $m_t = t$. The current drift of the Schrödinger bridge is

$$\tilde{v}(x, t) = \frac{\sigma^2}{\sigma^2 t + c} x + \frac{c}{\sigma^2 t + c} - \frac{\sigma^2}{2} \nabla \ln \tilde{\rho}_t(x),$$

where $\tilde{\rho}_t$ has the form

$$\tilde{\rho}_t = (2\pi)^{-1/2} \exp \left[-\frac{(x-t)^2}{2q(t)} \right].$$

Hence,

$$\tilde{v}(x, t) = \frac{\sigma^2}{\sigma^2 t + c} x + \frac{c}{\sigma^2 t + c} - \frac{\sigma^2}{2} \frac{t-x}{q(t)}.$$

As $\sigma^2 \searrow 0$, $c \rightarrow -2$, $q(t) \rightarrow 1$ and $\tilde{v}(x, t) \rightarrow 1, \forall x, \forall t$ (while $\tilde{u}(x, t) = \frac{\sigma^2}{2} \nabla \ln \tilde{\rho}(x, t) \rightarrow 0$), which is just the optimal control of the corresponding optimal transport problem. This is in agreement with the general theory [31], [29].

X. NUMERICAL EXAMPLE

We consider highly overdamped Brownian motion in a force field. Then, in a very strong sense [36, Theorem 10.1], the Smoluchowski model in configuration variables is a good approximation of the full Ornstein-Uhlenbeck model in phase space. We are interested in planar Brownian motion in the quadratic potential

$$V(x) = \frac{1}{2} x' 3 I_2 x = \frac{1}{2} [x_1, x_2] \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Taking the mass of the particle to be equal to one, the planar evolution of the Brownian particle is given by the Smoluchowski equation

$$dx(t) = -\nabla V(x(t)) dt + \sqrt{\epsilon} dw(t), \quad -\nabla V(x) = Ax, \quad A = \begin{bmatrix} -3 & 0 \\ 0 & -3 \end{bmatrix}, \quad (80)$$

where w is a standard, two-dimensional Wiener process. The observed distributions of the particle at the two end-points in time are normal with mean and variance

$$m_0 = \begin{bmatrix} -5 \\ -5 \end{bmatrix}, \quad \text{and} \quad \Sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

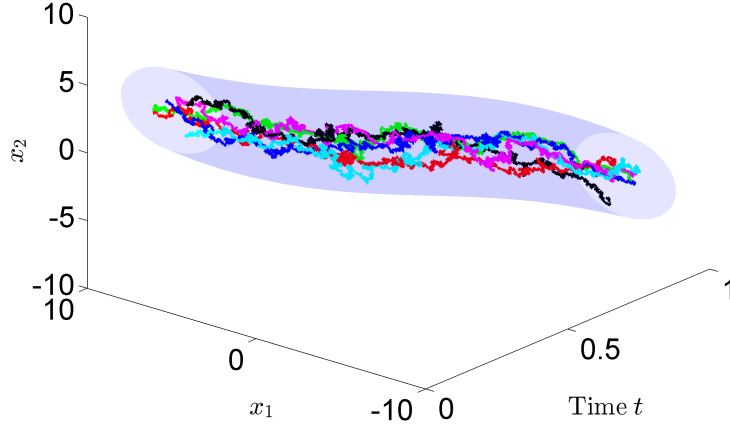


Fig. 1. Interpolation based on Schrödinger bridge with $\epsilon = 9$

at $t = 0$, and

$$m_1 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \text{ and } \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

at $t = 1$, respectively. We then seek to interpolate the density of the particle at intermediate points by solving the corresponding Schrödinger bridge problem where (80) plays the role of an a priori evolution.

Figure 1 depicts the flow between the two one-time marginals for the Schrödinger bridge when $\epsilon = 9$. The transparent tube represent the “ 3σ region”

$$(x' - m'_t)\Sigma_t^{-1}(x - m_t) \leq 9.$$

Typical sample paths are shown in the figure. Similarly, Figures 2 and 3 depict the corresponding flows for $\epsilon = 4$ and $\epsilon = 0.01$, respectively. Figure 4 is the limit that represents optimal mass transport with prior velocity field $v(x, t) = Ax$; the sample paths are smooth curves that follow optimal transportation paths. As $\epsilon \searrow 0$, the paths of the bridge process resemble those of the corresponding optimal transport process for $\epsilon = 0$. For comparison, we also provide in Figure 5 the interpolation corresponding to optimal transport without a prior, which is given by a constant speed translation.

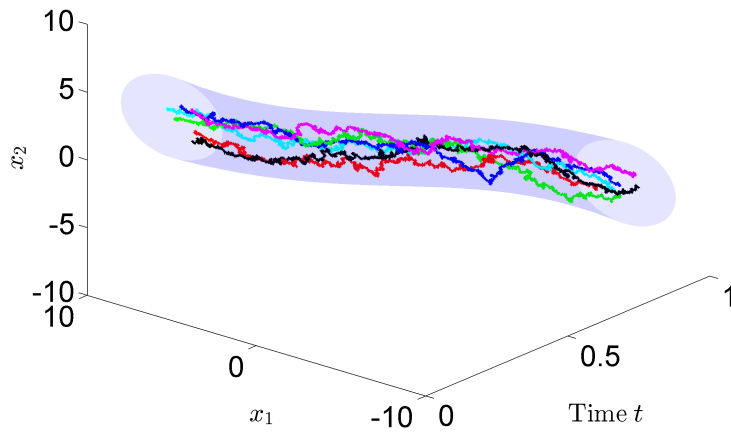


Fig. 2. Interpolation based on Schrödinger bridge with $\epsilon = 4$

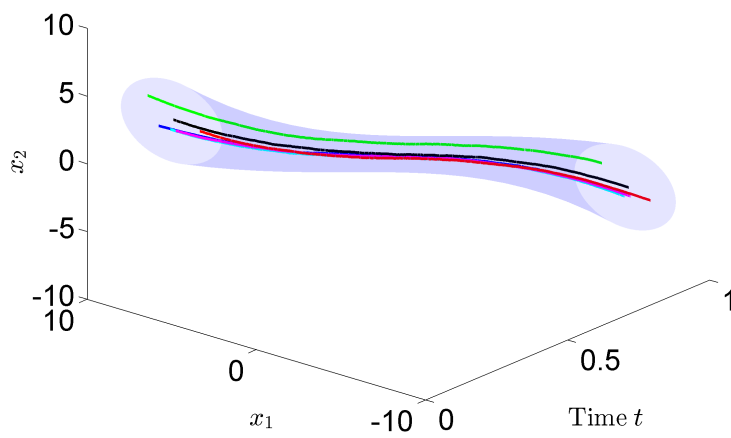


Fig. 3. Interpolation based on Schrödinger bridge with $\epsilon = 0.01$

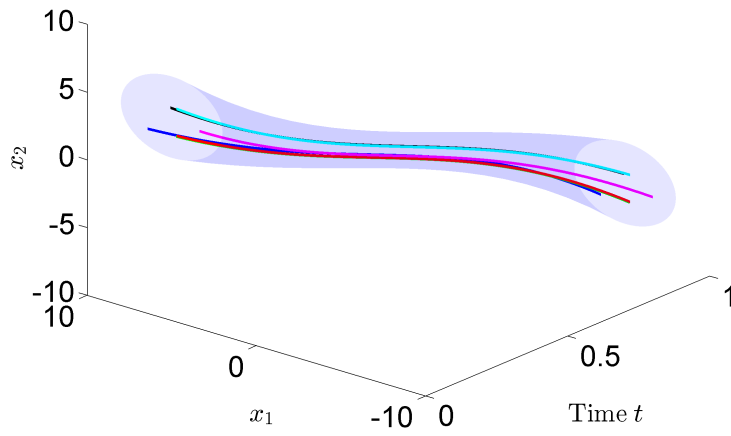


Fig. 4. Interpolation based on optimal transport with prior

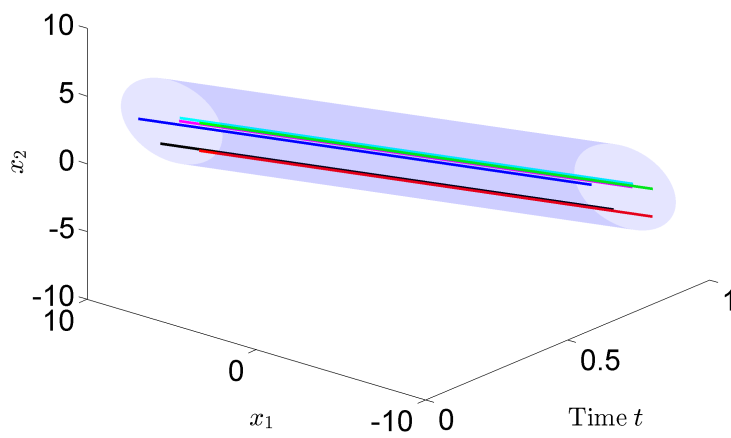


Fig. 5. Interpolation based on optimal transport without prior

ACKNOWLEDGMENT

Part of the research of M.P. was conducted during a stay at the Courant Institute of Mathematical Sciences of the New York University whose hospitality is gratefully acknowledged.

REFERENCES

- [1] L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2nd ed. 2008.
- [2] S. Angenent, S. Haker, and A. Tannenbaum. "Minimizing Flows for the Monge–Kantorovich Problem." *SIAM journal on mathematical analysis* 35.1 (2003): 61-97.
- [3] J. Benamou and Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, *Numerische Mathematik*, **84**, no. 3, 2000, 375-393.
- [4] A. Beurling, An automorphism of product measures, *Ann. Math.* **72** (1960), 189-200.
- [5] Y. Chen and T.T. Georgiou, Stochastic bridges of linear systems, <http://arxiv.org/abs/1407.3421>.
- [6] Y. Chen, T.T. Georgiou and M. Pavon, Optimal steering of a linear stochastic system to a final probability distribution, <http://arxiv.org/abs/1408.2222>.
- [7] Y. Chen, T. Georgiou and M. Pavon, Optimal steering of inertial particles diffusing anisotropically with losses, <http://arxiv.org/abs/1410.1605>.
- [8] Y. Chen, T.T. Georgiou and M. Pavon, Optimal steering of a linear stochastic system to a final probability distribution, part II <http://arxiv.org/abs/1410.3447>.
- [9] Y. Chen, T.T. Georgiou and M. Pavon, Fast cooling for a system of stochastic oscillators, <http://arxiv.org/abs/1411.1323>.
- [10] P. Dai Pra, A stochastic control approach to reciprocal diffusion processes, *Applied Mathematics and Optimization*, **23** (1), 1991, 313-329.
- [11] P. Dai Pra and M. Pavon, On the Markov processes of Schroedinger, the Feynman-Kac formula and stochastic control, in *Realization and Modeling in System Theory - Proc. 1989 MTNS Conf.*, M.A.Kaashoek, J.H. van Schuppen, A.C.M. Ran Eds., Birkaeuser, Boston, 1990, 497- 504.
- [12] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, second ed., Applied Math., vol 38, Springer-Verlag, 1998.
- [13] R. Fillieger and M.-O. Hongler, Relative entropy and efficiency measure for diffusion-mediated transport processes, *J. Physics A: Mathematical and General* **38** (2005), 1247-1255.
- [14] R. Fillieger, M.-O. Hongler and L. Streit, Connection between an exactly solvable stochastic optimal control problem and a nonlinear reaction-diffusion equation, *J. Optimiz. Theory Appl.* **137** (2008), 497-505.
- [15] H. Föllmer, Time reversal on Wiener Space, in *Stochastic Processes - Mathematics and Physics*, Lecture Notes in Mathematics (Springer-Verlag, New York, 1986), Vol. 1158, pp. 119-129.
- [16] H. Föllmer, Random fields and diffusion processes, in: *École d'Été de Probabilités de Saint-Flour XV-XVII*, edited by P. L. Hennequin, Lecture Notes in Mathematics, Springer-Verlag, New York, 1988, vol.1362,102-203.
- [17] W. Fleming and M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Second Edition, Springer (2006.)
- [18] R. Fortet, Résolution d'un système d'équations de M. Schrödinger, *J. Math. Pure Appl.* IX (1940), 83-105.
- [19] W. Gangbo, R.J. McCann, The geometry of optimal transportation, *Acta Mathematica* 177.2 (1996): 113-161.

- [20] T. T. Georgiou and M. Pavon, Positive contraction mappings for classical and quantum Schroedinger systems, 2014, arXiv:1405.6650v2, submitted for publication.
- [21] R. Graham, Path integral methods in nonequilibrium thermodynamics and statistics, in *Stochastic Processes in Nonequilibrium Systems*, L. Garrido, P. Seglar and P.J.Shepherd Eds., Lecture Notes in Physics 84, Springer-Verlag, New York, 1978, 82-138.
- [22] U.G.Haussmann and E.Pardoux, Time reversal of diffusions, *The Annals of Probability* **14**, 1986, 1188.
- [23] D.B.Hernandez and M. Pavon, Equilibrium description of a particle system in a heat bath, *Acta Applicandae Mathematicae* **14** (1989), 239-256.
- [24] B. Jamison, The Markov processes of Schrödinger, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **32** (1975), 323-331.
- [25] X. Jiang, Z. Luo, and T. Georgiou, Geometric methods for spectral analysis, *IEEE Transactions on Signal Processing*, **60**, no. 3, 2012, 1064-1074.
- [26] R. Jordan, D. Kinderlehrer, and F. Otto, The variational formulation of the Fokker-Planck equation, *SIAM J. Math. Anal.* **29**: 1-17 (1998).
- [27] L. Kantorovich, On the transfer of masses, Dokl. Akad. Nauk. SSSR, vol. 37, pp. 227-229, 1942.
- [28] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1988.
- [29] C. Léonard, A survey of the Schroedinger problem and some of its connections with optimal transport, *Discrete Contin. Dyn. Syst. A*, 2014, **34** (4): 1533-1574.
- [30] C. Léonard, From the Schrödinger problem to the Monge-Kantorovich problem, *J. Funct. Anal.*, 2012, **262**, 1879-1920.
- [31] T. Mikami, Monge's problem with a quadratic cost by the zero-noise limit of h-path processes, *Probab. Theory Relat. Fields*, **129**, (2004), 245-260.
- [32] T. Mikami and M. Thieullen, Duality theorem for the stochastic optimal control problem., *Stoch. Proc. Appl.*, 116, 1815-1835 (2006).
- [33] T. Mikami and M. Thieullen, Optimal Transportation Problem by Stochastic Optimal Control, *SIAM Journal of Control and Optimization*, **47**, N. 3, 1127-1139 (2008).
- [34] G. Monge, Mémoire sur la théorie des déblais et des remblais. (Open Library): De l'Imprimerie Royale, 1781.
- [35] M. Nagasawa, Stochastic variational principle of Schrödinger processes, in *Seminar on Stochastic Processes*, Çinlar et al. Eds., Birkhäuser, 1989, 165-175.
- [36] E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton, 1967.
- [37] L. Ning, T.T. Georgiou, and A. Tannenbaum, Matrix-valued Monge-Kantorovich Optimal Mass Transport, *IEEE Trans. Aut. Contr.*, 2014 (DOI 10.1109/TAC.2014.2350171).
- [38] R. Jordan, D. Kinderlehrer, and F. Otto, The variational formulation of the Fokker-Planck equation, *SIAM journal on mathematical analysis* 29.1 (1998): 1-17.
- [39] M. Pavon, Stochastic control and nonequilibrium thermodynamical systems, *Appl. Math. and Optimiz.* **19** (1989), 187-202.
- [40] M. Pavon, Hamilton's principle in stochastic mechanics, *J. Math. Physics* **36** (1995), 6774-6800.
- [41] M. Pavon and F. Ticozzi, Discrete-time classical and quantum Markovian evolutions: Maximum entropy problems on path space, *J. Math. Phys.*, **51**, 042104-042125 (2010) doi:10.1063/1.3372725.
- [42] M. Pavon and A. Wakolbinger, On free energy, stochastic control, and Schroedinger processes, Modeling, Estimation and Control of Systems with Uncertainty, G.B. Di Masi, A.Gombani, A.Kurzanski Eds., Birkauer, Boston, 1991, 334-348.
- [43] S. Rachev and L. Rüschendorf, Mass Transportation Problems: Theory, vol. 1. New York: Springer-Verlag, 1998.

- [44] E. Schrödinger, Über die Umkehrung der Naturgesetze, *Sitzungsberichte der Preuss Akad. Wissen.* Berlin, Phys. Math. Klasse (1931), 144-153.
- [45] E. Schrödinger, Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique, *Ann. Inst. H. Poincaré* 2, 269 (1932).
- [46] E. G. Tabak and G. Trigila, Data-Driven Optimal Transport, submitted to CPAM, 2014.
- [47] C. Villani, *Topics in optimal transportation*, AMS, 2003, vol. 58.
- [48] C. Villani, *Optimal transport: old and new*, Vol. 338. Springer, 2008.
- [49] A. Wakolbinger, Schroedinger bridges from 1931 to 1991, in *Proc. of the 4th Latin American Congress in Probability and Mathematical Statistics*, Mexico City 1990, *Contribuciones en probabilidad y estadística matematica*, **3** (1992), 61-79.