Optimal transport over a linear dynamical system

Yongxin Chen, Tryphon T. Georgiou and Michele Pavon

Abstract—We consider the problem of steering an initial probability density for the state vector of a linear system to a final one, in finite time, using minimum energy control. In the case where the dynamics correspond to an integrator ($\dot{x}(t) = u(t)$) this amounts to a Monge-Kantorovich Optimal Mass Transport (OMT) problem. In general, we show that the problem can again be reduced to solving an OMT problem and that it has a unique solution. In parallel, we study the optimal steering of the state-density of a linear stochastic system with white noise disturbance; this is known to correspond to a *Schrödinger bridge*. As the white noise intensity tends to zero, the flow of densities converges to that of the deterministic dynamics and can serve as a way to compute the solution of its deterministic counterpart. The solution can be expressed in closed-form for Gaussian initial and final state densities in both cases.

Keywords: Optimal mass transport, Schrödinger bridges, stochastic linear systems, optimal control.

I. INTRODUCTION

We are interested in stochastic control problems to steer the probability density of the state vector of a linear system between an initial and a final distribution for two cases, i) with and ii) without stochastic disturbance. That is, we consider the linear dynamics

$$dx(t) = A(t)x(t)dt + B(t)u(t)dt + \sqrt{\epsilon}B(t)dw(t) \quad (1)$$

where w is a Wiener process, u is a control input, x is the state process, and A, B is a controllable pair of continuous matrices, for the two cases where i) $\epsilon > 0$ and ii) $\epsilon = 0$. In either case, the state is a random vector with an initial distribution μ_0. Our task is to determine a minimum energy input that drives the system to a final state distribution μ_1 over the time interval $[0, 1]$, that is, the minimum of

$$\mathbb{E}\{\int_0^1 \|u(t)\|^2 dt\} \quad (2)$$

subject to μ_1 being the probability distribution of the state vector at the terminal time $t = 1$.

When the state distribution represents density of particles whose position obeys $\dot{x}(t) = u(t)$ (i.e., $A(t) \equiv 0$, $B(t) \equiv I$, and $\epsilon = 0$) the problem reduces to the classical Optimal Mass Transport (**OMT**) problem with quadratic cost [1], [2]. Historically, the modern formulation of OMT is due to Leonid Kantorovich [3] and has been the focus of dramatic developments because of its relevance in many diverse fields including economics, physics, engineering, and probability [4], [5], [6], [7], [2], [1], [8], [9], [10], [11]. Thus, the above problem, for $\epsilon = 0$, represents a generalization of OMT to deal with particles obeying known "prior" *non-trivial* dynamics while being steered between two end-point distributions – we refer to this as

the problem of *OMT with prior dynamics* (**OMT-wpd**). The problem of OMT-wpd was first introduced in our previous work [12] for the case where $B(t) \equiv I$. The difference of course to the classical OMT is that, here, the linear dynamics are arbitrary and may facilitate or hinder transport. Applications are envisioned in the steering of particle beams through time-varying potential, the steering of swarms (UAV's, large collection of microsatelites, ensemble control, etc.), as well as in the modeling of the flow and collective motion of particles, clouds, platoons, flocking of insects, birds, fish, etc. between end-point distributions [13] and the interpolation/morphing of distributions [14]. From a controls perspective, "important limitations standing in the way of the wider use of optimal control can be circumvented by explicitly acknowledging that in most situations the apparatus implementing the control policy will be judged on its ability to cope with a distribution of initial states, rather than a single state" as pointed out by R. Brockett in [15, page 23].

In the case where $\epsilon > 0$ and a stochastic disturbance is present, the flow of "particles" is dictated by dynamics as well as by Brownian diffusion. The corresponding stochastic control problem to steer the state density function between the end-point distributions has been shown to be equivalent to the so-called *Schrödinger bridge problem* [16]. The Schrödinger bridge problem, in its original formulation [17], [18], [19], seeks a probability law on path space with given two end-point marginals which is close to a Markovian *prior* in the sense of relative entropy. Important contributions were due to Fortet, Beurling, Jamison and Föllmer [20], [21], [22], [23]. Schrödinger's original vision was to provide a formulation of Quantum Mechanics based on diffusion processes which was accomplished in various versions of Stochastic Mechanics [24], [25]. More recent attempts to directly connect the Schrödinger bridge problem to Quantum Mechanics can be found in [26], [27], [28]. Another closely related area of research has been that of *reciprocal processes,* with important engineering applications in, e.g., image processing and other fields [29], [22], [30], [31], [32], [33], [34].

Renewed interest in Schrödinger bridges was sparked after a close relationship to stochastic control was recognized [16], [35], [36]. Recently, the present authors [37], [38], [39] have provided an attractive, implementable solution to the Schrödinger bridge problem for the Gauss-Markov case as well as extended the earlier theory in several directions. In particular, these include the cases of degenerate noise, differing noise and control channels, infinite horizon, and anisotropic diffusing particles with losses. Certain physics applications have also been developed in [40]. The Schrödinger bridge problem can be seen as a stochastic version of OMT due to the presence of the diffusive term in the dynamics. As a result, its solution is more well behaved due to the smoothing property of the Laplacian. On the other hand, it follows from [41], [42], [43], [44] that for the special case $A(t) \equiv 0$ and $B(t) \equiv I$, the solution to the Schrödinger bridge problem tends to that of the OMT when "slowing down" the diffusion by taking $\epsilon \to 0$. These two facts suggest the Schrödinger bridge problem as a means to construct approximate solutions to OMT for both, the standard one as well as the problem of OMT with prior dynamics.

In this paper, we continue the work initiated in [12] on the

connection between the bridge and OMT problems. In particular, we provide the first solution to the problem of optimally steering between two given probability densities for the state-vector of a linear noise-free dynamical model. As noted, the importance of such problems has been advocated by R. Brockett in [15], [45], who solved the controllability problem for the Gauss-Markov case and provided necessary conditions for a certain type of optimality. The optimal transport problem with prior linear dynamics was also considered in [46]. In our setting, we show that the controllability result for the steering between (possibly non-Gaussian) densities can be obtained as a direct byproduct of the existence of optimal mass transport maps.

The present work begins with an expository prologue on OMT (Section II). We then develop the theory of OMT-wpd (Section III) and establish that OMT-wpd always has a unique solution. Next we discuss in parallel the theory of the Schödinger bridge problem for linear dynamics and arbitrary end-point marginals (Section IV). We focus on the connection between the two problems and in Theorem 3 we establish that the solution to the OMT-wpd is indeed the limit, in a suitable sense, of the corresponding solution to the Schrödinger bridge problem. In Section V we specialize to the case of linear dynamics with Gaussian marginals, where closed-form solutions are available for both problems. The form of solution underscores the connection between the two and that the OMT-wpd is the limit of the Schrödinger bridge problem when the diffusion term vanishes. In Section VI we work out two academic examples to highlight the relation between the two problems (OMT and Schrödinger bridge).

II. OPTIMAL MASS TRANSPORT

Consider two nonnegative measures μ_0, μ_1 on $\mathbb{R}^n$ having equal total mass. These may represent probability distributions, distribution of resources, etc. Without loss of generality, we take μ_0 and μ_1 to be probability distributions in this paper. In the original formulation of OMT, due to Gaspard Monge, a transport (measurable) map

$$T \: : \: \mathbb{R}^n \to \mathbb{R}^n \: : \: x \mapsto T(x)$$

is sought that specifies where mass $\mu_0(dx)$ at x must be transported so as to match the final distribution in the sense that $T_\sharp \mu_0 = \mu_1$, i.e. μ_1 is the "push-forward" of μ_0 under T meaning

$$\mu_1(B) = \mu_0(T^{-1}(B))$$

for every Borel set in $\mathbb{R}^n$. Moreover, the map must incur minimum cost of transportation

$$\int c(x, T(x))\mu_0(dx).$$

Here, $c(x, y)$ represents the transportation cost per unit mass from point x to point y and in this section it will be taken as $c(x, y) = \frac{1}{2}\|x - y\|^2$.

The dependence of the transportation cost on T is highly nonlinear and a minimum may not exist. This fact complicated early analyses to the problem due to Abel and others [1]. A new chapter opened in 1942 when Leonid Kantorovich presented a relaxed formulation. In this, instead of seeking a transport map, we seek a joint distribution $\Pi(\mu_0, \mu_1)$ on the product space $\mathbb{R}^n \times \mathbb{R}^n$ so that the marginals along the two coordinate directions coincide with μ_0 and μ_1 respectively. The joint distribution $\Pi(\mu_0, \mu_1)$ is refered to as "coupling" of μ_0 and μ_1[1]. Thus, in the Kantorovich formulation we

---

[1]Throughout the paper it will be tacitly assumed that μ_0 and μ_1 possess a finite second moment.

---

seek

$$\inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{2}\|x - y\|^2 \pi(dxdy). \tag{3}$$

When the optimal Monge-map T exists, the support of the coupling is precisely the graph of T, see [1].

Formulation (3) represents a "static" end-point formulation, i.e., focusing on "what goes where". Ingenious insights due to Benamou and Brenier [2] and [47] led to a fluid dynamic formulation of OMT. An elementary derivation of the above was presented in [12] which we now follow. OMT is first cast as a stochastic control problem with atypical boundary constraints:

$$\inf_{v \in \mathcal{V}} \mathbb{E}\left\{ \int_0^1 \frac{1}{2}\|v(t, x^v(t))\|^2 dt \right\}, \tag{4a}$$

$$\dot{x}^v(t) = v(t, x^v(t)), \tag{4b}$$

$$x^v(0) \sim \mu_0, \quad x^v(1) \sim \mu_1. \tag{4c}$$

Here $\mathcal{V}$ represents the family of admissible Markov feedback control laws. We call a control law $v(t, x)$ admissible if the corresponding controlled system (4b) has a unique solution for almost every deterministic initial condition at $t = 0$. Requiring $v(t, \cdot)$ to be uniformly Lipschitz continuous on $[0, 1]$ is a sufficient condition, but it is not necessary.

From this point on we assume that μ_0 and μ_1 are absolutely continuous, i.e., $\mu_0(dx) = \rho_0(x)dx$, $\mu_1(dy) = \rho_1(y)dy$ with ρ_0, ρ_1 corresponding density functions. If $x^v(t)$ also has a absolutely continuous distribution, namely, $x^v(t) \sim \rho(t, x)dx$, then ρ satisfies weakly[2] the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0 \tag{5}$$

expressing the conservation of probability mass, where $\nabla \cdot$ denotes the divergence of a vector field, and

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

As a consequence, (4) is recast as a "fluid-dynamics" problem [2]:

$$\inf_{(\rho, v)} \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2}\|v(t, x)\|^2 \rho(t, x)dtdx, \tag{6a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \tag{6b}$$

$$\rho(0, x) = \rho_0(x), \quad \rho(1, y) = \rho_1(y). \tag{6c}$$

The minimum is taken over all the pairs ρ, v satisfying (6b)-(6c) and other technical assumptions, see [1, Theorem 8.1], [8, Chapter 8].

A. Solutions to OMT

For the case where μ_0, μ_1 are absolutely continuous ($\mu_0(dx) = \rho_0(x)dx$ and $\mu_1(dy) = \rho_1(y)dy$) it is a standard result that OMT has a unique solution [48], [1], [9] and that an optimal transport T map exists and is the gradient of a convex function ϕ, i.e.,

$$y = T(x) = \nabla\phi(x). \tag{7}$$

By virtue of the fact that the push-forward of μ_0 under $\nabla\phi$ is μ_1, this function satisfies a particular case of the Monge-Ampère equation [1, p.126], [2, p.377], namely, $\det(H\phi(x))\rho_1(\nabla\phi(x)) = \rho_0(x)$, where

---

[2]In the sense that $\int_{[0,1] \times \mathbb{R}^n} [(\partial f/\partial t + v \cdot \nabla f)\rho]dtdx = 0$ for smooth functions f with compact support.

$H\phi$ is the Hessian matrix of ϕ, which is a fully nonlinear second-order elliptic equation. The computation of ϕ has received attention only recently [2], [14], [49] where numerical schemes have been developed.

Having the optimal mass transport map T, as in (7), the optimal coupling is

$$\pi = (\mathrm{Id} \times T)_\sharp \mu_0,$$

where Id stands for the identity map, and the displacement of the mass along the path from $t = 0$ to $t = 1$ is

$$\mu_t = (T_t)_\sharp \mu_0, \quad T_t(x) = (1-t)x + tT(x) \tag{8a}$$

while μ_t is absolutely continuous with Radon-Nikodym derivative with respect to the Lebesgue measure

$$\rho(t, x) = \frac{d\mu_t}{dx}(x). \tag{8b}$$

Accordingly, the optimal control strategy of (4) is given by

$$v(t, x) = T \circ T_t^{-1}(x) - T_t^{-1}(x), \tag{9}$$

and the pair ρ, v solves (6). Here $\circ$ denotes the composition of maps. By (8a) T_t is the gradient of a uniformly convex function for $0 \leq t < 0$, so T_t is injective and therefore (9) is well-defined on $T_t(\mathbb{R}^n)$. The values $v(t, x)$ outside $T_t(\mathbb{R}^n)$ do not play any role.

An alternative expression for the optimal control (9) can be established using standard optimal control theory, and this is summarized in the following proposition [1, Theorem 5.51].

Proposition 1: Given marginal distributions $\mu_0(dx) = \rho_0(x)dx, \mu_1(dx) = \rho_1(x)dx$, let $\psi(t, x)$ be defined by the Hopf-Lax representation

$$\psi(t, x) = \inf_y \left\{ \psi(0, y) + \frac{\|x - y\|^2}{2t} \right\}, \quad t \in (0, 1]$$

with

$$\psi(0, x) = \phi(x) - \frac{1}{2}\|x\|^2$$

and ϕ as in (7). Then $v(t, x) := \nabla \psi(t, x)$ exists almost everywhere and it solves (4).

III. Optimal Mass Transport With Prior Dynamics

The OMT problem

$$\inf_{T: \mu_1 = T_\sharp \mu_0} \int_{\mathbb{R}^n} c(x, T(x))\mu_0(dx), \tag{10}$$

and the relaxed version, namely, the Monge-Kantorovich problem

$$\inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y)\pi(dxdy) \tag{11}$$

have also been studied for general cost $c(x, y)$ that derives from an action functional

$$c(x, y) = \inf_{x(\cdot) \in \mathcal{X}_{xy}} \int_0^1 L(t, x(t), \dot{x}(t))dt, \tag{12}$$

where the Lagrangian $L(t, x, p)$ is strictly convex and superlinear in the velocity variable p, see [9, Chapter 7], [50, Chapter 1], [51] and $\mathcal{X}_{xy}$ is the family of absolutely continuous paths with $x(0) = x$ and $x(1) = y$. Existence and uniqueness of an optimal transport map T has been established[3] for general cost functionals as in (12). It is

easy to see that the choice $c(x, y) = \frac{1}{2}\|x - y\|^2$ is the special case where $L(t, x, p) = \frac{1}{2}\|p\|^2$. Another interesting special case is when

$$L(t, x, p) = \frac{1}{2}\|p - v(t, x)\|^2. \tag{13}$$

This has been motivated by a transport problem "with prior" associated to the velocity field $v(t, x)$ [12, Section VII]. There the prior was thought to reflect a solution to a "nearby" problem that needs to be adjusted so as to be consistent with updated estimates for marginals.

An alternative motivation for (13) is to address transport in an ambient flow field $v(t, x)$. In this case, assuming the control has the ability to steer particles in all directions, transport will be effected according to dynamics $\dot{x}(t) = v(t, x) + u(t)$ where $u(t)$ represents control effort and

$$\int_0^1 \frac{1}{2}\|u(t)\|^2 dt = \int_0^1 \frac{1}{2}\|\dot{x}(t) - v(t, x)\|^2 dt$$

represents corresponding quadratic cost (energy). Thus, it is of interest to consider more general dynamics where the control does not affect directly all state directions. One such example is the problem to steer inertial particles in phase space through force input (see [37] and [38] where similar problems have been considered for dynamical systems with stochastic excitation).

Therefore, herein, we consider a natural generalization of OMT where the transport paths are required to satisfy dynamical constraints. We focus our attention on linear dynamics and, consequently, cost of the form

$$c(x, y) = \inf_{u \in \mathcal{U}} \int_0^1 \tilde{L}(t, x(t), u(t))dt, \text{ where} \tag{14a}$$
$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \tag{14b}$$
$$x(0) = x, \quad x(1) = y, \tag{14c}$$

and $\mathcal{U}$ is a suitable class of controls[4]. This formulation extends the transportation problem in a similar manner as optimal control generalizes the classical calculus of variations [52] (albeit herein only for linear dynamics). It is easy to see that (13) corresponds to $A(t) = 0$ and $B(t)$ the identity matrix in (14). When $B(t)$ is invertible, (14) reduces to (12) by a change of variables, taking

$$L(t, x, p) = \tilde{L}(t, x, B(t)^{-1}(p - A(t)x)).$$

However, when $B(t)$ is not invertible, an analogous change of variables leads to a Lagrangian $L(t, x, p)$ that fails to satisfy the classical conditions (strict convexity and superlinearity in p). Therefore, in this case, the existence and uniqueness of an optimal transport map T has to be established independently. We do this for the case where $\tilde{L}(t, x, u) = \|u\|^2/2$ corresponding to power. A similar problem has been studied in [46].

We now formulate the corresponding stochastic control problem. The system dynamics

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \tag{15}$$

are assumed to be controllable and the initial state $x(0)$ is a random vector with probability density ρ_0. Here, A and B are continuous maps taking values in $\mathbb{R}^{n \times n}$ and $\mathbb{R}^{n \times m}$, repectively. In the stochastic control formulation, we seek a minimum energy feedback control law $u(t, x)$ that steers the system to a final state $x(1)$ having distribution

---

[3]OMT has also been studied and similar results established for $\mathbb{R}^n$ replaced by a Riemannian manifold.

[4]Note that we use a common convention to denote by x a point in the state space and by $x(t)$ a state trajectory.

$\rho_1(x)dx$. That is, we address the following:

$$\inf_{u \in \mathcal{U}} \mathbb{E} \left\{ \int_0^1 \frac{1}{2} \|u(t, x^u)\|^2 dt \right\}, \tag{16a}$$

$$\dot{x}^u(t) = A(t)x^u(t) + B(t)u(t, x^u(t)), \tag{16b}$$

$$x^u(0) \sim \rho_0(x)dx, \quad x^u(1) \sim \rho_1(y)dy, \tag{16c}$$

where $\mathcal{U}$ is the family of admissible Markov feedback control laws. Recall that we call a control law $u(t, x)$ admissible if the corresponding controlled system (16b) has a unique solution for almost every deterministic initial condition at $t = 0$.

We next show that (16) is indeed a reformulation of (11) with generalized cost (14) when $\tilde{L}(t, x, u) = \|u\|^2/2$. First we note the cost is equal to

$$c(x, y) = \min_{x(\cdot) \in \mathcal{X}_{xy}} \int_0^1 \hat{L}(t, x(t), \dot{x}(t))dt, \tag{17}$$

where

$$\hat{L}(t, x, v) = \begin{cases} \frac{1}{2}(v - A(t)x)'(B(t)B(t)')^\dagger (v - A(t)x), \\ \quad \text{if } v - A(t)x \in \mathcal{R}(B(t)), \\ \infty \text{ otherwise} \end{cases}$$

with $\dagger$ denoting pseudo-inverse and $\mathcal{R}(\cdot)$ "the range of". If the minimizer of (17) exists, which will be denoted as $x^*(\cdot)$, then any probabilistic average of the action relative to absolutely continuous trajectories starting at x at time 0 and ending in y at time 1 cannot give a lower value. Thus, the probability measure on $\mathcal{X}_{xy}$ concentrated on the path $x^*(\cdot)$ solves the following problem

$$\inf_{P_{xy} \in \mathbb{D}(\delta_x, \delta_y)} \mathbb{E}_{P_{xy}} \left\{ \int_0^1 \hat{L}(t, x(t), \dot{x}(t))dt \right\}, \tag{18}$$

where $\mathbb{D}(\delta_x, \delta_y)$ are the probability measures on $C[0, 1]$ paths whose initial and final one-time marginals are Dirac's deltas concentrated at x and y, respectively.

Let u be a feasible control strategy in (16), and $x^u(\cdot)$ be the corresponding controlled process. This process induces a probability measure P in $\mathbb{D}(\mu_0, \mu_1)$, namely a measure on the path space $C[0, 1]$ whose one-time marginals at 0 and 1 are μ_0 and μ_1, respectively. The measure P can be disintgrated as [42], [8]

$$P = \int_{\mathbb{R}^n \times \mathbb{R}^n} P_{xy} \, \pi(dxdy), \tag{19}$$

where $P_{xy} \in \mathbb{D}(\delta_x, \delta_y)$ and $\pi \in \Pi(\mu_0, \mu_1)$. Then the control energy in (16) is greater than or equal to

$$\mathbb{E}_P \left\{ \int_0^1 \hat{L}(t, x(t), \dot{x}(t))dt \right\}$$

$$= \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathbb{E}_{P_{xy}} \left\{ \int_0^1 \hat{L}(t, x(t), \dot{x}(t))dt \right\} \pi(dxdy)$$

$$\geq \int_{\mathbb{R}^n \times \mathbb{R}^n} c(x, y)\pi(dxdy), \tag{20}$$

which shows that the minimum of (16) is bounded below by the minimum of (11) with cost in (14) or equivalently (17). In the next subsection we will construct a control strategy such that the joint measure π in (19) solves (11) and P_{xy} is concentrated on the path $x^*(\cdot)$ for π-almost every pair of intial position x and terminal position y. Therefore, the stochastic optimal control problem (16) is indeed a reformulation of the OMT (11) with the general cost in (14), and we refer to both of them as OMT-wpd.

Once again, formally, the stochastic control formulation (16) suggests the "fluid-dynamics" version:

$$\inf_{(\rho, u)} \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2} \|u(t, x)\|^2 \rho(t, x)dtdx, \tag{21a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot ((A(t)x + B(t)u)\rho) = 0, \tag{21b}$$

$$\rho(0, x) = \rho_0(x), \quad \rho(1, y) = \rho_1(y). \tag{21c}$$

Establishing rigorously the equivalence between (21) and OMT-wpd (16) is a delicate issue. We expect the equivalence can be shown along the lines of [1, Theorem 8.1], [8, Chapter 8], but this is not our focus in the present paper.

Naturally, for the trivial prior dynamics $A(t) \equiv 0$ and $B(t) \equiv I$, the OMT-wpd reduces to the classical OMT and the solution $\{\rho(t, \cdot) \mid 0 \leq t \leq 1\}$ is the *displacement interpolation* of the two marginals [47]. In the next subsection, we show directly that Problem (16) has a unique solution.

A. Solutions to OMT-wpd

Let $\Phi(t, s)$ be the state transition matrix of (15) from s to t, and

$$M(t, s) = \int_s^t \Phi(t, \tau)B(\tau)B(\tau)'\Phi(t, \tau)'d\tau \tag{22}$$

be the controllability Gramian of the system which, by the controllability assumption, is positive definite for all $0 \leq s < t \leq 1$; we denote $\Phi_{10} := \Phi(1, 0)$ and $M_{10} := M(1, 0)$. Recall [53], [54] that for linear dynamics (15) and given boundary conditions $x(0) = x$, $x(1) = y$, the least energy $c(x, y)$ and the corresponding optimal control input can be given in closed-form, namely

$$c(x, y) = \int_0^1 \frac{1}{2}\|u^*(t)\|^2 dt = \frac{1}{2}(y - \Phi_{10}x)'M_{10}^{-1}(y - \Phi_{10}x) \tag{23}$$

where

$$u^*(t) = B(t)'\Phi(1, t)'M_{10}^{-1}(y - \Phi_{10}x). \tag{24}$$

The corresponding optimal trajectory is

$$x^*(t) = \Phi(t, 1)M(1, t)M_{10}^{-1}\Phi_{10}x + M(t, 0)\Phi(1, t)'M_{10}^{-1}y. \tag{25}$$

The OMT-wpd problem with this cost is

$$\inf_\pi \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{2}(y - \Phi_{10}x)'M_{10}^{-1}(y - \Phi_{10}x)\pi(dxdy), \tag{26a}$$

$$\pi(dx \times \mathbb{R}^n) = \rho_0(x)dx, \ \pi(\mathbb{R}^n \times dy) = \rho_1(y)dy, \tag{26b}$$

where π is a measure on $\mathbb{R}^n \times \mathbb{R}^n$.

Problem (26) can be converted to the standard Kantorovich formulation (3) of the OMT by a transformation of coordinates. Indeed, consider the linear map

$$C : \begin{bmatrix} x \\ y \end{bmatrix} \longrightarrow \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} M_{10}^{-1/2}\Phi_{10}x \\ M_{10}^{-1/2}y \end{bmatrix} \tag{27}$$

and set $\hat{\pi} = C_\sharp \pi$. Clearly, (26a-26b) become

$$\inf_{\hat{\pi}} \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{2}\|\hat{y} - \hat{x}\|^2 \hat{\pi}(d\hat{x}d\hat{y}), \tag{28a}$$

$$\hat{\pi}(d\hat{x} \times \mathbb{R}^n) = \hat{\rho}_0(\hat{x})d\hat{x}, \ \hat{\pi}(\mathbb{R}^n \times d\hat{y}) = \hat{\rho}_1(\hat{y})d\hat{y}, \tag{28b}$$

where

$$\hat{\rho}_0(\hat{x}) = |M_{10}|^{1/2}|\Phi_{10}|^{-1}\rho_0(\Phi_{10}^{-1}M_{10}^{1/2}\hat{x}),$$

$$\hat{\rho}_1(\hat{y}) = |M_{10}|^{1/2}\rho_1(M_{10}^{1/2}\hat{y}).$$

Problem (28) is now a standard OMT with quadratic cost function and we know that the optimal transport map $\hat{T}$ for this problem exists [1]. It is the gradient of a convex function ϕ, i.e.,

$$\hat{T} = \nabla\phi, \tag{29}$$

and the optimal $\hat{\pi}$ is concentrated on the graph of $\hat{T}$ [48]. The solution to the original problem (26) can now be determined using $\hat{T}$, and it is

$$\pi = (\mathrm{Id} \times T)_\sharp \mu_0$$

with

$$y = T(x) = M_{10}^{1/2}\hat{T}(M_{10}^{-1/2}\Phi_{10}x). \tag{30}$$

From the above argument we can see that, with cost function (14) and $\tilde{L}(t,x,u) = \|u\|^2/2$, the OMT problem (10) and its relaxed version (11) are equivalent.

The map T, together with the optimal trajectories in (25) for fixed end-points, leads to the one-time marginals

$$\mu_t = (T_t)_\sharp\mu_0, \tag{31a}$$

where

$$T_t(x) = \Phi(t,1)M(1,t)M_{10}^{-1}\Phi_{10}x + M(t,0)\Phi(1,t)'M_{10}^{-1}T(x), \tag{31b}$$

and

$$\rho(t,x) = \frac{d\mu_t}{dx}(x). \tag{31c}$$

Note that (31b) generalizes (8a) to the present setting. In this case, we refer to the parametric family of one-time marginals as *displacement interpolation with prior dynamics*. Combining the optimal map T with (24), we obtain the optimal control strategy

$$u(t,x) = B(t)'\Phi(1,t)'M_{10}^{-1}[T \circ T_t^{-1}(x) - T_t^{-1}(x)]. \tag{32}$$

Again T_t is injective for $0 \leq t < 1$, so the above control strategy is well-defined on $T_t(\mathbb{R}^n)$. $T_0 = \mathrm{Id}$ is of course injective. To see T_t is an injection for $0 < t < 1$, assume that there are two different points $x \neq y$ such that $T_t(x) = T_t(y)$. Then

$$
\begin{aligned}
0 &= (x-y)'\Phi(t,0)'M(t,0)^{-1}(T_t(x) - T_t(y))' \\
&= (x-y)'\Phi(t,0)'M(t,0)^{-1}\Phi(t,1)M(1,t)M_{10}^{-1}\Phi_{10}(x-y) + \\
&\quad (x-y)'\Phi_{10}'M_{10}^{-1/2}(\nabla\phi(M_{10}^{-1/2}\Phi_{10}x) - \nabla\phi(M_{10}^{-1/2}\Phi_{10}y)).
\end{aligned}
$$

The second term is nonnegative due to the convexity of ϕ. The first term is equal to

$$(x-y)'\left(\Phi(t,0)'M(t,0)^{-1}\Phi(t,0) - \Phi_{10}'M_{10}^{-1}\Phi_{10}\right)(x-y),$$

which is positive since

$$
\begin{aligned}
&\Phi(t,0)'M(t,0)^{-1}\Phi(t,0) - \Phi_{10}'M_{10}^{-1}\Phi_{10} \\
&= \left(\int_0^t \Phi(0,\tau)B(\tau)B(\tau)'\Phi(0,\tau)'d\tau\right)^{-1} \\
&\quad - \left(\int_0^1 \Phi(0,\tau)B(\tau)B(\tau)'\Phi(0,\tau)'d\tau\right)^{-1}
\end{aligned}
$$

is positive definite for all $0 < t < 1$. Since the control (32) is consistent with both the optimal coupling π and the optimal trajectories (25), it achieves the minimum of (11), which is of course is the minimum of (16) based on (20).

An alternative expression for the optimal control (32) can be derived as follows using standard optimal control theory. Consider the following deterministic optimal control problem

$$\inf_{u \in \mathcal{U}} \int_0^1 \frac{1}{2}\|u(t,x^u)\|^2 dt - \psi_1(x^u(1)), \tag{33a}$$

$$\dot{x}^u(t) = A(t)x^u(t) + B(t)u(t) \tag{33b}$$

for some terminal cost $-\psi_1$. The dynamic programming principle [55] gives the value function (cost-to-go function) $-\psi(t,x)$ as

$$-\psi(t,x) = \inf_{u:x(t)=x} \int_t^1 \frac{1}{2}\|u(t,x^u)\|^2 dt - \psi_1(x^u(1)). \tag{34}$$

The associated dynamic programming equation is

$$\inf_{u\in\mathbb{R}^m}\left[\frac{1}{2}\|u\|^2 - \frac{\partial\psi}{\partial t} - \nabla\psi \cdot (A(t)x + B(t)u)\right] = 0. \tag{35}$$

Point-wise minimization yields the Hamilton-Jacobi-Bellman equation

$$\frac{\partial\psi}{\partial t} + x'A(t)'\nabla\psi + \frac{1}{2}\nabla\psi'B(t)B(t)'\nabla\psi = 0 \tag{36a}$$

with boundary condition

$$\psi(1,y) = \psi_1(y), \tag{36b}$$

and the corresponding optimal control is

$$u(t,x) = B(t)'\nabla\psi(t,x). \tag{37}$$

When the value function $-\psi(t,x)$ is smooth, it solves the Hamilton-Jacobi-Bellman equation (36). In this case, if the optimal control (37) drives the controlled process from initial distribution μ_0 to terminal distribution μ_1, then this u in fact solves the OMT-wpd (16). In general, one cannot expect (36) to have a classical solution and has to be content with viscosity solutions [56], [55]. Typically, one can prove existence by including a vanishingly small regularization term with a Laplacian [57, Section 10.1]. Here, however, it is possible to give an explicit expression for the value function based only on the dynamic programming principle (34). This is summarized in the following proposition.

Proposition 2: Given marginal distributions $\mu_0(dx) = \rho_0(x)dx, \mu_1(dx) = \rho_1(x)dx$, let $\psi(t,x)$ be defined by the formula

$$
\begin{aligned}
\psi(t,x) = \inf_y \Big\{ &\psi(0,y) + \\
&\frac{1}{2}(x - \Phi(t,0)y)'M(t,0)^{-1}(x - \Phi(t,0)y)\Big\}
\end{aligned}
\tag{38}
$$

with

$$\psi(0,x) = \phi(M_{10}^{-1/2}\Phi_{10}x) - \frac{1}{2}x'\Phi_{10}'M_{10}^{-1}\Phi_{10}x$$

and ϕ as in (29). Then $u(t,x) := B(t)'\nabla\psi(t,x)$ exists almost everywhere and it solves (16).

Proof: The proof is given in Appendix A. ∎

IV. SCHRÖDINGER BRIDGES AND THEIR ZERO-NOISE LIMIT

In 1931/32, Schrödinger [17], [18] treated the following problem: A large number N of i.i.d. Brownian particles in $\mathbb{R}^n$ is observed to have at time $t = 0$ an empirical distribution approximately equal to $\rho_0(x)dx$, and at some later time $t = 1$ an empirical distribution approximately equal to $\rho_1(x)dx$. Suppose that $\rho_1(x)$ considerably differs from what it should be according to the law of large numbers, namely

$$\int q^B(0,x,1,y)\rho_0(x)dx,$$

where

$$q^B(s,x,t,y) = (2\pi)^{-n/2}(t-s)^{-n/2}\exp\left(-\frac{\|x-y\|^2}{2(t-s)}\right)$$

denotes the Brownian transition probability density. It is apparent that the particles have been transported in an unlikely way. But of

the many unlikely ways in which this could have happened, which one is the most likely? In view of Sanov's theorem, see Föllmer [23], Schrödinger's question reduces to determining a probability law $\mathcal{P}(\cdot)$ on $C[0,1]$, the continuous paths over the specified interval, that minimizes the relative entropy

$$\mathcal{S}(\mathcal{P},\mathcal{Q}) := \int_{C[0,1]} \log\left(\frac{d\mathcal{P}}{d\mathcal{Q}}\right) d\mathcal{P}.$$

Here $\mathcal{Q}(\cdot)$ is the probability law induced by the prior Markovian evolution and $\mathcal{P}(\cdot)$ is chosen among probability laws that are absolutely continuous with respect to $\mathcal{Q}(\cdot)$ and have the prescribed marginals[5]. Evidently, this is an abstract problem on an infinite-dimensional space. The solution to this optimization problem is referred to as the *Schrödinger bridge*. Existence of the minimizer has been proven in various degrees of generality by Fortet [20], Beurling [21], Jamison [22], Föllmer [23]. Jamison's result is stated below.

Theorem 1: Given two probability measures $\mu_0(dx) = \rho_0(x)dx$ and $\mu_1(dy) = \rho_1(y)dy$ on $\mathbb{R}^n$ and the continuous, everywhere positive Markov kernel $q(s,x,t,y)$, there exists a unique pair of σ-finite measure $(\hat{\varphi}_0(x)dx, \varphi_1(x)dx)$ on $\mathbb{R}^n$ such that the measure $\mathcal{P}_{01}$ on $\mathbb{R}^n \times \mathbb{R}^n$ defined by

$$\mathcal{P}_{01}(E) = \int_E q(0,x,1,y)\hat{\varphi}_0(x)\varphi_1(y)dxdy \tag{39}$$

has marginals μ_0 and μ_1. Furthermore, the Schrödinger bridge from μ_0 to μ_1 induces the distribution flow

$$\mathcal{P}_t(dx) = \varphi(t,x)\hat{\varphi}(t,x)dx \tag{40a}$$

with

$$\varphi(t,x) = \int q(t,x,1,y)\varphi_1(y)dy \tag{40b}$$

$$\hat{\varphi}(t,x) = \int q(0,y,t,x)\hat{\varphi}_0(y)dy. \tag{40c}$$

The flow (40) is referred to as the *entropic interpolation with prior q* between μ_0 and μ_1, or simply entropic interpolation, when it is clear what the Markov kernel q is. An efficient numerical algorithm to obtain the pair $(\hat{\varphi}_0, \varphi_1)$ and thereby solve the Schrödinger bridge problem is given in [58].

For the case of non-degenerate Markov processes, a connection between the Schrödinger problem and stochastic control was drawn in [16], see also [35] and [59]. In particular, for the case of a Brownian kernel, it was shown there that the one-time marginals $\rho(t,x)$ for Schrödinger's problem are the densities of the optimal state vector in the stochastic control problem

$$\inf_{v\in\mathcal{V}} \mathbb{E}\left\{\int_0^1 \frac{1}{2}\|v(t,x^v)\|^2 dt\right\}, \tag{41a}$$

$$dx^v(t) = v(t,x^v(t))dt + dw(t), \tag{41b}$$

$$x^v(0) \sim \rho_0(x)dx, \quad x^v(1) \sim \rho_1(y)dy. \tag{41c}$$

Here $\mathcal{V}$ is the class of finite energy Markov controls. Using a localization argument [60, p. 98] one can prove that for every $v \in \mathcal{V}$ the controlled process has a weak solution in $[0,1]$ [61]. This reformulation of the Schrödinger problem builds on the fact that the relative entropy between x^v and x^0 (zero control) in (41b) is bounded above by the control energy, namely,

$$\mathcal{S}(\mathcal{P}_{x^v},\mathcal{P}_{x^0}) \leq \mathbb{E}\left\{\int_0^1 \frac{1}{2}\|v(t,x^v(t))\|^2 dt\right\},$$

[5] In the above, $\frac{d\mathcal{P}}{d\mathcal{Q}}$ denotes the Radon-Nikodym derivative.

where $\mathcal{P}_{x^v}, \mathcal{P}_{x^0}$ denote the measures induced by x^v and x^0, respectively. The proof is based on Girsanov theorem, see [16], [61]. The optimal control to (41) is given by

$$v(t,x) = \nabla \log \varphi(t,x)$$

with φ in (40b), see [16]. The stochastic control problem (41) has the following equivalent formulation [39], [62]:

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2}\|v(t,x)\|^2 \rho(t,x)dtdx, \tag{42a}$$

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(v\rho) - \frac{1}{2}\Delta\rho = 0, \tag{42b}$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y). \tag{42c}$$

Here, the infimum is over smooth fields v and ρ solves weakly of the corresponding Fokker-Planck equation (42b). The entropic interpolation is $\mathcal{P}_t(dx) = \rho(t,x)dx$.

An alternative equivalent reformulation given in [12] is

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2}\|v(t,x)\|^2 + \frac{1}{8}\|\nabla\log\rho(t,x)\|^2\right]\rho(t,x)dtdx, \tag{43a}$$

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(v\rho) = 0, \tag{43b}$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y), \tag{43c}$$

where the Laplacian in the dynamical constraint is traded for a "Fisher information" regularization term in the cost functional. It answers at once a question posed by E. Carlen in 2006 investigating the connections between optimal transport and Nelson's stochastic mechanics [63]. Although the form in (43) is quite appealing, for the purposes of this paper we will use only (42).

Formulation (42) is quite similar to OMT (6) except for the presence of the Laplacian in (42b). It has been shown [43], [44], [41], [42] that the OMT problem is, in a suitable sense, indeed the limit of the Schrödinger problem when the diffusion coefficient of the reference Brownian motion goes to zero. In particular, the minimizers of the Schrödinger problems converge to the unique solution of OMT, see below.

Theorem 2: Given two probability measures $\mu_0(dx) = \rho_0(x)dx, \mu_1(dy) = \rho_1(y)dy$ on $\mathbb{R}^n$ with finite second moment, let $\mathcal{P}_{01}^{B,\epsilon}$ be the solution of the Schrödinger problem with Markov kernel

$$q^{B,\epsilon}(s,x,t,y) = (2\pi)^{-n/2}((t-s)\epsilon)^{-n/2}\exp\left(-\frac{\|x-y\|^2}{2(t-s)\epsilon}\right) \tag{44}$$

and marginals μ_0,μ_1, and let $\mathcal{P}_t^{B,\epsilon}$ be the corresponding entropic interpolation. Similarly, let π be the solution to the OMT problem (3) with the same marginal distributions, and μ_t the corresponding displacement interpolation. Then, $\mathcal{P}_{01}^{B,\epsilon}$ converges weakly[6] to π and $\mathcal{P}_t^{B,\epsilon}$ converges weakly to μ_t, as ϵ goes to 0.

To build some intuition on the relation between OMT and Schrödinger bridges, consider $dx(t) = \sqrt{\epsilon}dw(t)$ with $w(t)$ being the standard Wiener process; the Markov kernel of $x(t)$ is $q^{B,\epsilon}$ in (44). The corresponding Schrödinger bridge problem with the law of

[6] A sequence $\{P_n\}$ of probability measures on a metric space $\mathcal{S}$ converges weakly to a measure P if $\int_{\mathcal{S}} fdP_n \to \int_{\mathcal{S}} fdP$ for every bounded, continuous function f on the space.

$x(t)$ as prior, is equivalent to

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2\epsilon} \|v(t,x)\|^2 \rho(t,x) dt dx, \qquad (45a)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) - \frac{\epsilon}{2} \Delta \rho = 0, \qquad (45b)$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y). \qquad (45c)$$

Note that the solution exists for all ϵ and coincides with the solution of the problem to minimize the cost functional

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

instead, i.e., "rescaling" (45a) by removing the factor $1/\epsilon$. Now observe that the only difference between (45) after removing the scaling $1/\epsilon$ in the cost functional and the OMT formulation (6) is the regularization term $\frac{\epsilon}{2} \Delta \rho$ in (45b). Thus, formally, the constraint (45b) becomes (6b) as ϵ goes to 0.

Below we discuss a general result that includes the case when the zero-noise limit of Schrödinger bridges corresponds to OMT-wpd. This problem has been studied in [41] in a more abstract setting based on Large Deviation Theory [64]. Here we consider the special case that is connected to our OMT-wpd formulation. To this end, we begin with the Markov kernel corresponding to the process

$$dx(t) = A(t)x(t)dt + \sqrt{\epsilon}B(t)dw(t). \qquad (46)$$

Notice that the corresponding transition kernel (see (62) in Appendix B) is everywhere positive because of the controllability assumption. Moreover, $\varphi(t,x)$ for $0 \le t < 1$ satisfying (40b) is also everywhere positive and smooth.

Motivated by the previous case, we consider the following stochastic control problem,

$$\inf_{u \in \mathcal{U}} \mathbb{E}\left\{ \int_0^1 \frac{1}{2} \|u(t,x^u)\|^2 dt \right\}, \qquad (47a)$$

$$dx^u(t) = A(t)x^u(t)dt + B(t)u(t,x^u)dt + \sqrt{\epsilon}B(t)dw(t), \quad (47b)$$

$$x^u(0) \sim \rho_0(x)dx, \quad x^u(1) \sim \rho_1(y)dy. \qquad (47c)$$

Here, $\mathcal{U}$ is the set of admissible Markov controls such that for each $u \in \mathcal{U}$ the controlled process admits a weak solution in $[0,1]$ and the control has finite energy. By a general version of Girsanov theorem [65, Chapter IV.4] and the contraction property of relative entropy [66], we have

$$\mathcal{S}(\mathcal{P}_{x^u}, \mathcal{P}_{x^0}) \le \mathbb{E}\left\{ \int_0^1 \frac{1}{2\epsilon} \|u(t,x^u(t))\|^2 dt \right\},$$

where $\mathcal{P}_{x^u}, \mathcal{P}_{x^0}$ denote the measures induced by x^u and x^0 (zero control) in (47b), respectively.

Let $\varphi, \hat{\varphi}$ be as in (40b) with the Markov kernel corresponding to (46) (and given in (62) in Appendix B). We claim that, under the technical assumptions that i) $\int_{\mathbb{R}^n} \varphi(0,x)\mu_0(dx) < \infty$ and ii) $\mathcal{S}(\mu_1, \hat{\varphi}(1,\cdot)) < \infty$, the optimal solution to (47) is

$$u(t,x) = \epsilon B(t)' \nabla \log \varphi(t,x). \qquad (48)$$

The assumption i) guarantees that the local martingale $\varphi(t,x(t))$, where x is the uncontrolled evolution (46), is actually a martingale. The assumption ii) implies that the control (48) has finite energy. For both statements see [16, Theorem 3.2], whose proof carries verbatim. While these conditions i) and ii) are difficult to verify in general, they are satisfied when both μ_0 and μ_1 have compact support (c.f. [16, Proposition 3.1]).

Then, by the argument in [16, Theorem 2.1] and in view of the equivalence between existence of weak solutions to stochastic differential equations (SDEs) and solutions to the martingale problem (see [67, Theorem 4.2], [61, p. 314]), it follows that with $u(t,x)$ as in (48) the SDE (47b) has a weak solution. By substituting (48) in the Fokker-Planck equation it can be seen that the corresponding controlled process satisfies the marginals (47c). In fact, the density flow ρ coincides with the one-time marginals of the Schrödinger bridge (40a).

Finally, to see that (48) is optimal, we use a completion of squares argument. To this end, consider the equivalent problem of minimizing the cost functional

$$J(u) = \mathbb{E}\left\{ \int_0^1 \frac{1}{2\epsilon} \|u(t)\|^2 dt - \log \varphi(1,x(1)) + \log(\varphi(0,x(0))) \right\}$$

in (47a) (the boundary terms are constant over the admissible path-space probability distributions, cf. (47c)). Using Ito's rule, and the fact that u in (48) has finite energy, a standard calculation [68] shows

$$\begin{aligned} J(u) &= \mathbb{E}\left\{ \int_0^1 \frac{1}{2\epsilon} \|u(t)\|^2 dt - d\log \varphi(t,x(t)) \right\} \\ &= \mathbb{E}\left\{ \int_0^1 \frac{1}{2\epsilon} \|u(t) - \epsilon B(t)' \nabla \log \varphi(t,x(t))\|^2 dt \right\}, \end{aligned}$$

from which we readily conclude that (48) is the optimal control law.

The entropic interpolation $\mathcal{P}_t(dx) = \rho(t,x)dx$ can now be obtained by solving

$$\inf_{(\rho,u)} \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2} \|u(t,x)\|^2 \rho(t,x) dt dx, \qquad (49a)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot ((A(t)x + B(t)u)\rho) - \frac{\epsilon}{2} \sum_{i,j=1}^n \frac{\partial^2 (a(t)_{ij}\rho)}{\partial x_i \partial x_j} = 0, \quad (49b)$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y), \qquad (49c)$$

where $a(t) = B(t)B(t)'$, see [39], [40]. Comparing (49) with (21) we see that the only difference is the extra term

$$\frac{\epsilon}{2} \sum_{i,j=1}^n \frac{\partial^2 (a(t)_{ij}\rho)}{\partial x_i \partial x_j}$$

in (49b) as compared to (21b).

Formally, (49b) converges to (21b) as ϵ goes to 0. This suggests that the minimizer of the OMT-wpd might be obtained as the limit of the joint initial-final time distribution of solutions to the Schrödinger bridge problems as the diffusivity goes to zero. This result is stated next and can be proved based on the result in [41] together with the Freidlin-Wentzell Theory [64, Section 5.6] (a large deviation principle on sample path space). As kindly noted by a referee, it can also be derived by showing tightness of solution laws of the family of martingale problems by applying, e.g., [69, Theorem 9.4, p. 145], and appealing to the fact that martingale property w.r.t. natural filtration is preserved under weak convergence, along with well-posedness of the limiting martingale problem. For completeness we include in Appendix B an alternative detailed proof.

Theorem 3: Given two probability measures $\mu_0(dx) = \rho_0(x)dx, \mu_1(dy) = \rho_1(y)dy$ on $\mathbb{R}^n$ with finite second moment, let $\mathcal{P}_{01}^\epsilon$ be the solution of the Schrödinger problem with reference Markov evolution (46) and marginals μ_0, μ_1, and let $\mathcal{P}_t^\epsilon$ be the corresponding entropic interpolation. Similarly, let π be the solution to (26) with the same marginal distributions, and μ_t the corresponding

displacement interpolation. Then, $\mathcal{P}_{01}^\epsilon$ converges weakly to π and $\mathcal{P}_t^\epsilon$ converges weakly to μ_t as ϵ goes to 0.

An important consequence of this theorem is that one can now develop numerical algorithms for the general problem of OMT with prior dynamics, and in particular for the standard OMT, by solving the Schrödinger problem for a vanishing ϵ. This approach appears particular promising in view of recent work [70] that provides an effective computational scheme to solve the Schrödinger problem by computing the pair $(\hat{\varphi}_0, \varphi_1)$ in Theorem 1 as the fixed point of an iteration. This is now being developed for diffusion processes in [58]. See also [71], [72] for similar works in discrete space setting, which has a wide range of applications. This approach to obtain approximate solutions to general OMT problems, via solutions to Schrödinger problems with vanishing noise, is illustrated in the examples of Section VI. It should also be noted that OMT problems are known to be computationally challenging in high dimensions, and specialized algorithms have been developed [2], [14]. The present approach suggests a totally new computational scheme.

V. Gaussian marginals

We now consider the correspondence between Schrödinger bridges and OMT-wpd for the special case where the marginals are Gaussian distributions. That the OMT-wpd solution corresponds to the zero-noise limit of the Schrödinger bridges is of course a consequence of Theorem 3, but in this case, we can obtain explicit expressions in closed-form and this is the point of this section.

Consider the reference evolution

$$dx(t) = A(t)x(t)dt + \sqrt{\epsilon}B(t)dw(t) \tag{50}$$

and the two marginals

$$\rho_0(x) = \frac{1}{\sqrt{(2\pi)^n|\Sigma_0|}} \exp\left[-\frac{1}{2}(x-m_0)'\Sigma_0^{-1}(x-m_0)\right], \tag{51a}$$

$$\rho_1(x) = \frac{1}{\sqrt{(2\pi)^n|\Sigma_1|}} \exp\left[-\frac{1}{2}(x-m_1)'\Sigma_1^{-1}(x-m_1)\right], \tag{51b}$$

where, as usual, the system with matrices $(A(t), B(t))$ is controllable. In our previous work [37], [38], we derived a "closed-form" expression for the Schrödinger bridge, namely,

$$\begin{aligned} dx(t) = &(A(t) - B(t)B(t)'\Pi_\epsilon(t))x(t)dt \\ &+ B(t)B(t)'m(t)dt + \sqrt{\epsilon}B(t)dw(t) \end{aligned} \tag{52}$$

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

$$\dot{\Pi}_\epsilon(t) + A(t)'\Pi_\epsilon(t) + \Pi_\epsilon(t)A(t) - \Pi_\epsilon(t)B(t)B(t)'\Pi_\epsilon(t) = 0, \tag{53}$$

$$\begin{aligned} \Pi_\epsilon(0) = \Sigma_0^{-1/2}\big[&\frac{\epsilon}{2}I + \Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2} \\ &- (\frac{\epsilon^2}{4}I + \Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Sigma_1 M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2}\big]\Sigma_0^{-1/2}, \end{aligned} \tag{54}$$

and

$$m(t) = \hat{\Phi}(1,t)'\hat{M}(1,0)^{-1}(m_1 - \hat{\Phi}(1,0)m_0), \tag{55}$$

where $\hat{\Phi}(t,s), \hat{M}(t,s)$ satisfy

$$\frac{\partial\hat{\Phi}(t,s)}{\partial t} = (A(t) - B(t)B(t)'\Pi_\epsilon(t))\hat{\Phi}(t,s), \quad \hat{\Phi}(t,t) = I$$

and

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

The probability law $\mathcal{P}^\epsilon(\cdot)$ induced on path space by the stochastic process in (52) is indeed the solution of the Schrödinger bridge problem with prior corresponding to the law induced by the stochastic process in (50) and marginals (51).

Next we consider the zero-noise limit by letting ϵ go to 0. In the case where $A(t) \equiv 0, B(t) \equiv I$, the Schrödinger bridges converge to the solution of the OMT. In general, when $A(t) \not\equiv 0, B(t) \not\equiv I$, by taking $\epsilon = 0$ in (54) we obtain

$$\begin{aligned} \Pi_0(0) = \Sigma_0^{-1/2}[&\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2} \\ &- (\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Sigma_1 M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2}]\Sigma_0^{-1/2}, \end{aligned} \tag{56}$$

and the corresponding limiting process

$$\begin{aligned} dx(t) = &(A(t) - B(t)B(t)'\Pi_0(t))x(t)dt + B(t)B(t)'m(t)dt, \\ &x(0) \sim (m_0, \Sigma_0) \end{aligned} \tag{57}$$

with $\Pi_0(t), m(t)$ satisfying (53), (55) and (56). In fact $\Pi_0(t)$ has the explicit expression

$$\begin{aligned} \Pi_0(t) = &-M(t,0)^{-1} - M(t,0)^{-1}\Phi(t,0) \\ &\times \Big[\Phi_{10}'M_{10}^{-1}\Phi_{10} - \Sigma_0^{-1/2}(\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Sigma_1 M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2}\Sigma_0^{-1/2} \\ &- \Phi(t,0)'M(t,0)^{-1}\Phi(t,0)\Big]^{-1}\Phi(t,0)'M(t,0)^{-1}. \end{aligned} \tag{58}$$

As indicated earlier, Theorem 3 already implies that (57) yields an optimal solution to (21). Here we give an alternative proof by "completion of squares", a technique used in optimal control theory.

Proposition 3: Given Gaussian marginal distributions as in (51), the linear control

$$u(t,x) = -B(t)'\Pi_0(t)x + B(t)'m(t), \tag{59}$$

with Π_0 in (53) and m in (55) solves (16).

Proof: We show first that u in (59) is a feasible control by proving that the corresponding probability density function ρ satisfies the boundary condition (51), and second, that this control u is the optimal one.

The controlled process (57) is linear with gaussian initial condition, hence $x(t)$ is a gaussian process. We claim that density of $x(t)$ is

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

where

$$n(t) = \hat{\Phi}(t,0)m_0 + \int_0^t \hat{\Phi}(t,\tau)B(\tau)B(\tau)'m(\tau)d\tau$$

and

$$\begin{aligned} \Sigma(t) = &M(t,0)\Phi(0,t)'\Sigma_0^{-1/2} \\ &\times \Big[-\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2} + (\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Sigma_1 M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2} \\ &+ \Sigma_0^{1/2}\Phi(t,0)'M(t,0)^{-1}\Phi(t,0)\Sigma_0^{1/2}\Big]^2 \Sigma_0^{-1/2}\Phi(0,t)M(t,0) \end{aligned} \tag{60}$$

for $t \in (0,1]$. It is obvious that $\mathbb{E}\{x(t)\} = n(t)$ and it is also immediate that

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

Straightforward but lengthy computations show that $\Sigma(t)$ satisfies the Lyapunov differential equation

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

Hence, $\Sigma(t)$ is the covariance of $x(t)$. Now, observing that

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

and

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

allows us to conclude that ρ satisfies $\rho(1,x) = \rho_1(x)$.

For the second part, consider the OMT-wpd (16) with augmented cost function

$$J(u) = \mathbb{E}\{\int_0^1 \frac{1}{2}\|u(t)\|^2 dt + \frac{1}{2}x(1)'\Pi_0(1)x(1)$$
$$- \frac{1}{2}x(0)'\Pi_0(0)x(0) - m(1)'x(1) + m(0)'x(0)\}.$$

This doesn't change the minimizer because the extra terms are constant under the fixed boundary distributions. Observing

$$J(u) = \mathbb{E}\{\int_0^1 \frac{1}{2}\|u(t)\|^2 dt + \frac{1}{2}d(x(t)'\Pi_0(t)x(t))$$
$$- d(m(t)'x(t))\}$$
$$= \mathbb{E}\{\int_0^1 \frac{1}{2}\|u(t) + B(t)'\Pi_0(t)x(t) - B(t)'m(t)\|^2 dt\}$$
$$+ \int_0^1 \frac{1}{2}m(t)'B(t)B(t)'m(t)dt,$$

it is easy to see that u in (59) achieves the minimum of $J(u)$. This completes the proof. ∎

VI. NUMERICAL EXAMPLES

We present two examples. The first one is on steering a collection of *inertial* particles in a 2-dimensional phase space between Gaussian marginal distributions at the two end-points of a time interval. We use the closed-form control presented in Section V. The second example is on steering distributions in a one-dimensional state-space with specified prior dynamics and more general marginal distributions. In both examples, we observe that the entropic interpolations converge to the displacement interpolation as the diffusion coefficient goes to zero.

A. Gaussian marginals

Consider a large collection of inertial particles moving in a 1-dimension configuration space (i.e., for each particle, the position $x(t) \in \mathbb{R}$). The position x and velocity v of particles are assumed
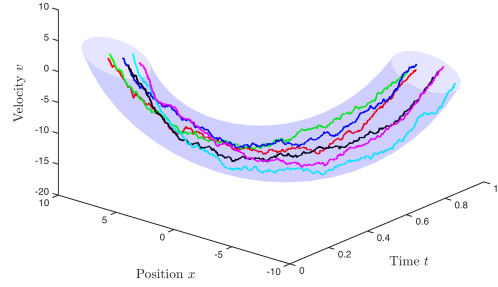

Fig. 1: Interpolation based on Schrödinger bridge with $\epsilon = 9$

to be jointly normally distributed in the 2-dimensional phase space $((x,v) \in \mathbb{R}^2)$ with mean and variance

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

at $t = 0$. We seek to steer the particles to a new joint Gaussian distribution with mean and variance

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

at $t = 1$. The problem to steer the particles provides also a natural way to interpolate these two end-point marginals by providing a flow of one-time marginals at intermediary points $t \in [0,1]$.

When the particles experience stochastic forcing, their trajectories correspond to a Schrödinger bridge with reference evolution

$$d\begin{pmatrix} x(t) \\ v(t) \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{pmatrix} x(t) \\ v(t) \end{pmatrix}dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\sqrt{\epsilon}dw(t).$$

In particular, we are interested in the behavior of trajectories when the random forcing is negligible compared to the "deterministic" drift.

Figure 1 depicts the flow of the one-time marginals of the Schrödinger bridge with $\epsilon = 9$. The transparent tube represents the 3σ region

$$(\xi(t)' - m_t')\Sigma_t^{-1}(\xi(t) - m_t) \leq 9, \quad \xi(t) = \begin{bmatrix} x(t) \\ v(t) \end{bmatrix}$$

and the curves with different color stand for typical sample paths of the Schrödinger bridge. Similarly, Figures 2 and 3 depict the corresponding flows for $\epsilon = 4$ and $\epsilon = 0.01$, respectively. The interpolating flow in the absence of stochastic disturbance, i.e., for the optimal transport with prior, is depicted in Figure 4; the sample paths are now smooth as compared to the corresponding sample paths with stochastic disturbance. As $\epsilon \searrow 0$, the paths converge to those corresponding to optimal transport and $\epsilon = 0$. For comparison, we also provide in Figure 5 the interpolation corresponding to optimal transport without prior, i.e., for the trivial dynamics $A(t) \equiv 0$ and $B(t) \equiv I$, which is precisely a constant speed translation.

B. General marginals

Consider now a large collection of particles obeying

$$dx(t) = -2x(t)dt + u(t)dt$$
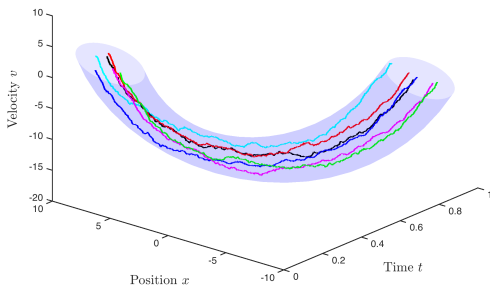
Fig. 2: Interpolation based on Schrödinger bridge with $\epsilon = 4$

Fig. 5: Interpolation based on OMT

Fig. 3: Interpolation based on Schrödinger bridge with $\epsilon = 0.01$
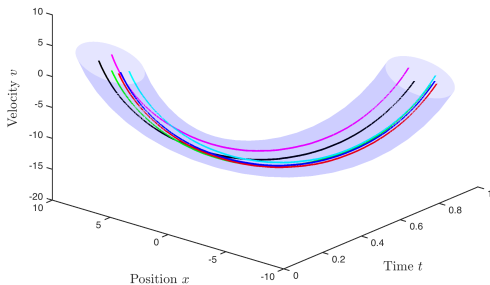
Fig. 6: Marginal distributions

in 1-dimensional state space with marginal distributions

$$\rho_0(x) = \begin{cases} 0.2 - 0.2\cos(3\pi x) + 0.2 & \text{if } 0 \leq x < 2/3 \\ 5 - 5\cos(6\pi x - 4\pi) + 0.2 & \text{if } 2/3 \leq x \leq 1, \end{cases}$$

and

$$\rho_1(x) = \rho_0(1 - x).$$

These are shown in Figure 6 and, obviously, are not Gaussian. Once again, our goal is to steer the state of the system (equivalently, the particles) from the initial distribution ρ_0 to the final ρ_1 using minimum energy control. That is, we need to solve the problem of OMT-wpd. In this 1-dimensional case, just like in the classical OMT
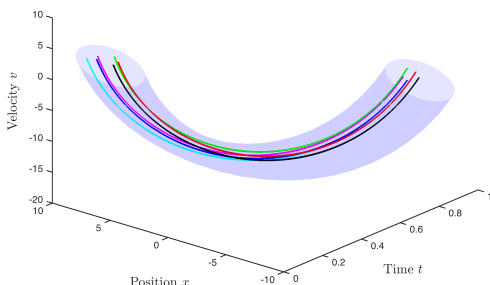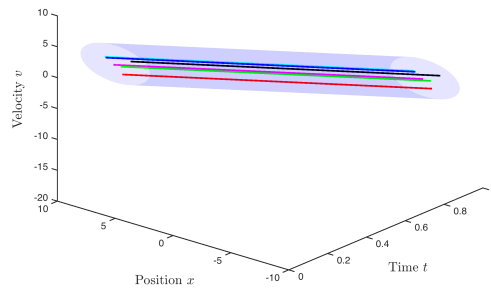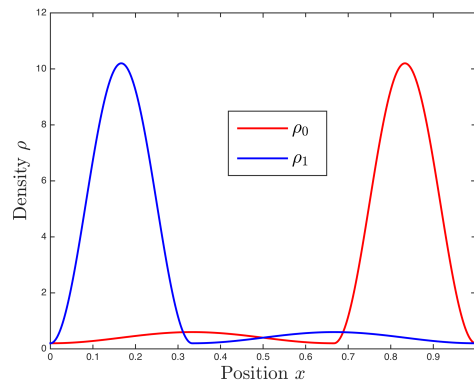
problem, the optimal transport map $y = T(x)$ between the two end-points can be determined from[7]

$$\int_{-\infty}^{x} \rho_0(y)dy = \int_{-\infty}^{T(x)} \rho_1(y)dy.$$

The interpolation flow ρ_t, $0 \leq t \leq 1$ can then be obtained using (31). Figure 7 depicts the solution of OMT-wpd. For comparison, we also show the solution of the classical OMT in figure 8 where the particles move on straight lines.

Finally, we assume a stochastic disturbance,

$$dx(t) = -2x(t)dt + u(t)dt + \sqrt{\epsilon}dw(t),$$

with $\epsilon > 0$. Figure 9–12 depict minimum energy flows for diffusion coefficients $\sqrt{\epsilon} = 0.5, \ 0.15, \ 0.05, \ 0.01$, respectively. As $\epsilon \to 0$, it is seen that the solution to the Schrödinger problem converges to the solution of the problem of OMT-wpd as expected.

VII. RECAP

The problem to steer the random state of a dynamical system between given probability distributions can be equally well be seen as the control problem to simultaneously herd a collection of particles obeying the given dynamics, or as the problem to identify a potential that effects such a transition. The former is seen to have applications

Fig. 4: Interpolation based on OMT-wpd

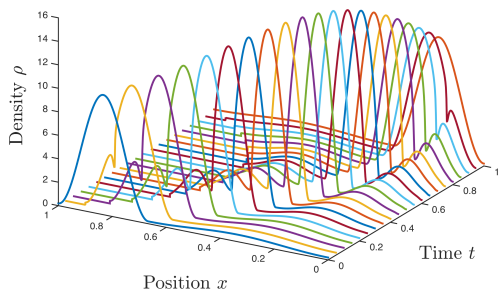[7]In this 1-dimensional case, (30) is a simple rescaling and, therefore, $T(\cdot)$ inherits the monotonicity of $\hat{T}(\cdot)$.
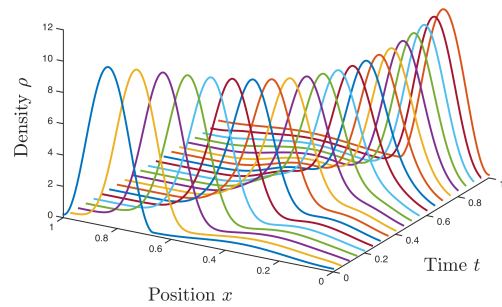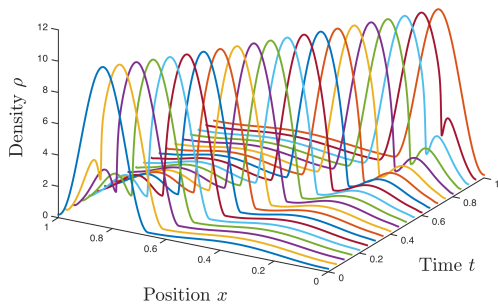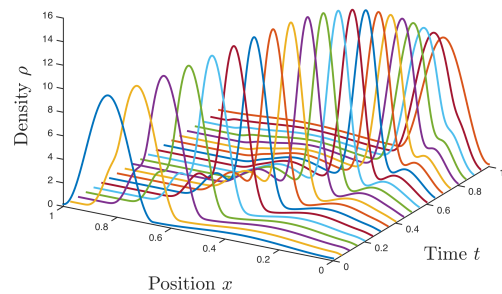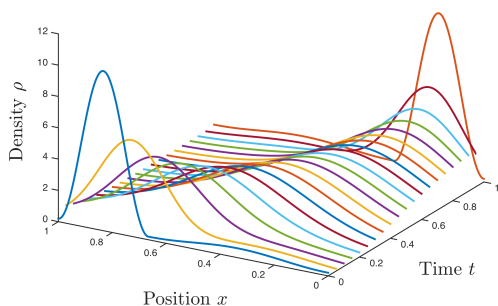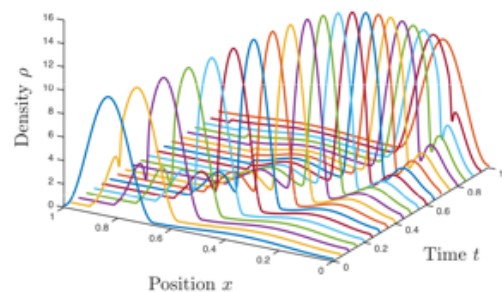
Fig. 7: Interpolation based on OMT-wpd

Fig. 10: Interpolation based on Schrödinger bridge with $\sqrt{\epsilon} = 0.15$

Fig. 8: Interpolation based on OMT

Fig. 11: Interpolation based on Schrödinger bridge with $\sqrt{\epsilon} = 0.05$

in the control of uncertain systems, system of particles, etc. The latter is seen as either a modeling or a system identification problem, where e.g., the collective response of particles is observed and the prior dynamics need to be adjusted by postulating a suitable potential so as to be consistent with observed marginals. When the dynamics are trivial (the state matrix is zero and the input matrix is the identity), the problem reduces to the classical OMT problem. Herein we presented a generalization to nontrivial linear dynamics. A version of both viewpoints where an added stochastic disturbance is present relates to the problem of constructing the so-called Schrödinger bridge between two end-point marginals. In fact, Schrödinger's bridge problem was

conceived as a modeling problem to identify a probability law on path space that is closest to a prior and is consistent with the marginals. Its stochastic control reformulation in the 90's has led to a rapidly developing subject. The present work relates OMT as a limit to Schrödinger bridges, when the stochastic disturbance goes to zero, and discusses the generalization of both to the setting where the prior linear dynamics are quite general. It opens the way to employ the efficient iterative techniques recently developed for Schrödinger bridges [70] to the computationally challenging OMT (with or without prior dynamics). This is the topic of [58].

Fig. 9: Interpolation based on Schrödinger bridge with $\sqrt{\epsilon} = 0.5$

Fig. 12: Interpolation based on Schrödinger bridge with $\sqrt{\epsilon} = 0.01$

APPENDIX

A. Proof of Proposition 2

The velocity field associated with $u(t,x) = B(t)'\nabla\psi(t,x)$ is

$$v(t,x) = A(t)x + B(t)B(t)'\nabla\psi(t,x),$$

which is well-defined almost everywhere (as it will be shown below that ψ is indeed differentiable almost everywhere). Since we already know from previous discussion that T_t in (31b) gives the trajectories associated with the optimal transportation plan, it suffices to show

$$v(t,\cdot) \circ T_t = dT_t/dt,$$

that is, $v(t,x)$ is the velocity field associated with the trajectories $(T_t)_{0 \le t \le 1}$. We next prove $v(t,\cdot) \circ T_t = dT_t/dt$.

For $0 < t < 1$, formula (38) can be rewritten as

$$g(x) = \sup_y \left\{ x'M(t,0)^{-1}\Phi(t,0)y - f(y) \right\},$$

with

$$
\begin{aligned}
g(x) &= \frac{1}{2}x'M(t,0)^{-1}x - \psi(t,x) \\
f(y) &= \frac{1}{2}y'\Phi(t,0)'M(t,0)^{-1}\Phi(t,0)y + \psi(0,y).
\end{aligned}
$$

The function

$$
\begin{aligned}
f(y) &= \frac{1}{2}y'\Phi(t,0)'M(t,0)^{-1}\Phi(t,0)y + \psi(0,y) \\
&= \frac{1}{2}y'\left[\Phi(t,0)'M(t,0)^{-1}\Phi(t,0) - \Phi_{10}'M_{10}^{-1}\Phi_{10}\right]y \\
&\quad + \phi(M_{10}^{-1/2}\Phi_{10}y)
\end{aligned}
$$

is uniformly convex since ϕ is convex and the matrix

$$
\begin{aligned}
&\Phi(t,0)'M(t,0)^{-1}\Phi(t,0) - \Phi_{10}'M_{10}^{-1}\Phi_{10} \\
&= \left(\int_0^t \Phi(0,\tau)B(\tau)B(\tau)'\Phi(0,\tau)'d\tau\right)^{-1} \\
&\quad - \left(\int_0^1 \Phi(0,\tau)B(\tau)B(\tau)'\Phi(0,\tau)'d\tau\right)^{-1}
\end{aligned}
$$

is positive definite. Hence, f, g, ψ are differentiable almost everywhere, and from a similar argument to the case of Legendre transform, we obtain

$$\nabla g \circ (M(t,0)\Phi(0,t)'\nabla f(x)) = M(t,0)^{-1}\Phi(t,0)x,$$

for all $x \in \mathbb{R}^n$. It follows

$$
\begin{aligned}
&(M(t,0)^{-1} - \nabla\psi(t,\cdot)) \circ \Big(M(t,0)\Phi(0,t)' \\
&\quad \times \left[\Phi(t,0)'M(t,0)^{-1}\Phi(t,0)x + \nabla\psi(0,x)\right] \Big) \\
&= M(t,0)^{-1}\Phi(t,0)x.
\end{aligned}
$$

After some cancellations it yields

$$
\begin{aligned}
&\nabla\psi(t,\cdot) \circ \Phi(t,0)x + \nabla\psi(t,\cdot) \circ M(t,0)\Phi(0,t)'\nabla\psi(0,x) \\
&- \Phi(0,t)'\nabla\psi(0,x) = 0. \quad (61)
\end{aligned}
$$

On the other hand, since

$$T(x) = M_{10}^{-1/2}\nabla\phi(M_{10}^{-1/2}\Phi_{10}x) = M_{10}\Phi_{01}'\nabla\psi(0,x) + \Phi_{10}x,$$

we have

$$
\begin{aligned}
T_t(x) &= \Phi(t,1)M(1,t)M_{10}^{-1}\Phi_{10}x + M(t,0)\Phi(1,t)'M_{10}^{-1}T(x) \\
&= (\Phi(t,1)M(1,t) + M(t,0)\Phi(0,t)')M_{10}^{-1}\Phi_{10}x \\
&\quad + M(t,0)\Phi(0,t)'\nabla\psi(0,x) \\
&= \Phi(t,0)x + M(t,0)\Phi(0,t)'\nabla\psi(0,x).
\end{aligned}
$$

The fact that $(\Phi(t,1)M(1,t) + M(t,0)\Phi(0,t)')M_{10}^{-1}\Phi_{10} = \Phi(t,0)$ follows by substituting expressions for the Grammians from (22). It now follows that

$$
\begin{aligned}
\frac{dT_t(x)}{dt} &= A(t)\Phi(t,0)x + A(t)M(t,0)\Phi(0,t)'\nabla\psi(0,x) \\
&\quad + B(t)B(t)'\Phi(0,t)'\nabla\psi(0,x).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&v(t,\cdot) \circ T_t(x) - \frac{dT_t(x)}{dt} \\
&= \left[A(t) + B(t)B(t)'\nabla\psi(t,\cdot)\right] \circ \left[\Phi(t,0)x \right. \\
&\quad \left. + M(t,0)\Phi(0,t)'\nabla\psi(0,x)\right] \\
&\quad - \left[A(t)\Phi(t,0)x + A(t)M(t,0)\Phi(0,t)'\nabla\psi(0,x) \right. \\
&\quad \left. + B(t)B(t)'\Phi(0,t)'\nabla\psi(0,x)\right] \\
&= B(t)B(t)' \left\{\nabla\psi(t,\cdot) \circ \Phi(t,0)x \right. \\
&\quad + \nabla\psi(t,\cdot) \circ M(t,0)\Phi(0,t)'\nabla\psi(0,x) \\
&\quad \left. - \Phi(0,t)'\nabla\psi(0,x)\right\} = 0,
\end{aligned}
$$

by (61), which completes the proof.

B. Proof of Theorem 3

The Markov kernel of (46) is

$$
\begin{aligned}
q^\epsilon(s,x,t,y) &= (2\pi\epsilon)^{-n/2}|M(t,s)|^{-1/2} \quad (62) \\
&\times \exp\left(-\frac{1}{2\epsilon}(y - \Phi(t,s)x)'M(t,s)^{-1}(y - \Phi(t,s)x)\right).
\end{aligned}
$$

Comparing this and the Brownian kernel $q^{B,\epsilon}$ we obtain

$$
\begin{aligned}
q^\epsilon(s,x,t,y) &= (t-s)^{n/2}|M(t,s)|^{-1/2} \\
&\times q^{B,\epsilon}(s, M(t,s)^{-1/2}\Phi(t,s)x, t, M(t,s)^{-1/2}y).
\end{aligned}
$$

Now define two new marginal distributions $\hat{\rho}_0$ and $\hat{\rho}_1$ through the coordinates transformation C in (27),

$$
\begin{aligned}
\hat{\rho}_0(x) &= |M_{10}|^{1/2}|\Phi_{10}|^{-1}\rho_0(\Phi_{10}^{-1}M_{10}^{1/2}x), \\
\hat{\rho}_1(x) &= |M_{10}|^{1/2}\rho_1(M_{10}^{1/2}x).
\end{aligned}
$$

Let $(\hat{\varphi}_0, \varphi_1)$ be a pair that solves the Schrödinger bridge problem with kernel q^ϵ and marginals ρ_0, ρ_1, and define $(\hat{\varphi}_0^B, \varphi_1^B)$ as

$$
\begin{aligned}
\hat{\varphi}_0(x) &= |\Phi_{10}|\hat{\varphi}_0^B(M_{10}^{-1/2}\Phi_{10}x), \quad (63a) \\
\varphi_1(x) &= |M_{10}|^{-1/2}\varphi_1^B(M_{10}^{-1/2}x), \quad (63b)
\end{aligned}
$$

then the pair $(\hat{\varphi}_0^B, \varphi_1^B)$ solves the Schrödinger bridge problem with kernel $q^{B,\epsilon}$ and marginals $\hat{\rho}_0, \hat{\rho}_1$. To verify this, we need only to show that the joint distribution

$$\mathcal{P}_{01}^{B,\epsilon}(E) = \int_E q^{B,\epsilon}(0,x,1,y)\hat{\varphi}_0^B(x)\varphi_1^B(y)dxdy$$

matches the marginals $\hat{\rho}_0, \hat{\rho}_1$. This follows from

$$\int_{\mathbb{R}^n} q^{B,\epsilon}(0,x,1,y)\hat{\varphi}_0^B(x)\varphi_1^B(y)dy$$

$$= \int_{\mathbb{R}^n} q^{B,\epsilon}(0,x,1,M_{10}^{-1/2}y)\hat{\varphi}_0^B(x)\varphi_1^B(M_{10}^{-1/2}y)d(M_{10}^{-1/2}y)$$

$$= |M_{10}|^{1/2}|\Phi_{10}|^{-1}\int_{\mathbb{R}^n} q^\epsilon(0,\Phi_{10}^{-1}M_{10}^{1/2}x,1,y)$$

$$\times \hat{\varphi}_0(\Phi_{10}^{-1}M_{10}^{1/2}x)\varphi_1(y)dy$$

$$= |M_{10}|^{1/2}|\Phi_{10}|^{-1}\rho_0(\Phi_{10}^{-1}M_{10}^{1/2}x) = \hat{\rho}_0(x),$$

and

$$\int_{\mathbb{R}^n} q^{B,\epsilon}(0,x,1,y)\hat{\varphi}_0^B(x)\varphi_1^B(y)dx$$

$$= \int_{\mathbb{R}^n} q^{B,\epsilon}(0,M_{10}^{-1/2}\Phi_{10}x,1,y)\hat{\varphi}_0^B(M_{10}^{-1/2}\Phi_{10}x)$$

$$\times \varphi_1^B(y)d(M_{10}^{-1/2}\Phi_{10}x)$$

$$= |M_{10}|^{1/2}\int_{\mathbb{R}^n} q^\epsilon(0,x,1,M_{10}^{1/2}y)\hat{\varphi}_0(x)\varphi_1(M_{10}^{1/2}y)dx$$

$$= |M_{10}|^{1/2}\rho_1(M_{10}^{1/2}y) = \hat{\rho}_1(y).$$

Compare $\mathcal{P}_{01}^{B,\epsilon}$ with $\mathcal{P}_{01}^\epsilon$ it is not difficult to find out that $\mathcal{P}_{01}^{B,\epsilon}$ is a push-forward of $\mathcal{P}_{01}^\epsilon$, that is,

$$\mathcal{P}_{01}^{B,\epsilon} = C_\sharp \mathcal{P}_{01}^\epsilon.$$

On the other hand, let π^B be the solution to classical OMT (3) with marginals $\hat{\rho}_0, \hat{\rho}_1$, then

$$\pi^B = C_\sharp \pi.$$

Now since $\mathcal{P}_{01}^{B,\epsilon}$ weakly converge to π^B from Theorem 2, we conclude that $\mathcal{P}_{01}^\epsilon$ weakly converge to π as ϵ goes to 0.

We next show $\mathcal{P}_t^\epsilon$ weakly converges to μ_t as ϵ goes to 0 for all t. The corresponding path space measure μ can be expressed as

$$\mu(\cdot) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \delta_{\gamma^{xy}}(\cdot)\,\pi(dxdy),$$

where γ^{xy} is the minimum energy path (25) connecting x, y, and $\delta_{\gamma^{xy}}$ is the Dirac measure concentrated on γ^{xy}. Similarly, the Schrödinger bridge $\mathcal{P}^\epsilon$ can be decomposed [42] as

$$\mathcal{P}^\epsilon(\cdot) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathcal{Q}_{xy}^\epsilon(\cdot)\,\mathcal{P}_{01}^\epsilon(dxdy),$$

where $\mathcal{Q}_{xy}^\epsilon$ is the pinned bridge [73] (a generalization of Brownian bridge) associated with (46) conditioned on $x^\epsilon(0) = x$ and $x^\epsilon(1) = y$, and it has the stochastic differential equation representation

$$dx^\epsilon(t) = (A(t) - B(t)B(t)'\Phi(1,t)'M(1,t)^{-1}\Phi(1,t))x^\epsilon(t)dt$$
$$+ B(t)B(t)'\Phi(1,t)'M(1,t)^{-1}ydt + \sqrt{\epsilon}B(t)dw(t) \quad (64)$$

with initial value $x^\epsilon(0) = x$. As ϵ goes to zero, $\mathcal{Q}_{xy}^\epsilon$ tends to concentrate on the solution of

$$dx^0(t) = (A(t) - B(t)B(t)'\Phi(1,t)'M(1,t)^{-1}\Phi(1,t))x^0(t)dt$$
$$+ B(t)B(t)'\Phi(1,t)'M(1,t)^{-1}ydt,\ x^0(0) = x,$$

which is γ^{xy}. The linear stochastic differential equation (64) represents a Gaussian process. It has the following explicit expression

$$x^\epsilon(t) = \gamma^{xy}(t) + \sqrt{\epsilon}\int_0^t \tilde{\Phi}(t,\tau)B(\tau)dw(\tau),\ 0 \le t < 1, \quad (65)$$

where $\tilde{\Phi}$ is the transition matrix of the dynamics (64), and $x^\epsilon(t)$ converges to y almost surely as t goes to 1, see [73]. From (65) it is easy to see that the autocovariance of $x^\epsilon(\cdot)$ depends linearly on ϵ

and therefore goes to 0 as $\epsilon \to 0$. Combining this and the fact $x^\epsilon(\cdot)$ is a Gaussian process we conclude that the set of processes $x^\epsilon(\cdot)$ is tight [74, Theorem 7.3] and their finite dimensional distributions converge weakly to those of $x^0(\cdot)$. Hence, $\mathcal{Q}_{xy}^\epsilon$ converges weakly to $\delta_{\gamma^{xy}}$ [74, Theorem 7.1] as ϵ goes to 0.

We finally claim that $\mathcal{P}_t^\epsilon$ weakly converges to μ_t as ϵ goes to 0 for each t. To see this, choose a bounded, uniformly continuous[8] function h and define

$$g^\epsilon(x,y) := \langle \mathcal{Q}_{xy,t}^\epsilon, h\rangle,$$

$$g(x,y) := \langle \delta_{\gamma^{xy},t}, h\rangle$$
$$= h(\gamma^{xy}(t)),$$

where $\langle\cdot,\cdot\rangle$ denotes the integration of the function against the measure. From (25) it is immediate that g is a bounded continuous functions of x, y. Since $\mathcal{Q}_{xy,t}^\epsilon$ is a Gaussian distribution with mean $\gamma^{xy}(t)$ and covariance which is independent of x, y and tends to zero as $\epsilon \to 0$ based on (65), $g^\epsilon \to g$ uniformly as $\epsilon \to 0$. It follows

$$\langle \mathcal{P}_t^\epsilon, h\rangle - \langle \mu_t, h\rangle = \langle \mathcal{P}_{01}^\epsilon, g^\epsilon\rangle - \langle \pi, g\rangle$$
$$= (\langle \mathcal{P}_{01}^\epsilon, g\rangle - \langle \pi, g\rangle) + \langle \mathcal{P}_{01}^\epsilon, g^\epsilon - g\rangle.$$

Both summands tend to zero as $\epsilon \to 0$, the first due to weak convergence of $\mathcal{P}_{01}^\epsilon$ to π and the second due to the uniform convergence of g^ϵ to g. This completes the proof.

REFERENCES

[1] C. Villani, *Topics in Optimal Transportation*. American Mathematical Soc., 2003, no. 58.

[2] J.-D. Benamou and Y. Brenier, "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem," *Numerische Mathematik*, vol. 84, no. 3, pp. 375–393, 2000.

[3] L. V. Kantorovich, "On the transfer of masses," in *Dokl. Akad. Nauk. SSSR*, vol. 37, no. 7-8, 1942, pp. 227–229.

[4] W. Gangbo and R. J. McCann, "The geometry of optimal transportation," *Acta Mathematica*, vol. 177, no. 2, pp. 113–161, 1996.

[5] R. Jordan, D. Kinderlehrer, and F. Otto, "The variational formulation of the Fokker–Planck equation," *SIAM journal on mathematical analysis*, vol. 29, no. 1, pp. 1–17, 1998.

[6] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory*. Springer, 1998, vol. 1.

[7] L. C. Evans and W. Gangbo, *Differential equations methods for the Monge-Kantorovich mass transfer problem*. American Mathematical Soc., 1999, vol. 653.

[8] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2006.

---

[8]To guarantee weak convergence, it suffices to have bounded, uniform continuous test functions by Portmanteau Theorem [74].

[9] C. Villani, *Optimal Transport: Old and New*. Springer, 2008, vol. 338.

[10] R. J. McCann and N. Guillen, "Five lectures on optimal transportation: geometry, regularity and applications," *Analysis and geometry of metric measure spaces: lecture notes of the séminaire de Mathématiques Supérieure (SMS) Montréal*, pp. 145–180, 2011.

[11] L. Ning, T. T. Georgiou, and A. Tannenbaum, "Matrix-valued Monge-Kantorovich optimal mass transport," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 3906–3911.

[12] Y. Chen, T. T. Georgiou, and M. Pavon, "On the relation between optimal transport and schrödinger bridges: A stochastic control viewpoint," *Journal of Optimization Theory and Applications*, pp. 1–21, 2014.

[13] N. E. Leonard and E. Fiorelli, "Virtual leaders, artificial potentials and coordinated control of groups," in *Decision and Control, 2001. Proceedings of the 40th IEEE Conference on*, vol. 3. IEEE, 2001, pp. 2968–2973.

[14] S. Angenent, S. Haker, and A. Tannenbaum, "Minimizing flows for the Monge–Kantorovich problem," *SIAM journal on mathematical analysis*, vol. 35, no. 1, pp. 61–97, 2003.

[15] R. W. Brockett, "Optimal control of the Liouville equation," *AMS IP Studies in Advanced Mathematics*, vol. 39, p. 23, 2007.

[16] P. Dai Pra, "A stochastic control approach to reciprocal diffusion processes," *Applied mathematics and Optimization*, vol. 23, no. 1, pp. 313–329, 1991.

[17] E. Schrödinger, "Über die Umkehrung der Naturgesetze," *Sitzungsberichte der Preuss Akad. Wissen. Phys. Math. Klasse, Sonderausgabe*, vol. IX, pp. 144–153, 1931.

[18] E. Schrödinger, "Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique," in *Annales de l'institut Henri Poincaré*, vol. 2, no. 4. Presses universitaires de France, 1932, pp. 269–310.

[19] A. Wakolbinger, "Schrödinger bridges from 1931 to 1991," in *Proc. of the 4th Latin American Congress in Probability and Mathematical Statistics, Mexico City*, 1990, pp. 61–79.

[20] R. Fortet, "Résolution d'un système d'équations de M. Schrödinger," *J. Math. Pures Appl.*, vol. 83, no. 9, 1940.

[21] A. Beurling, "An automorphism of product measures," *The Annals of Mathematics*, vol. 72, no. 1, pp. 189–200, 1960.

[22] B. Jamison, "Reciprocal processes," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 30, pp. 65–86, 1974.

[23] H. Föllmer, "Random fields and diffusion processes," in *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*. Springer, 1988, pp. 101–203.

[24] E. Nelson, *Dynamical theories of Brownian motion*. Princeton university press Princeton, 1967, vol. 17.

[25] E. Nelson, *Quantum fluctuations*. Princeton University Press Princeton, 1985.

[26] J. Zambrini, "Stochastic mechanics according to E. Schrödinger," *Physical Review A*, vol. 33, no. 3, p. 1532, 1986.

[27] B. C. Levy and A. J. Krener, "Stochastic mechanics of reciprocal diffusions," *Journal of Mathematical Physics*, vol. 37, no. 2, pp. 769–802, 1996.

[28] M. Pavon, "Quantum Schrödinger bridges," in *Directions in Mathematical Systems Theory and Optimization*. Springer, 2003, pp. 227–238.

[29] S. Bernstein, "Sur les liaisons entre les grandeurs aléatoires," *Verh. Internat. Math.-Kongr., Zurich*, pp. 288–309, 1932.

[30] A. J. Krener, "Reciprocal diffusions and stochastic differential equations of second order?" *Stochastics: An International Journal of Probability and Stochastic Processes*, vol. 24, no. 4, pp. 393–422, 1988.

[31] B. C. Levy, R. Frezza, and A. J. Krener, "Modeling and estimation of discrete-time gaussian reciprocal processes," *Automatic Control, IEEE Transactions on*, vol. 35, no. 9, pp. 1013–1023, 1990.

[32] A. J. Krener, R. Frezza, and B. C. Levy, "Gaussian reciprocal processes and self-adjoint stochastic differential equations of second order," *Stochastics and stochastic reports*, vol. 34, no. 1-2, pp. 29–56, 1991.

[33] B. Levy and A. Krener, "Kinematics and dynamics of reciprocal diffusions," *J. Math. Phys*, vol. 34, no. 1846, p. 1875, 1993.

[34] F. P. Carli, A. Ferrante, M. Pavon, and G. Picci, "A maximum entropy solution of the covariance extension problem for reciprocal processes," *Automatic Control, IEEE Transactions on*, vol. 56, no. 9, pp. 1999–2012, 2011.

[35] P. Dai Pra and M. Pavon, "On the Markov processes of Schrödinger, the Feynman–Kac formula and stochastic control," in *Realization and Modelling in System Theory*. Springer, 1990, pp. 497–504.

[36] M. Pavon and A. Wakolbinger, "On free energy, stochastic control, and Schrödinger processes," in *Modeling, Estimation and Control of Systems with Uncertainty*. Springer, 1991, pp. 334–348.

[37] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part I," *arXiv:1408.2222*, IEEE Trans. on Automatic Control, *to appear*, 2016.

[38] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part II," *arXiv:1410.3447*, IEEE Trans. on Automatic Control, *to appear*, 2016.

[39] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of inertial particles diffusing anisotropically with losses," in *Proc. American Control Conf. (arXiv:1410.1605v1)*, 2015, pp. 1252–1257.

[40] Y. Chen, T. T. Georgiou, and M. Pavon, "Fast cooling for a system of stochastic oscillators," *Journal of Mathematical Physics*, vol. 56, no. 11, p. 113302, 2015.

[41] C. Léonard, "From the Schrödinger problem to the Monge–Kantorovich problem," *Journal of Functional Analysis*, vol. 262, no. 4, pp. 1879–1920, 2012.

[42] C. Léonard, "A survey of the Schrödinger problem and some of its connections with optimal transport," *Dicrete Contin. Dyn. Syst. A*, vol. 34, no. 4, pp. 1533–1574, 2014.

[43] T. Mikami, "Monge's problem with a quadratic cost by the zero-noise limit of h-path processes," *Probability theory and related fields*, vol. 129, no. 2, pp. 245–260, 2004.

[44] T. Mikami and M. Thieullen, "Optimal transportation problem by stochastic optimal control," *SIAM Journal on Control and Optimization*, vol. 47, no. 3, pp. 1127–1139, 2008.

[45] R. Brockett, "Notes on the control of the liouville equation," in *Control of Partial Differential Equations*. Springer, 2012, pp. 101–129.

[46] A. Hindawi, J.-B. Pomet, and L. Rifford, "Mass transportation with LQ cost functions," *Acta applicandae mathematicae*, vol. 113, no. 2, pp. 215–229, 2011.

[47] R. J. McCann, "A convexity principle for interacting gases," *Advances in mathematics*, vol. 128, no. 1, pp. 153–179, 1997.

[48] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Communications on pure and applied mathematics*, vol. 44, no. 4, pp. 375–417, 1991.

[49] J.-D. Benamou, B. D. Froese, and A. M. Oberman, "Numerical solution of the optimal transportation problem using the Monge-Ampere equation," *Journal of Computational Physics*, vol. 260, pp. 107–126, 2014.

[50] A. Figalli, *Optimal transportation and action-minimizing measures*. Publications of the Scuola Normale Superiore, Pisa, Italy, 2008.

[51] P. Bernard and B. Buffoni, "Optimal mass transportation and Mather theory," *J. Eur. Math. Soc.*, vol. 9, pp. 85–121, 2007.

[52] W. Fleming and R. Rishel, *Deterministic and Stochastic Optimal Control*. Springer, 1975.

[53] E. B. Lee and L. Markus, *Foundations of optimal control theory*. Wiley, 1967.

[54] M. Athans and P. Falb, *Optimal Control: An Introduction to the Theory and Its Applications*. McGraw-Hill, 1966.

[55] W. H. Fleming and H. M. Soner, *Controlled Markov processes and viscosity solutions*. Springer Science & Business Media, 2006, vol. 25.

[56] M. G. Crandall, H. Ishii, and P.-L. Lions, "User's guide to viscosity solutions of second order partial differential equations," *Bulletin of the American Mathematical Society*, vol. 27, no. 1, pp. 1–67, 1992.

[57] L. C. Evans, *Partial differential equations*. American Mathematical Soc., 1998, vol. 19.

[58] Y. Chen, T. T. Georgiou, and M. Pavon, "Entropic and displacement interpolation: a computational approach using the hilbert metric," *arXiv:1506.04255v1*, 2015.

[59] A. Blaquière, "Controllability of a Fokker-Planck equation, the Schrödinger system, and a related stochastic optimal control (revised version)," *Dynamics and Control*, vol. 2, no. 3, pp. 235–253, 1992.

[60] R. Van Handel, "Stochastic calculus, filtering, and stochastic control," *Course notes., URL http://www. princeton. edu/ ̃ rvan/acm217/ACM217. pdf*, 2007.

[61] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*. Springer, 1988.

[62] I. Gentil, C. Léonard, and L. Ripani, "About the analogy between optimal transport and minimal entropy," *arXiv:1510.08230*, 2015.

[63] E. Carlen, "Stochastic mechanics: a look back and a look ahead," in *Diffusion, Quantum Theory and Radically Elementary Mathematics*, W. G. Faris, Ed. Princeton University Press, 2006, vol. 47, pp. 117–139.

[64] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*. Springer Science & Business Media, 2009, vol. 38.

[65] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes*. Elsevier, 2014.

[66] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[67] M. Fischer, "Markov processes and martingale problems," http://www.math.unipd.it/ ̃fischer/Didattica/MarkovMP.pdf, 2012, notes for a graduate course in mathematics, University of Padua.

[68] A. Wakolbinger, "A simplified variational characterization of Schrödinger processes," *Journal of mathematical physics*, vol. 30, no. 12, pp. 2943–2946, 1989.

[69] S. N. Ethier and T. G. Kurtz, *Markov processes: characterization and convergence*. John Wiley & Sons, 2009, vol. 282.

[70] T. T. Georgiou and M. Pavon, "Positive contraction mappings for classical and quantum Schrödinger systems," *J. Math. Phys. (arXiv:1405.6650v2)*, vol. 56, p. 033301, 2015.

[71] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Advances in Neural Information Processing Systems*, 2013, pp. 2292–2300.

[72] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, "Iterative bregman projections for regularized transportation problems," *SIAM Journal on Scientific Computing*, vol. 37, no. 2, pp. A1111–A1138, 2015.

[73] Y. Chen and T. T. Georgiou, "Stochastic bridges of linear systems," *arXiv:1407.3421*, IEEE Trans. on Automatic Control, *to appear*, 2016.

[74] P. Billingsley, *Convergence of probability measures*. John Wiley & Sons, 1999.

Yongxin Chen received his BSc in Mechanical Engineering from Shanghai Jiao Tong university, China, in 2011. He obtained his Ph.D. in Mechanical Engineering from University of Minnesota in 2016 under the supervision of Tryphon Georgiou, with a Ph.D. minor in Mathematics. He is now a postdoc fellow in the Department of Electrical and Computer Engineering in University of Minnesota. He is interested in the application of mathematics in engineering and theoretical physics. His current research focuses on linear dynamical systems, stochastic processes and optimal mass transport theory.

Tryphon T. Georgiou received the Diploma in Mechanical and Electrical Engineering from the National Technical University of Athens, Greece, in 1979 and the Ph.D. degree from the University of Florida, Gainesville, in 1983. He is a faculty in the Department of Electrical and Computer Engineering at the University of Minnesota and the Vincentine Hermes-Luh Chair. He is a recipient of the George S. Axelby Outstanding Paper award of the IEEE Control Systems Society for the years 1992, 1999, and 2003, a Fellow of the Institute of Electrical and Electronic Engineers (IEEE), and a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA).

Michele Pavon was born in Venice, Italy, on October 12, 1950. He received the Laurea degree from the University of Padova, Padova, Italy, in 1974, and the Ph.D. degree from the University of Kentucky, Lexington, in 1979, both in mathematics. After service in the Italian Army, he was on the research staff of LADSEB-CNR, Padua, Italy, for six years.Since July 1986, he has been a Professor at the School of Engineering, the University of Padova. He has visited several institutions in Europe, Northern America and Asia. His present research interests include maximum entropy and optimal transport problems.