Title
Optimal Transport Over a Linear Dynamical System

Journal
IEEE Transactions on Automatic Control, 62(5)

ISSN
0018-9286

Authors
Chen, Yongxin
Georgiou, Tryphon T
Pavon, Michele

Publication Date
2017

DOI
10.1109/tac.2016.2602103

Optimal transport over a linear dynamical system

Yongxin Chen, Tryphon 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.

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

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

where w is a Wiener process, u is a control input, x is the state process, and (A, B) is a controllable pair of 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[1] $[0,\, 1]$, that is, the minimum of

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

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

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[2] with quadratic cost [3], [7]. 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 [10] 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, 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 [11], and the interpolation/morphing of distributions [12].

[1] There is no loss in generality having time window $[0, 1]$ instead of, the more general $[t_0, t_1]$. This is done for notational convenience.

[2] Historically, the modern formulation of OMT is due to Leonid Kantorovich [1] and has been the focus of dramatic developments because of its relevance in many diverse fields including economics, physics, engineering, and probability [2], [3], [4], [5], [6], [7], [8], [3], [4], [9]. Kantorovich's contributions and the impact of the OMT to resource allocation was recognized with the Nobel Prize in Economics in 1975.

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 recently shown to be equivalent to the so-called *Schrödinger bridge problem*[3] [20], [23], [24]. The Schödinger 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 [25], [26], [27], [28] 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 solutions to OMT for both, the standard one as well as the problem of OMT with prior dynamics.

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. In the original formulation of OMT, due to Gaspar 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 [3]. 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

[3]The Schrödinger bridge problem, in its original formulation [13], [14], [15], seeks a probability law on path space with given two end-point marginals which is close to a Markovian *prior* distribution in the sense of large deviations (minimum relative entropy). Early important contributions were due to Fortet, Beurling, Jamison and Föllmer [16], [17], [18], [19] while renewed interest was sparked after a close relationship to stochastic control was recognized [20], [21], [22].

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. Thus, in the Kantorovich formulation we seek

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

When the optimal Monge-map T exists, the support of the coupling is precisely the graph of T, see [3].

Formulation (3) represents a "static" end-point formulation, i.e., focusing on "what goes where". Ingenious insights due to Benamou and Brenier [7] and [29] led to a fluid dynamic formulation of OMT. An elementary derivation of the above was presented in [10] 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 continuous feedback control laws. 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, and accordingly a distribution for $x^v(t) \sim \rho(t,x)dx$. Then, ρ satisfies weakly[4] the continuity equation

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

expressing the conservation of probability mass 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 reformulated as a "fluid-dynamics" problem [7]:

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

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

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

A. Solutions to OMT

Assuming that μ_0, μ_1 are absolutely continuous ($d\mu_0(dx) = \rho_0(x)dx$ and $d\mu_1(dx) = \rho_1(x)dx$) it is a standard result that OMT has a unique solution [30], [3], [4] 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 [3, p.126], [7, p.377], namely, $\det(H\phi(x))\rho_1(\nabla\phi(x)) = \rho_0(x)$, where $H\phi$ is the Hessian matrix of ϕ, which is a fully nonlinear second-order elliptic equation. The computation of ϕ has received attention only recently [7], [12] where numerical schemes have been developed. We will appeal to the availability of ϕ in the sequel without being concerned about its explicit computation.

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

Having T, 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 derivative

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

Then, $v(t,x) = T \circ T_t^{-1}(x) - T_t^{-1}(x)$ and $\rho(t,x)$ together solve (6). Here $\circ$ denotes the composition of maps.

B. Variational analysis

In this subsection we briefly recapitulate the sufficient optimality conditions for a pair $(\rho(\cdot,\cdot), v(\cdot,\cdot))$ to be a solution of (6) from [7] (see also [10, Section II] for an alternative elementary derivation).

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

$$\frac{\partial \rho^*}{\partial t} + \nabla \cdot (\nabla \psi \rho^*) = 0, \quad \rho^*(0,x) = \rho_0(x), \tag{9a}$$

where ψ is a solution of the Hamilton-Jacobi equation

$$\frac{\partial \psi}{\partial t} + \frac{1}{2}\|\nabla \psi\|^2 = 0. \tag{9b}$$

If in addition

$$\rho^*(1,x) = \rho_1(x), \tag{9c}$$

then the pair (ρ^*, v^*) with $v^*(t,x) = \nabla \psi(t,x)$ is a solution of (6).

The stochastic nature of the Benamou-Brenier formulation (6) 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 (9b) and has to be content with a viscosity solution. Let ψ be a viscosity solution of (9b) that admits the Hopf-Lax representation [3, p. 174][26, p. 4]

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

with

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

and ϕ as in (7), then this ψ together with the displacement interpolation ρ in (8) is a solution to (9).

III. Optimal mass transport with prior dynamics

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

$$c(x,y) = \inf_{x \in \mathcal{X}_{xy}} \int_0^1 L(t, x(t), \dot{x}(t)) dt, \tag{10}$$

where the Lagrangian $L(t,x,p)$ is strictly convex and superlinear in the velocity variable p, see [4, Chapter 7], [31, Chapter 1], [32]. Existence and uniqueness of an optimal transport map T has been established[5]

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

for general cost functionals as in (10). 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

$$L(t,x,p) = \frac{1}{2}\|p - v(t,x)\|^2. \tag{11}$$

This has been motivated by a transport problem "with prior" associated to the velocity field $v(t,x)$ [10, 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 (11) 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 [23] and [33] 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

$$c(x,y) = \inf_{\mathcal{U}} \int_0^1 \tilde{L}(t, x(t), u(t))dt, \text{ where} \tag{12a}$$

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \tag{12b}$$

$$x(0) = x, \quad x(1) = y, \tag{12c}$$

and $\mathcal{U}$ is a suitable class of controls[6]. This formulation extends the transportation problem in a similar manner as optimal control generalizes the classical calculus of variations [34] (albeit herein only for linear dynamics). It is easy to see that (11) corresponds to $A(t) = 0$ and $B(t)$ the identity matrix in (12). When $B(t)$ is invertible, (12) reduces to (10) 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.

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

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \tag{13}$$

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

are assumed to be controllable and the initial state $x(0)$ to be a random vector with probability density ρ_0. We seek a minimum energy continuous feedback control law $u(t,x)$ that steers the system to a final state $x(1)$ having distribution $\rho_1(x)dx$. That is, we address the following:

$$\inf_{u\in\mathcal{U}} \mathbb{E}\left\{\int_0^1 \frac{1}{2}\|u(t,x^u)\|^2 dt\right\}, \tag{14a}$$

$$\dot{x}^u(t) = A(t)x^u(t) + B(t)u(t), \tag{14b}$$

$$x^u(0) \sim \mu_0, \quad x^u(1) \sim \mu_1, \tag{14c}$$

where $\mathcal{U}$ is the family of continuous feedback control laws. Once again, this can be recast in a "fluid-dynamics" version in terms of the sought one-time probability density functions of the state vector:

$$\inf_{(\rho,u)} \int_{\mathbb{R}^n}\int_0^1 \frac{1}{2}\|u(t,x)\|^2 \rho(t,x)dtdx, \tag{15a}$$

$$\frac{\partial \rho}{\partial t} + \nabla\cdot((A(t)x + B(t)u)\rho) = 0, \tag{15b}$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y). \tag{15c}$$

Naturally, for the trivial prior dynamics $A(t) \equiv 0$ and $B(t) \equiv I$, the problem reduces to the classical OMT and the solution $\{\rho(t,\cdot) \mid 0 \le t \le 1\}$ is the *displacement interpolation* of the two marginals [29]. In the rest of the section, we show directly that Problem (15) has a unique solution.

A. Solutions to OMT-wpd

Let $\Phi(t_1,t_0)$ be the state transition matrix of (13) from t_0 to t_1, with $\Phi_{10} := \Phi(1,0)$, and

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

be the controllability Gramian of the system. Recall [35], [36] that for linear dynamics (13) and given boundary conditions $x(0) = x$, $x(1) = y$, the least energy and the corresponding control input can be given in closed-form, namely

$$\inf \int_0^1 \frac{1}{2}\|u(t)\|^2 dt = \frac{1}{2}(y - \Phi_{10}x)'M_{10}^{-1}(y - \Phi_{10}x) \tag{16}$$

which is attained for

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

and the corresponding optimal trajectory

$$x(t) = \Phi(t,1)M(1,t)M_{10}^{-1}\Phi_{10}x + M(t,0)\Phi(1,t)'M_{10}^{-1}y. \tag{17}$$

Problem (14) can now be written as

$$\inf_{\pi} \int_{\mathbb{R}^n\times\mathbb{R}^n} \frac{1}{2}(y - \Phi_{10}x)'M_{10}^{-1}(y - \Phi_{10}x)d\pi(x,y) \tag{18a}$$

$$\int_{dx}\int_{y\in\mathbb{R}^n} d\pi(x,y) = \rho_0(x)dx, \quad \int_{dy}\int_{x\in\mathbb{R}^n} d\pi(x,y) = \rho_1(y)dy, \tag{18b}$$

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

Problem (18) can be converted to the Kantorovich formulation (3) of the OMT by a transformation of coordinates. Indeed, consider the linear map

$$C : \begin{bmatrix} x \\ y \end{bmatrix} \longrightarrow \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} M_{10}^{-1/2}\Phi_{10}x \\ M_{10}^{-1/2}y \end{bmatrix} \tag{19}$$

and set

$$\hat{\pi} = C_\sharp \pi.$$

Clearly, (18a-18b) become

$$\inf_{\hat{\pi}} \int_{\mathbb{R}^n\times\mathbb{R}^n} \frac{1}{2}\|\hat{y}-\hat{x}\|^2 d\hat{\pi}(\hat{x},\hat{y}) \tag{20a}$$

$$\int_{d\hat{x}}\int_{\hat{y}\in\mathbb{R}^n} d\hat{\pi}(\hat{x},\hat{y}) = \hat{\rho}_0(\hat{x})d\hat{x}, \quad \int_{d\hat{y}}\int_{\hat{x}\in\mathbb{R}^n} d\hat{\pi}(\hat{x},\hat{y}) = \hat{\rho}_1(\hat{y})d\hat{y}, \tag{20b}$$

where

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

Problem (20) is now a standard OMT with quadratic cost function and we know that the optimal transport map $\hat{T}$ for this problem exists. It is the gradient of a convex function ϕ, i.e.,

$$\hat{T} = \nabla\phi, \tag{21}$$

and the optimal $\hat{\pi}$ is concentrated on the graph of $\hat{T}$ [30]. The solution to the original problem (20) can now be determined using $\hat{T}$, and it is

$$y = T(x) = M_{10}^{1/2}\hat{T}(M_{10}^{-1/2}\Phi_{10}x). \tag{22}$$

The one-time marginals can be readily computed as the push-forward

$$\mu_t = (T_t)_\sharp \mu_0, \tag{23a}$$

where

$$T_t(x) = \Phi(t,1)M(1,t)M_{10}^{-1}\Phi_{10}x + M(t,0)\Phi(1,t)'M_{10}^{-1}T(x), \tag{23b}$$

and

$$\rho(t,x) = d\mu_t(x)/dx. \tag{23c}$$

In this case, we refer to the parametric family of one-time marginals as *displacement interpolation with prior dynamics*.

B. Variational analysis

In this section we present a variational analysis for the OMT-wpd (15) analogous to that for the OMT problem [7], [10, Section II]. The analysis provides conditions for a pair $(\rho(\cdot,\cdot), v(\cdot,\cdot))$ to be a solution to OMT-wpd and will be used in Section V to prove optimality of the solution of the OMT-wpd with Gaussian marginals.

Let $\mathcal{P}_{\rho_0\rho_1}$ be the family of flows of probability densities satisfying the boundary conditions and $\mathcal{U}$ be the family of continuous feedback control laws $u(\cdot,\cdot)$. Consider the unconstrained minimization over $\mathcal{P}_{\rho_0\rho_1} \times \mathcal{U}$ of the Lagrangian

$$\mathcal{L}(\rho, u, \lambda) = \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2}\|u(t,x)\|^2 \rho(t,x) + \lambda(t,x) \left(\frac{\partial \rho}{\partial t} + \nabla \cdot ((A(t)x + B(t)u)\rho) \right) \right] dtdx, \tag{24}$$

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

$$\inf_{(\rho,u)\in\mathcal{P}_{\rho_0\rho_1}\times\mathcal{U}} \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2}\|u(t,x)\|^2 + \left(-\frac{\partial \lambda}{\partial t} - \nabla\lambda \cdot (A(t)x + B(t)u) \right) \right] \rho(t,x) dtdx. \tag{25}$$

Pointwise minimization with respect to u for each fixed flow of probability densities ρ gives

$$u^*(t,x) = B(t)'\nabla\lambda(t,x). \tag{26}$$

Substituting into (25), we get

$$J(\rho,\lambda) = -\int_{\mathbb{R}^n} \int_0^1 \left[\frac{\partial \lambda}{\partial t} + A(t)x \cdot \nabla\lambda + \frac{1}{2}\nabla\lambda \cdot B(t)B(t)'\nabla\lambda \right] \rho(t,x) dtdx. \tag{27}$$

As in Section II-B, we get the following sufficient conditions for optimality:

Proposition 2: Consider ρ^* that satisfies

$$\frac{\partial \rho^*}{\partial t} + \nabla \cdot [(A(t)x + B(t)B(t)'\nabla\psi)\rho^*] = 0, \quad \rho^*(0,x) = \rho_0(x), \tag{28a}$$

where ψ is a solution of the Hamilton-Jacobi equation

$$\frac{\partial \psi}{\partial t} + x'A(t)'\nabla\psi + \frac{1}{2}\nabla\psi' B(t)B(t)'\nabla\psi = 0. \tag{28b}$$

If in addition

$$\rho^*(1,x) = \rho_1(x), \tag{28c}$$

then the pair $(\rho^*(t,x), u^*(t,x) = B(t)'\nabla\psi(t,x))$ is a solution to problem (15).

It turns out that (28) always admits a solution. In fact, one solution can be constructed as follows.

Proposition 3: Given dynamics (13) and 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

$$\psi(t,x) = \inf_y \left\{ \psi(0,y) + \frac{1}{2}(x - \Phi(t,0)y)'M(t,0)^{-1}(x - \Phi(t,0)y) \right\} \tag{29}$$

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 (21). Moreover, let $\rho(t,\cdot) = (T_t)_\sharp\rho_0$ be the displacement interpolation as in (23). Then this pair (ψ,ρ) is a solution to (28).

Proof: First we show that (29) satisfies (28b). Let $H(t, x, \nabla\psi)$ be the Hamiltonian of the Hamilton-Jacobi equation (28b), that is,

$$H(t, x, \nabla\psi) = x' A(t)' \nabla\psi + \frac{1}{2}\nabla\psi' B(t)B(t)'\nabla\psi,$$

and define

$$\begin{aligned} L(t, x, v) &= \sup_p \{p \cdot v - H(t, x, p)\} \\ &= \begin{cases} \frac{1}{2}(v - A(t)x)'(B(t)B(t)')^{\dagger}(v - A(t)x) & \text{if } v - A(t)x \in \mathcal{R}(B) \\ \infty & \text{otherwise,} \end{cases} \end{aligned}$$

where $\dagger$ denotes pseudo-inverse and $\mathcal{R}(\cdot)$ denotes "the range of". Then the *Bellman principle of optimality* [34] yields a particular solution of (28b)

$$\begin{aligned} \psi(t, x) &= \inf_y \left\{ \psi_0(y) + \int_0^t L(\tau, \xi(\tau), \dot{\xi}(\tau)), \quad \xi(t) = x, \xi(0) = y \right\} \\ &= \inf_y \left\{ \psi_0(y) + \frac{1}{2}(x - \Phi(t,0)y)' M(t,0)^{-1}(x - \Phi(t,0)y) \right\}. \end{aligned}$$

This shows that (29) is indeed a solution of (28b).

Next we show (ψ, ρ) is a (weak) solution of (28a). Define

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

then (28a) becomes a linear transport equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \tag{30}$$

with velocity field $v(t, x)$. We claim

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

that is, $v(t, x)$ is the velocity field associated with the trajectories (T_t). If this claim is true, then the linear transport equation (30) follows from a standard argument [3, p. 167]. Moreover, the terminal condition (28c) follows since $\rho(1, \cdot) = T_\sharp \rho_0$. We next prove the claim. Formula (29) 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

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

is 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 semi-definite. Hence, from a similar argument to the case of Legendre transform, we obtain

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

It follows

$$(M(t,0)^{-1} - \nabla\psi(t,\cdot)) \circ \left\{M(t,0)\Phi(0,t)'\left[\Phi(t,0)'M(t,0)^{-1}\Phi(t,0)x + \nabla\psi(0,x)\right]\right\} = M(t,0)^{-1}\Phi(t,0)x.$$

After some cancellations it yields

$$\nabla\psi(t,\cdot) \circ \Phi(t,0)x + \nabla\psi(t,\cdot) \circ M(t,0)\Phi(0,t)'\nabla\psi(0,x) - \Phi(0,t)'\nabla\psi(0,x) = 0.$$

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

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

from which it follows

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

Therefore,

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

which completes the proof. ■

IV. Schrödinger bridges and their zero-noise limit

In 1931/32, Schrödinger [13], [14] 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? The process that is consistent with the observed marginals and fulfils Schrödinger's requirement is referred to as the Schrödinger bridge.

This problem has a long history [15]. In particular, Föllmer [19] showed that the solution to Schrödinger's problem corresponds to a probability law $\mathcal{P}^B$ on path space that minimizes the relative entropy with respect to the Wiener measure among all laws with given initial and terminal distributions, $\rho_0(x)dx$ and $\rho_1(x)dx$, respectively, and proved that the minimizer always exists. Beurling [17] and Jamison [18] generalized the idea of the Schrödinger bridge by changing the Wiener measure to a more general reference measure induced by a Markov process. Jamison's result is stated below.

Theorem 1: Given two probability measures $\mu_0(dx) = \rho_0(x)dx$ and $\mu_1(dx) = \rho_1(x)dx$ 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

$$\mathcal{P}_{01}(E) = \int_E q(0, x, 1, y)\hat{\varphi}_0(x)\varphi_1(y)dxdy \tag{31}$$

has marginals μ_0 and μ_1. Furthermore, the Schrödinger bridge from μ_0 to μ_1 is determined via the distribution flow

$$\mathcal{P}_t(dx) = \varphi(t, x)\hat{\varphi}(t, x)dx \tag{32a}$$

with

$$\varphi(t, x) = \int q(t, x, 1, y)\varphi_1(y)dy \tag{32b}$$

$$\hat{\varphi}(t, x) = \int q(0, y, t, x)\hat{\varphi}_0(y)dy. \tag{32c}$$

The flow (32) 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 [37].

For the case of non-degenerate Markov processes, a connection between the Schrödinger problem and stochastic optimal control was drawn by Dai Pra [20]. In particular, for the case of a Brownian kernel, he showed that the one-time marginals $\rho(t, x)$ for Schrödinger's problem can be obtained as solutions to

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_0^1 \frac{1}{2}\|v(t, x)\|^2 \rho(t, x)dtdx, \tag{33a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) - \frac{1}{2}\Delta\rho = 0, \tag{33b}$$

$$\rho(0, x) = \rho_0(x), \quad \rho(1, y) = \rho_1(y). \tag{33c}$$

Here, (33a) is the infimum of the expected cost while (33b) is the corresponding Fokker-Planck equation. The entropic interpolation is $\mathcal{P}_t(dx) = \rho(t, x)dx$.

An alternative equivalent reformulation given in [10] is

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n} \int_0^1 \left[\frac{1}{2}\|v(x, t)\|^2 + \frac{1}{8}\|\nabla \log \rho(x, t)\|^2\right] \rho(t, x)dtdx, \tag{34a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \tag{34b}$$

$$\rho(0, x) = \rho_0(x), \quad \rho(1, y) = \rho_1(y), \tag{34c}$$

where the Laplacian in the dynamical constraint is traded for a "Fisher information" regularization term in the cost functional. Although the form in (34) is quite appealing, for the purposes of this paper we will use only (33).

Formulation (33) is quite similar to OMT (6) except for the presence of the Laplacian in (33b). It has been shown [27], [28], [25], [26] 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 as explained below.

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

$$q^{B,\epsilon}(s,x,t,y) = (2\pi)^{-n/2}((t-s)\epsilon)^{-n/2}\exp\left(-\frac{\|x-y\|^2}{2(t-s)\epsilon}\right) \tag{35}$$

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[7] 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 (35). The corresponding Schrödinger bridge problem with the law of $x(t)$ as prior, is equivalent to

$$\inf_{(\rho,v)} \int_{\mathbb{R}^n}\int_0^1 \frac{1}{2\epsilon}\|v(t,x)\|^2\rho(t,x)dtdx, \tag{36a}$$

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(v\rho) - \frac{\epsilon}{2}\Delta\rho = 0, \tag{36b}$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y). \tag{36c}$$

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

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

instead, i.e., "rescaling" (36a) by removing the factor $1/\epsilon$. Now observe that the only difference between (36) 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 (36b). Thus, formally, the constraint (36b) 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 with (linear) dynamics. This problem has been studied in [25] in a more abstract setting based on Large Deviation Theory [38]. 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

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

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

$$\inf_{(\rho,u)} \int_{\mathbb{R}^n}\int_0^1 \frac{1}{2}\|u(t,x)\|^2\rho(t,x)dtdx, \tag{37a}$$

$$\frac{\partial\rho}{\partial t} + \nabla\cdot((A(t)x+B(t)u)\rho) - \frac{\epsilon}{2}\sum_{i,j=1}^n \frac{\partial^2(a(t)_{ij}\rho)}{\partial x_i\partial x_j} = 0, \tag{37b}$$

$$\rho(0,x) = \rho_0(x), \quad \rho(1,y) = \rho_1(y), \tag{37c}$$

[7] 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.

where $a(t) = B(t)B(t)'$, see [39], [24]. This result represents a slight generalization of Dai Pra's result [20] in that the stochastic differential equation corresponding to (37b) may be degenerate (i.e., $\mathrm{rank}(a(t)) \neq n$). Comparing (37) with (15) 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 (37b) as compared to (15b). Formally, (37b) converges to (15b) 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 [25] together with the Freidlin-Wentzell Theory [38, Section 5.6] (a large deviation principle on sample path space). In the Appendix, we also provide a direct proof which doesn't require a large deviation principle.

Theorem 3: Given two probability measures $\mu_0(dx) = \rho_0(x)dx, \mu_1(dx) = \rho_1(x)dx$ on $\mathbb{R}^n$ with finite second moment, let $\mathcal{P}_{01}^{\epsilon}$ be the solution of the Schrödinger problem with reference Markov process

$$dx(t) = A(t)x(t)dt + \sqrt{\epsilon}B(t)dw(t) \tag{38}$$

and marginals μ_0, μ_1, and let $\mathcal{P}_t^{\epsilon}$ be the corresponding entropic interpolation. Similarly, let π be the solution to (18) 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 we can now use the numerical algorithm in [37] which provides a solution to the Schrödinger problem, for a vanishing ϵ, as a means to solve the general problem of OMT with prior dynamics (and, in particular, the standard OMT [37]). This is highlighted in the examples of Section VI. It should be noted that the algorithm, which relies on computing the pair $(\hat{\varphi}_0, \varphi_1)$ in Theorem 1, is totally different from other numerical algorithms that solve standard OMT problems [7], [12].

V. Gaussian marginals

We now consider the correspondence between Schrödinger bridges and OMT-wpd for the special case where the marginals are normal 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

$$dx(t) = A(t)x(t)dt + \sqrt{\epsilon}B(t)dw(t) \tag{39}$$

and the two marginals

$$\rho_0(x) = \frac{1}{\sqrt{(2\pi)^n|\Sigma_0|}}\exp\left[-\frac{1}{2}(x-m_0)'\Sigma_0^{-1}(x-m_0)\right], \tag{40a}$$

$$\rho_1(x) = \frac{1}{\sqrt{(2\pi)^n|\Sigma_1|}}\exp\left[-\frac{1}{2}(x-m_1)'\Sigma_1^{-1}(x-m_1)\right], \tag{40b}$$

where, as usual, the system with matrices $(A(t), B(t))$ is controllable. In our previous work [23], [33], we derived a "closed-form" expression for the Schrödinger bridge when $m_0 = m_1 = 0$, namely,

$$dx(t) = (A(t) - B(t)B(t)'\Pi_\epsilon(t))x(t)dt + \sqrt{\epsilon}B(t)dw(t) \tag{41}$$

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

$$\dot{\Pi}_\epsilon(t) + A(t)'\Pi_\epsilon(t) + \Pi_\epsilon(t)A(t) - \Pi_\epsilon(t)B(t)B(t)'\Pi_\epsilon(t) = 0 \tag{42}$$

and the boundary condition

$$\Pi_\epsilon(0) = \Sigma_0^{-1/2}[\frac{\epsilon}{2}I + \Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2} - (\frac{\epsilon^2}{4}I + \Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Sigma_1 M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2}]\Sigma_0^{-1/2}. \tag{43}$$

When $m_0 \neq 0$ or $m_1 \neq 0$ the bridge becomes:

$$dx(t) = (A(t) - B(t)B(t)'\Pi_\epsilon(t))x(t)dt + B(t)B(t)'m(t)dt + \sqrt{\epsilon}B(t)dw(t) \tag{44}$$

where

$$m(t) = \hat{\Phi}(1,t)'\hat{M}(1,0)^{-1}(m_1 - \hat{\Phi}(1,0)m_0) \tag{45}$$

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

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

and

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

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 (43) we obtain

$$\Pi_0(0) = \Sigma_0^{-1/2}[\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2} - (\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Sigma_1 M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2}]\Sigma_0^{-1/2}, \tag{46}$$

and the corresponding limiting process

$$dx(t) = (A(t) - B(t)B(t)'\Pi_0(t))x(t)dt + B(t)B(t)'m(t)dt, \quad x(0) \sim (m_0, \Sigma_0) \tag{47}$$

with $\Pi_0(t), m(t)$ satisfying (42), (45) and (46). In fact $\Pi_0(t)$ has the explicit expression

$$\begin{aligned}\Pi_0(t) \;=\; & -M(t,0)^{-1} - M(t,0)^{-1}\Phi(t,0)\Big[\Phi_{10}'M_{10}^{-1}\Phi_{10} - \Sigma_0^{-1/2}(\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Sigma_1 M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2}\Sigma_0^{-1/2} \\ & -\Phi(t,0)'M(t,0)^{-1}\Phi(t,0)\Big]^{-1}\Phi(t,0)'M(t,0)^{-1}.\end{aligned} \tag{48}$$

As indicated earlier, Theorem 3 already implies that (47) yields an optimal solution to (15). Here we give an alternative proof by directly verifying that the corresponding displacement interpolation and the control satisfy the conditions of Proposition 2.

Proposition 4: Let $\rho(t,\cdot)$ be the probability density of the process $x(t)$ in (47), and

$$u(t,x) = -B(t)'\Pi_0(t)x + B(t)'m(t),$$

then the pair (ρ, u) is a solution of the problem (15) with prior dynamics (13) and marginals (40).

Proof: To prove that the pair (ρ, u) is a solution, we show first that ρ satisfies the boundary condition $\rho(1,x) = \rho_1(x)$, and second, that $u(t,x) = B(t)'\nabla\psi(t,x)$ for some ψ that satisfies the Hamilton-Jacobi equation (28b).

Equation (47) 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

$$\begin{aligned}\Sigma(t) &= M(t,0)\Phi(0,t)'\Sigma_0^{-1/2}\Big[-\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2} + (\Sigma_0^{1/2}\Phi_{10}'M_{10}^{-1}\Sigma_1 M_{10}^{-1}\Phi_{10}\Sigma_0^{1/2})^{1/2} \\ &\quad +\Sigma_0^{1/2}\Phi(t,0)'M(t,0)^{-1}\Phi(t,0)\Sigma_0^{1/2}\Big]^2\Sigma_0^{-1/2}\Phi(0,t)M(t,0) \qquad (49)\end{aligned}$$

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

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

and

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

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

For the second claim, let

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

with

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

Clearly, $u(t,x) = B(t)'\nabla\psi$ while

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

■

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 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 normal 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.

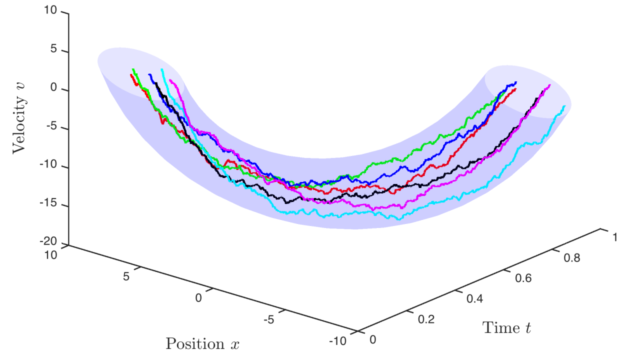

Fig. 1: Interpolation based on Schrödinger bridge with $\epsilon = 9$

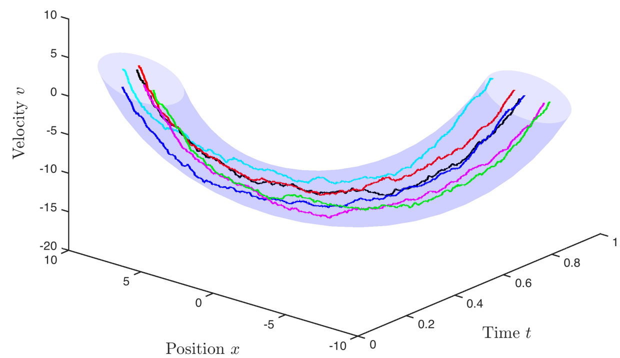

Fig. 2: Interpolation based on Schrödinger bridge with $\epsilon = 4$

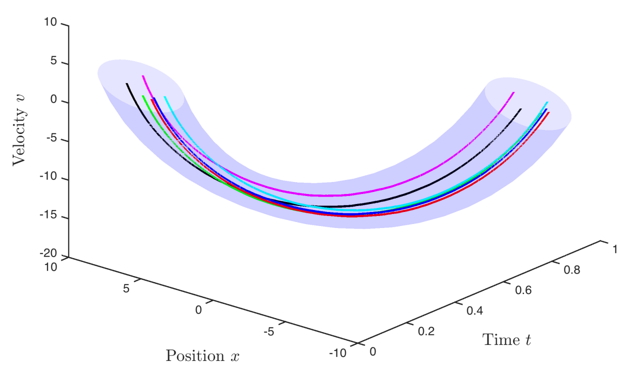

Fig. 3: Interpolation based on Schrödinger bridge with $\epsilon = 0.01$

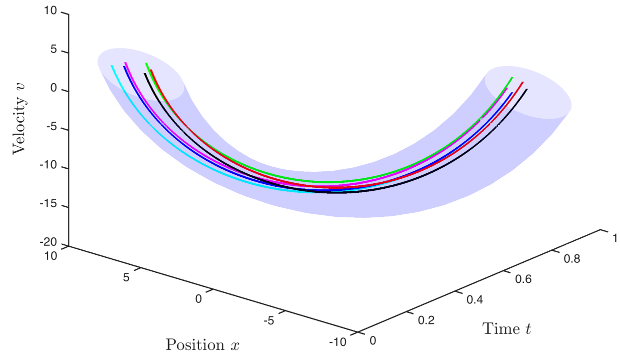

Fig. 4: Interpolation based on OMT-wpd

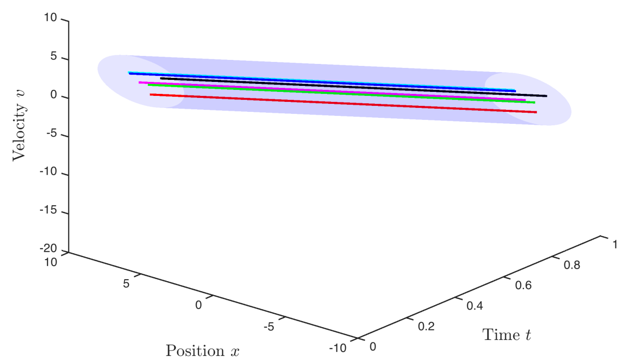

Fig. 5: Interpolation based on OMT

B. General marginals

Consider now a large collection of particles obeying

$$dx(t) = -2x(t)dt + u(t)dt$$

in 1-dimensional state space with marginal distributions

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

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

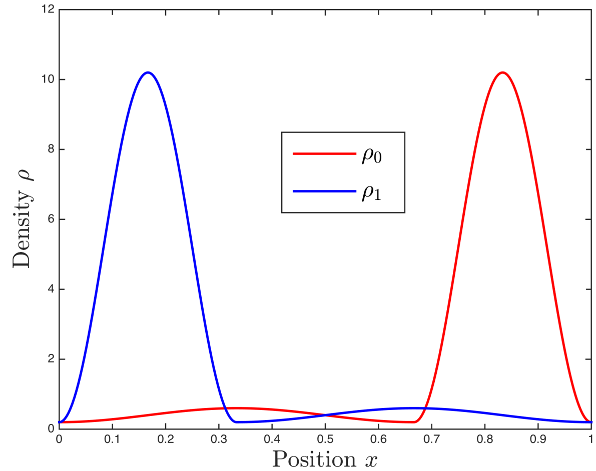

Fig. 6: Marginal distributions

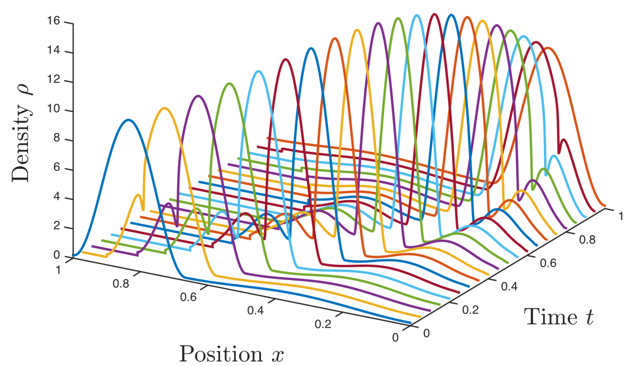

Fig. 7: Interpolation based on OMT-wpd

in the classical OMT problem, the optimal transport map $y = T(x)$ between the two end-points can be determined from[8]

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

The interpolation flow $\rho_t,\ 0 \leq t \leq 1$ can then be obtained using (23). 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 9–13 depict minimum energy flows for diffusion coefficients $\sqrt{\epsilon} = 0.5,\ 0.3,\ 0.15,\ 0.05,\ 0.01$, 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.

[8] In this 1-dimensional case, (22) is a simple rescaling and, therefore, $T(\cdot)$ inherits the monotonicity of $\hat{T}(\cdot)$.

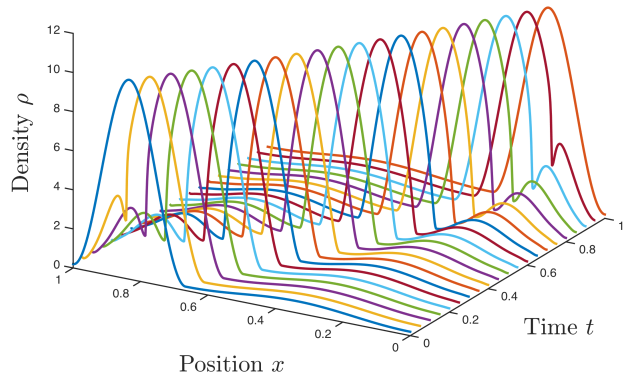

Fig. 8: Interpolation based on OMT

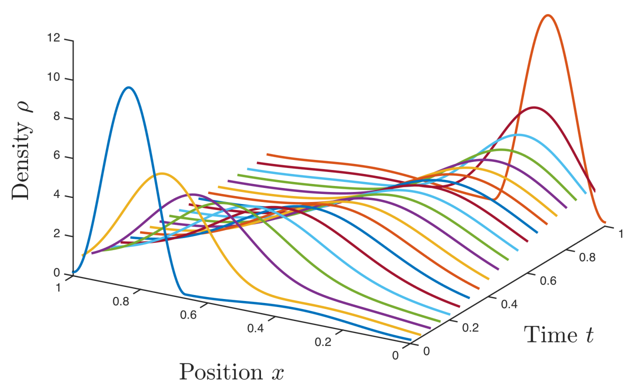

Fig. 9: Interpolation based on Schrödinger bridge with $\sqrt{\epsilon} = 0.5$

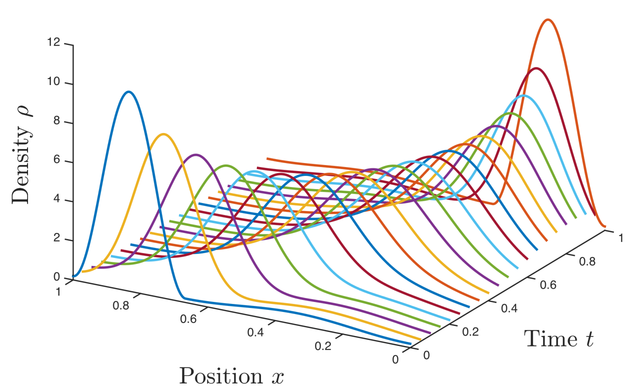

Fig. 10: Interpolation based on Schrödinger bridge with $\sqrt{\epsilon} = 0.3$

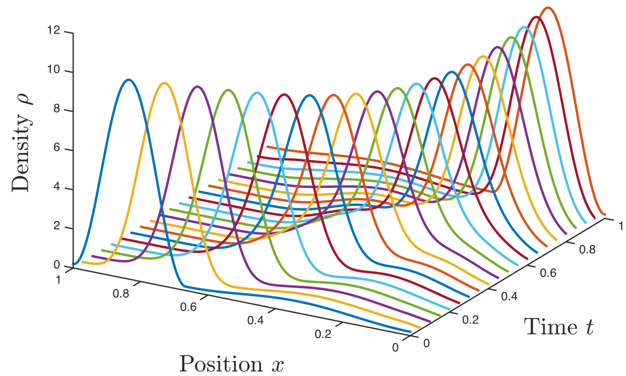

Fig. 11: Interpolation based on Schrödinger bridge with $\sqrt{\epsilon} = 0.15$

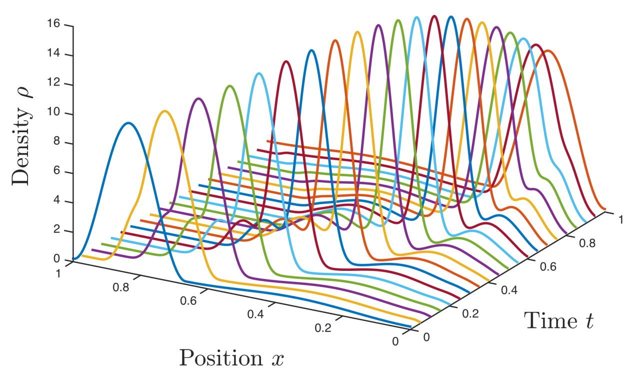

Fig. 12: Interpolation based on Schrödinger bridge with $\sqrt{\epsilon} = 0.05$

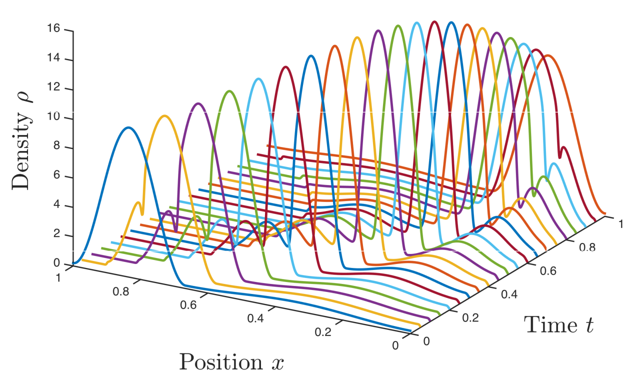

Fig. 13: Interpolation based on Schrödinger bridge with $\sqrt{\epsilon} = 0.01$

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 in the control of uncertain systems, system of particles, etc. The latter is seen as a modeling problem and 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 and the state matrix is zero while 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 (usually a Wiener measure) 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 to the computationally challenging OMT (with or without prior dynamics). This is the topic of [37].

Appendix: Proof of Theorem 3

Let q^ϵ be the Markov kernel of (38), then

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

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

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

Now define two new marginal distributions $\hat\rho_0$ and $\hat\rho_1$ through the coordinates transformation C in (19),

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

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

$$\hat\varphi_0(x) = |\Phi_{10}|\hat\varphi_0^B(M_{10}^{-1/2}\Phi_{10}x) \tag{50a}$$

$$\varphi_1(x) = |M_{10}|^{-1/2}\varphi_1^B(M_{10}^{-1/2}x), \tag{50b}$$

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

$$\begin{aligned}\int_{\mathbb{R}^n} q^{B,\epsilon}(0,x,1,y)\hat\varphi_0^B(x)\varphi_1^B(y)dy &= \int_{\mathbb{R}^n} q^{B,\epsilon}(0,x,1,M_{10}^{-1/2}y)\hat\varphi_0^B(x)\varphi_1^B(M_{10}^{-1/2}y)d(M_{10}^{-1/2}y)\\ &= |M_{10}|^{1/2}|\Phi_{10}|^{-1}\int_{\mathbb{R}^n} q^\epsilon(0,\Phi_{10}^{-1}M_{10}^{1/2}x,1,y)\hat\varphi_0(\Phi_{10}^{-1}M_{10}^{1/2}x)\varphi_1(y)dy\\ &= |M_{10}|^{1/2}|\Phi_{10}|^{-1}\rho_0(\Phi_{10}^{-1}M_{10}^{1/2}x) = \hat\rho_0(x),\end{aligned}$$

and

$$\begin{aligned}\int_{\mathbb{R}^n} q^{B,\epsilon}(0,x,1,y)\hat\varphi_0^B(x)\varphi_1^B(y)dx &= \int_{\mathbb{R}^n} q^{B,\epsilon}(0,M_{10}^{-1/2}\Phi_{10}x,1,y)\hat\varphi_0^B(M_{10}^{-1/2}\Phi_{10}x)\varphi_1^B(y)d(M_{10}^{-1/2}\Phi_{10}x)\\ &= |M_{10}|^{1/2}\int_{\mathbb{R}^n} q^\epsilon(0,x,1,M_{10}^{1/2}y)\hat\varphi_0(x)\varphi_1(M_{10}^{1/2}y)dx\\ &= |M_{10}|^{1/2}\rho_1(M_{10}^{1/2}y) = \hat\rho_1(y).\end{aligned}$$

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. The displacement interpolation μ can be decomposed as

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

where γ^{xy} is the minimum energy path (17) connecting x, y, and $\delta_{\gamma^{xy}}$ is the Dirac measure at γ^{xy} on the path space. Similarly, the entropic interpolation $\mathcal{P}^\epsilon$ can be decomposed as

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

where $\mathcal{Q}_{xy}^\epsilon$ is the pinned bridge [40] associated with (38) conditioned on $x(0)=x$ and $x(1)=y$. It has the stochastic differential equation representation

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

As ϵ goes to zero, it converges to

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

which is γ^{xy}. In other word, $\mathcal{Q}_{xy}^\epsilon$ weakly converges to $\delta_{\gamma^{xy}}$. This together with the fact that $\mathcal{P}_{01}^\epsilon$ weakly converges to π show that $\mathcal{P}_t^\epsilon$ weakly converges to μ_t as ϵ goes to 0.

REFERENCES

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

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

[3] C. Villani, *Topics in optimal transportation*. American Mathematical Soc., 2003, no. 58.

[4] ——, *Optimal transport: old and new*. Springer, 2008, vol. 338.

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

[6] 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.

[7] 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.

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

[9] 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.

[10] Y. Chen, T. Georgiou, and M. Pavon, "On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint," *arXiv preprint arXiv:1412.4430*, 2014.

[11] 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.

[12] 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.

[13] E. Schrödinger, *Über die umkehrung der naturgesetze*. Verlag Akademie der wissenschaften in kommission bei Walter de Gruyter u. Company, 1931.

[14] ——, "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.

[15] 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.

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

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

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

[19] 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.

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

[21] 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.

[22] 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.

[23] Y. Chen, T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution," *arXiv preprint arXiv:1408.2222*, 2014.

[24] ——, "Fast cooling for a system of stochastic oscillators," *arXiv preprint arXiv:1411.1323*, 2014.

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

[26] ——, "A survey of the Schrödinger problem and some of its connections with optimal transport," *arXiv preprint arXiv:1308.0215*, 2013.

[27] 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.

[28] 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.

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

[30] 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.

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

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

[33] Y. Chen, T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part II," *arXiv preprint arXiv:1410.3447*, 2014.

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

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

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

[37] Y. Chen, T. Georgiou, and M. Pavon, "A computational approach to optimal mass transport via the Schrödinger bridge problem," *in preparation*, 2015.

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

[39] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of inertial particles diffusing anisotropically with losses," *arXiv preprint arXiv:1410.1605*, 2014.

[40] Y. Chen and T. Georgiou, "Stochastic bridges of linear systems," *arXiv preprint arXiv:1407.3421*, 2014.