Title

Entropic and Displacement Interpolation: A Computational Approach Using the Hilbert Metric

Permalink

https://escholarship.org/uc/item/18j942nc

Journal

SIAM Journal on Applied Mathematics, 76(6)

ISSN

0036-1399

Authors

Chen, Yongxin
Georgiou, Tryphon
Pavon, Michele

Publication Date

2016

DOI

10.1137/16m1061382

Copyright Information

Peer reviewed

Entropic and displacement interpolation:
a computational approach using the Hilbert metric

Yongxin Chen, Tryphon T. Georgiou and Michele Pavon

Abstract

Monge-Kantorovich optimal mass transport (OMT) provides a blueprint for geometries in the space of positive densities – it quantifies the cost of transporting a mass distribution into another. In particular, it provides natural options for interpolation of distributions (*displacement interpolation*) and for modeling flows. As such it has been the cornerstone of recent developments in physics, probability theory, image processing, time-series analysis, and several other fields. In spite of extensive work and theoretical developments, the computation of OMT for large scale problems has remained a challenging task. An alternative framework for interpolating distributions, rooted in statistical mechanics and large deviations, is that of Schrödinger bridges (*entropic interpolation*). This may be seen as a stochastic regularization of OMT and can be cast as the stochastic control problem of steering the probability density of the state-vector of a dynamical system between two marginals. In this approach, however, the actual computation of flows had hardly received any attention. In recent work on Schrödinger bridges for Markov chains and quantum evolutions, we noted that the solution can be efficiently obtained from the fixed-point of a map which is contractive in the Hilbert metric. Thus, the purpose of this paper is to show that a similar approach can be taken in the context of diffusion processes which i) leads to a new proof of a classical result on Schrödinger bridges and ii) provides an efficient computational scheme for both, Schrödinger bridges and OMT. We illustrate this new computational approach by obtaining interpolation of densities in representative examples such as interpolation of images.

Index Terms

Optimal mass transport, Schrödinger bridges, Hilbert metric, interpolation of densities, image morphing
AMS Classification: 47H07, 47H09, 60J25, 34A34, 49J20

I. Introduction

The present paper is concerned with the problem of interpolating densities. Linear interpolation, which is widely used in signal analysis and image processing, has several drawbacks including fade-in fade-out effects, see e.g., [2], [24]. For this reason other methods have been pursued. The purpose of this work is to relate three relevant subjects, optimal mass transport, Schrödinger bridges, and the Hilbert metric, that provide a geometric framework for interpolation of densities. All these three subjects have a long and notable history. However, the connection between all three represents a recent development.

Among our three subjects, the one with the longest history is that of *Monge-Kantorovich Optimal Mass Transport* (OMT) originating in the work of Gaspard Monge in 1781 [33]. After more than a century of attempts by renowned mathematicians, the problem of transferring a mass from a starting distribution to a final one, while incurring minimal cost, found its modern formulation and solution in the 1940's: Leonid Kantorovich introduced duality and linear programming to address specifically this problem. In the past twenty years, Ambrosio, Benamou, Brenier, McCann, Gangbo, Jordan, Kinderlehrer, Otto, Rachev, Rüschendorf, Villani and several others drew connections with gradient flows, the heat equation, logarithmic Sobolev inequalities, natural geodesic structures for interpolating densities (*displacement*

interpolation), all with applications to several topics of great physical and engineering importance; see [19], [30], [25], [3], [1], [40], [41], [35]. Besides the intrinsic importance of optimal mass transport to the geometry of spaces and the multitude of applications, a significant impetus for recent work has been the need for effective computation [3], [2], [8] which is often challenging.

Historically, our next subject, *Hilbert's projective metric*, was introduced in 1895 [22]. In the 1950's Garrett Birkhoff and Hans Samelson realized the significance of the metric in proving the existence of positive eigenvectors to linear operators that leave a cone invariant via a contraction principle, see [27]. Birkhoff's version of the Hilbert metric was further developed for nonlinear operators and popularized by Bushell [7]; for an overview and a historical account of key developments on the subject see [26], [20]. Our usage of the Hilbert metric will be in the same spirit, to ascertain and also construct solutions to a system of equations using a contraction mapping principle.

Our third subject, commonly referred to as the *Schrödinger bridge problem*, pertains to steering the density of particles from an initial to a terminal one-time distribution with minimal energy. The original formulation due to Schrödinger in 1931/32 [38], [39] aimed at explaining the conundrum of quantum mechanics by providing a classical stochastic reformulation (a task partially achieved by Nelson [34] 35 years later). Schrödinger's original question was not phrased in terms of "steering," but as the question to identify the most likely flow of the density of particles between two points in time where their distribution is known [42]; see also [18], [4], [23] for early important contributions. It only later became clear that Schrödinger's question can also be posed as a minimum energy steering between marginals [14], [15]. This, in itself, has led to developments of engineering significance [11], [12], [9]. Flows of densities between end-point marginals specified by *Schrödinger bridges* provide what is referred to as *entropic interpolation* [28], [29] and can be seen as a "regularized" counterpart of OMT geodesics. Recent insights in the above references have extended and broadened the scope of *Schrödinger bridges* to problems of great engineering interest.

The structure of our paper and the connection between all three topics will proceed as follows. We first explain the topic of Monge-Kantorovich optimal mass transport for the special case of quadratic cost. We then briefly review Schrödinger bridges and explain how Monge-Kantorovich OMT is a limiting case when stochastic excitation goes to zero, following Mikami, Thieullen, and Léonard [31], [32], [28], [29]. Finally, we bring in the Hilbert metric, and exploit contractiveness of certain maps that underly Schrödinger's problem to construct solutions. In the process, we give an independent proof of a classical result due to Jamison on the existence of solutions to Schrödinger systems. The numerics and the speed of convergence hold the promise that this may provide an attractive approach to solving general OMT problems and displacement interpolation. It certainly provides a numerical approach to Schrödinger bridge problems and entropic interpolation, that has not received much attention in the past.

II. Background

We briefly discuss the three topics of interest and the links between them.

A. *Optimal mass transport*

Gaspar Monge's formulation begins with two non-negative measures μ_0, μ_1 on $\mathbb{R}^n$ having equal total mass. Throughout, these will be probability measures; when absolutely continuous we denote by $\rho_i(x)$ ($i \in \{0, 1\}$) the corresponding densities, i.e., $\mu_i(dx) = \rho_i(x)dx$. The OMT problem seeks a transportation plan (that is, a measurable map) T from $\mathbb{R}^n \to \mathbb{R}^n$ so that

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

and the cost of transportation

$$\int_{\mathbb{R}^n} \frac{1}{2} \|x - T(x)\|^2 \mu_0(dx) \tag{2}$$

is minimal.[1] The composition $\mu_0(T^{-1}(\cdot))$ is commonly referred to as the "push-forward" and denoted by $T_\sharp \mu_0$. Further, for the problem to be well-posed the distributions must have finite second moments.

In Kantorovich's formulation μ_0, μ_1 are thought of as marginals on the product space $\mathbb{R}^n \times \mathbb{R}^n$, and the transportation cost becomes

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{1}{2} \|x - y\|^2 \pi(dxdy) \tag{3}$$

where π, refered to as a "coupling" between μ_0 and μ_1, is a measure on the product space with marginals along the two coordinate directions equal to μ_0 and μ_1, respectively—the space of all such couplings will be denoted by $\Pi(\mu_0, \mu_1)$.

The subtle technical challenge that impeded satisfactory solutions of Monge's problem for over a century is the fact that an optimal transport plan T may not exist in general. When one exists, the optimal coupling has support on the graph of T, and this is the case when the two marginals μ_0 and μ_1 are absolutely continuous [40].

The existence of transportation map T provides a way to connect μ_0 and μ_1 through *(McCann's) displacement interpolation*

$$\mu_t(dx) = ((1 - t)x + tT(x))_\sharp \mu_0(dx), \quad 0 \le t \le 1. \tag{4}$$

This family of distributions, assuming T is an optimal transportation plan, is in fact a geodesic in the space of distributions when metrized by the Wasserstein 2-metric [41], [36], [28], [19]. We note in passing that interest in interpolating distributions arises in physical modeling, resource allocation, spectral analysis, filtering and smoothing of time-series, etc., see e.g. [25], [16], [35].

B. Schrödinger bridges

Schrödinger [38], [39] in 1931/32 sought to reconcile empirical observations of diffusive particles with known priors. More specifically he considered a large number of i.i.d. Brownian particles transitioning between two empirical distributions $\rho_0(x)dx$ and $\rho_1(x)dx$ at two end-points in time, $t = 0$ and $t = 1$, respectively. He hypothesized that $\rho_1(x)$ considerably differs from what we expect according to the law of large numbers, namely

$$\int_{\mathbb{R}^n} q(0, y, 1, x)\rho_0(y)dy,$$

where

$$q(s, y, t, x) = (2\pi)^{-n/2}(t - s)^{-n/2} \exp\left(-\frac{1}{2}\frac{\|x - y\|^2}{(t - s)}\right)$$

denotes the Brownian transition probability kernel, and sought to determine the "most likely" evolution of the cloud of particles from ρ_0 to ρ_1. In the language of large deviations, Schrödinger's question amounts to seeking a probability law on path space that agrees with the observed empirical marginals and is closest to the prior law of the Brownian diffusion in the sense of relative entropy [17].

Schrödinger's solution (called *Schrödinger bridge*), which was formally established by Fortet [18] and later on generalized by Beurling [4] and Jamison [23] to more general reference measures, is a follows (see also Föllmer [17]).

---

[1]G. Monge's original formulation sought to minimize $\int \|x - T(x)\|\mu_0(dx)$ instead of (2); the present work focuses on the latter which also draws on much richer a theory.

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 Markov kernel $q(s, y, t, x)$, there exists a unique pair of σ-finite measures $(\hat{\varphi}_0(x)dx, \varphi_1(x)dx)$ on $\mathbb{R}^n$ such that the measure $\mathcal{P}_{01}$ on $\mathbb{R}^n \times \mathbb{R}^n$ defined by

$$\mathcal{P}_{01}(E) = \int_E q(0, y, 1, x)\hat{\varphi}_0(y)\varphi_1(x)dydx \tag{5}$$

has marginals μ_0 and μ_1. Furthermore, The Schrödinger bridge between μ_0 and μ_1 has as one-time marginal

$$\mathcal{P}_t(dx) = \varphi(t, x)\hat{\varphi}(t, x)dx \tag{6}$$

at time t, with

$$\varphi(t, x) = \int q(t, x, 1, y)\varphi_1(y)dy \tag{7a}$$

$$\hat{\varphi}(t, x) = \int q(0, y, t, x)\hat{\varphi}_0(y)dy. \tag{7b}$$

Given marginal distributions $\mu_0(dx) = \rho_0(x)dx$ and $\mu_1(dx) = \rho_1(x)dx$, the distribution flow (6) 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. The pair $(\hat{\varphi}_0(\cdot), \varphi_1(\cdot)) = (\hat{\varphi}(0, \cdot), \varphi(1, \cdot))$ is the key to the Schrödinger problem; it is the unique solution to the so-called *Schrödinger system* which consists of the two linear equations (7a) with $t = 0$ and (7b) with $t = 1$, together with the two nonlinear coupling conditions

$$\rho_t(x) = \varphi(t, x)\hat{\varphi}(t, x), \quad t = 0, 1. \tag{8}$$

Its existence and uniqueness have been studied for a long time [18], [4], [23], [17].

C. The Hilbert metric

The Hilbert metric was introduced by David Hilbert in 1895 [22]. The form that the metric takes to quantify distance between rays in positive cones, as used herein, is due to Garrett Birkhoff [5]. The importance of the metric and subsequent developments are being discussed in [7]. See also a recent survey on its applications in analysis [27]. The Hilbert metric and certain key facts are presented next.

Definition 2: Let $\mathcal{S}$ be a real Banach space and let $\mathcal{K}$ be a closed solid cone (with nonempty interior $\mathcal{K}^+$) in $\mathcal{S}$, i.e., $\mathcal{K}$ is a closed subset of $\mathcal{S}$ with the properties (i) $\mathcal{K} + \mathcal{K} \subset \mathcal{K}$, (ii) $\alpha\mathcal{K} \subset \mathcal{K}$ for all real $\alpha \geq 0$ and (iii) $\mathcal{K} \cap -\mathcal{K} = \{0\}$. The cone $\mathcal{K}$ induces a partial order relation $\preceq$ in $\mathcal{S}$

$$x \preceq y \Leftrightarrow y - x \in \mathcal{K}.$$

For any $x, y \in \mathcal{K}^+ := \mathcal{K} \backslash \{0\}$, define

$$M(x, y) := \inf\{\lambda \mid x \preceq \lambda y\},$$
$$m(x, y) := \sup\{\lambda \mid \lambda y \preceq x\}$$

with the convention $\inf \emptyset = \infty$. Then the Hilbert metric is defined on $\mathcal{K}^+$ by

$$d_H(x, y) := \log\left(\frac{M(x, y)}{m(x, y)}\right).$$

Strictly speaking, the Hilbert metric is a projective metric since it remains invariant under scaling by positive constants, i.e., $d_H(x, y) = d_H(\lambda x, y) = d_H(x, \lambda y)$ for any $\lambda > 0$ and, thus, it actually measures distance between rays and not elements.

A map $\mathcal{E}: \mathcal{K} \to \mathcal{K}$ is said to be *positive* if $\mathcal{E}: \mathcal{K}^+ \to \mathcal{K}^+$. For such a map define its projective diameter

$$\Delta(\mathcal{E}) := \sup\{d_H(\mathcal{E}(x), \mathcal{E}(y)) \mid x, y \in \mathcal{K}^+\}$$

and the contraction ratio

$$\kappa(\mathcal{E}) := \inf\{\lambda \mid d_H(\mathcal{E}(x), \mathcal{E}(y)) \leq \lambda d_H(x, y) \ \forall x, y \in \mathcal{K}^+\}$$

For a positive map $\mathcal{E}$ which is also linear, we have $\kappa(\mathcal{E}) \leq 1$. In fact, Birkhoff's theorem gives the relation between $\Delta(\mathcal{E})$ and $\kappa(\mathcal{E})$ as

$$\kappa(\mathcal{E}) = \tanh(\frac{1}{4}\Delta(\mathcal{E})), \tag{9}$$

see [5], [7]. Also, another important relation that needs to be pointed out is

$$\Delta(\mathcal{E}) \leq 2\sup\{d_H(\mathcal{E}(x), 1) \mid x \in \mathcal{K}^+\}.$$

This follows directly from the definition of $\Delta(\mathcal{E})$ and the triangular inequality

$$d_H(\mathcal{E}(x), \mathcal{E}(y)) \leq d_H(\mathcal{E}(x), 1) + d_H(1, \mathcal{E}(y)), \quad x, y \in \mathcal{K}^+.$$

D. Connections between the three topics

The Monge-Kantorovich problem (3) turns out to be the limit of the Schrödinger problem as the diffusion coefficient of the reference Brownian motion tends to zero [31], [32], [28], [29], [10], [13]. The precise statement follows.

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

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

and marginals μ_0 and μ_1, and π be the solution to the Monge-Kantorovich problem (3) with the same marginal distributions. Then $\mathcal{P}^\epsilon_{01}$ weakly converges to π as ϵ goes to 0. Moveover, the entropic interpolation $\mathcal{P}^\epsilon_t$ also weakly converge to the displacement interpolation μ_t.

Thus, according to Theorem 3, the displacement interpolation between given marginals can be approximated by their entropic interpolation for a small enough diffusion coefficient.

For both, the Monge-Kantorovich problem as well as the Schrödinger bridge problem, the main objects are probability distributions. These are nonnegative, by definition, and thereby belong to a convex set (simplex) or a cone (if we dispense of the normalization). According to Birkhoff's theorem (9), as we noted, linear endomorphisms of a positive cone are contractive; a fact which is often the key in obtaining solutions of corresponding equations. Thus, the geometry underlying both problems is expected to be impacted by endowing distributions with a suitable version of the Hilbert metric. This is done in the next section.

III. SCHRÖDINGER SYSTEM REDUX

Herein, we provide a new proof for the existence and uniqueness of the solution to the Schrödinger system. To this end, we show that the solution $(\varphi, \hat{\varphi})$ to the Schrödinger system of equations is the unique fixed point of a contraction in the Hilbert metric. Thereby, we also obtain an efficient algorithm to solve the Schrödinger system. We begin by considering ρ_0 and ρ_1 having compact support.

Proposition 1: Suppose that, for $i \in \{0, 1\}$, $S_i \subset \mathbb{R}^n$ is a compact set, $\rho_i(x)dx$ is absolutely continuous probability measure (with respect to the Lebesgue measure) on the σ-field Σ_i of Borel sets of S_i, and that q is an everywhere continuous, strictly positive function on $S_0 \times S_1$. Then, there exist nonnegative functions $\varphi(0, \cdot), \hat{\varphi}(0, \cdot)$ defined on S_0 and $\varphi(1, \cdot), \hat{\varphi}(1, \cdot)$ defined on S_1 satisfying the following so-called *Schrödinger system* of equations:

$$\varphi(0, x) = \int_{S_1} q(0, x, 1, y)\varphi(1, y)dy, \tag{10a}$$

$$\hat{\varphi}(1, x) = \int_{S_0} q(0, y, 1, x)\hat{\varphi}(0, y)dy, \tag{10b}$$

$$\rho_0(x) = \varphi(0, x)\hat{\varphi}(0, x), \tag{10c}$$

$$\rho_1(x) = \varphi(1, x)\hat{\varphi}(1, x). \tag{10d}$$

Moreover, this solution is unique up to multiplication of $\varphi(0, \cdot)$ and $\varphi(1, \cdot)$ and division of $\hat{\varphi}(0, \cdot)$ and $\hat{\varphi}(1, \cdot)$ by the same positive constant.

In order to study the Schrödinger system (10) we consider

$$\mathcal{E} : \ \varphi(1, x_1) \mapsto \varphi(0, x_0) = \int_{S_1} q(0, x_0, 1, x_1)\varphi(1, x_1)dx_1 \tag{11a}$$

$$\mathcal{E}^\dagger : \ \hat{\varphi}(0, x_0) \mapsto \hat{\varphi}(1, x_1) = \int_{S_0} q(0, x_0, 1, x_1)\hat{\varphi}(0, x_0)dx_0 \tag{11b}$$

$$\hat{\mathcal{D}}_{\rho_0} : \ \varphi(0, x_0) \mapsto \hat{\varphi}(0, x_0) = \rho_0(x_0)/\varphi(0, x_0) \tag{11c}$$

$$\mathcal{D}_{\rho_1} : \ \hat{\varphi}(1, x_1) \mapsto \varphi(1, x_1) = \rho_1(x_1)/\hat{\varphi}(1, x_1). \tag{11d}$$

We also define

$$\mathcal{L}_\epsilon^p(S) := \{f \in \mathcal{L}^p(S) \mid f(x) \geq \epsilon, \forall x \in S\},$$

for $\epsilon \geq 0$, and

$$\mathcal{L}_+^p(S) := \bigcup_{\epsilon > 0} \mathcal{L}_\epsilon^p(S),$$

for $p \in \{1, 2, \infty\}$ and $S \in \{S_0, S_1\}$, and endow $\mathcal{L}_+^\infty(S)$ with the Hilbert metric with respect to the natural partial order of inequality between elements (almost everywhere).

Since the kernel q is positive and continuous on the compact set $S_0 \times S_1$, it is bounded from below and above, i.e., there exist $0 < \alpha \leq \beta < \infty$ such that

$$\alpha \leq q(0, x, 1, y) \leq \beta, \ \forall(x, y) \in S_0 \times S_1. \tag{12}$$

It follows that $\mathcal{E}, \mathcal{E}^\dagger$ map nonnegative integrable functions ($\mathcal{L}_0^1$), except the zero function, to functions that are strictly positive and bounded ($\mathcal{L}_+^\infty$). Conversely, since ρ_0 and ρ_1 are nonnegative and integrate to 1 (though, possibly unbounded), $\hat{\mathcal{D}}_{\rho_0}, \mathcal{D}_{\rho_1}$ map positive and bounded functions ($\mathcal{L}_+^\infty$) to nonnegative and integrable ones ($\mathcal{L}_0^1$). Thus, the Schrödinger system relates to the following circular diagram

$$
\begin{array}{ccc}
\hat{\varphi}(0, x_0) & \xrightarrow{\ \mathcal{E}^\dagger\ } & \hat{\varphi}(1, x_1) \\
\hat{\mathcal{D}}_{\rho_0} \uparrow & & \downarrow \mathcal{D}_{\rho_1} \\
\varphi(0, x_0) & \xleftarrow{\ \mathcal{E}\ } & \varphi(1, x_1)
\end{array}
$$

where $\varphi(0, x_0), \hat{\varphi}(1, x_1) \in \mathcal{L}_+^\infty$, while $\varphi(1, x_1), \hat{\varphi}(0, x_0) \in \mathcal{L}_0^1$ on the corresponding domains S_0, S_1, i.e., the circular diagram provides a correspondence between spaces as follows,

$$
\begin{array}{ccc}
\mathcal{L}_0^1(S_0) & \xrightarrow{\ \mathcal{E}^\dagger\ } & \mathcal{L}_+^\infty(S_1) \\
\hat{\mathcal{D}}_{\rho_0} \uparrow & & \downarrow \mathcal{D}_{\rho_1} \\
\mathcal{L}_+^\infty(S_0) & \xleftarrow{\ \mathcal{E}\ } & \mathcal{L}_0^1(S_1).
\end{array}
$$

We will focus on the composition $\mathcal{C} := \mathcal{E}^\dagger \circ \hat{\mathcal{D}}_{\rho_0} \circ \mathcal{E} \circ \mathcal{D}_{\rho_1}$, that is,

$$
\mathcal{C} : \ \mathcal{L}_+^\infty(S_1) \to \mathcal{L}_+^\infty(S_1)
$$

$$
: \ \hat{\varphi}(1, x_1) \xrightarrow{\ \mathcal{E}^\dagger \circ \hat{\mathcal{D}}_{\rho_0} \circ \mathcal{E} \circ \mathcal{D}_{\rho_1}\ } (\hat{\varphi}(1, x_1))_{\text{next}}
$$

and establish the following key lemma.

Lemma 4: The contraction ratio for $\mathcal{C}$ is $\kappa(\mathcal{C}) < 1$.

Proof: When using the Hilbert metric it is important to work on convex sets or cones with non-empty interior, cf. Definition 2. We note that this is not the case with $\mathcal{L}_0^1$ as it has an empty interior. Interestingly, this apparent difficulty in analyzing $\mathcal{C} = \mathcal{E}^\dagger \circ \hat{\mathcal{D}}_{\rho_0} \circ \mathcal{E} \circ \mathcal{D}_{\rho_1}$ can be circumvented if we factor $\mathcal{C}$ in a slightly different manner so that the factors map $\mathcal{L}_+^\infty$ into itself. It is noted that $\mathcal{L}_+^\infty$ is indeed a positive cone satisfying the condition (non-empty interior) in the definition of the Hilbert metric.

To this end, we first define

$$
\mathcal{D} : \ f(x) \mapsto f(x)^{-1}. \tag{13a}
$$

For simplicity of notation we use the same symbol $\mathcal{D}$ for the inversion of functions on either S_0 or S_1. Then, define

$$
\mathcal{E}_{\rho_1} : \ g(x_1) \mapsto \int_{S_1} q(0, x_0, 1, x_1) \rho_1(x_1) g(x_1) dx_1 \tag{13b}
$$

$$
\mathcal{E}_{\rho_1}^\dagger : \ h(x_0) \mapsto \int_{S_0} q(0, x_0, 1, x_1) \rho_0(x_0) h(x_0) dx_0. \tag{13c}
$$

In this way, $\mathcal{C}$ can instead be expressed as the composition

$$
\mathcal{C} = \mathcal{E}_{\rho_1}^\dagger \circ \mathcal{D} \circ \mathcal{E}_{\rho_1} \circ \mathcal{D},
$$

where now all operators map $\mathcal{L}_+^\infty$ to itself (albeit, with possibly different support, S_0 or S_1). Accordingly, the circular diagram can be redrawn as follows:

$$
\begin{array}{ccc}
h(x_0) & \xrightarrow{\ \mathcal{E}_{\rho_0}^\dagger\ } & \hat{\varphi}(1, x_1) \\
\mathcal{D} \uparrow & & \downarrow \mathcal{D} \\
\varphi(0, x_0) & \xleftarrow{\ \mathcal{E}_{\rho_1}\ } & g(x_1).
\end{array}
$$

We now show that

$$
\mathcal{E}_{\rho_1} \circ \mathcal{D} : \ \mathcal{L}_+^\infty(S_1) \to \mathcal{L}_+^\infty(S_0)
$$

is strictly contractive in the Hilbert metric. The fact that, $\mathcal{E}_{\rho_0}^\dagger \circ \mathcal{D} : \ \mathcal{L}_+^\infty(S_0) \to \mathcal{L}_+^\infty(S_1)$ is also contractive proceeds similarly.

Consider two functions $\hat{\varphi}_i(1, \cdot), \in \mathcal{L}_+^\infty(S_1)$, for $i \in \{1, 2\}$, and let

$$
\begin{aligned}
g_i(\cdot) &= \mathcal{D}(\hat{\varphi}_i(1, \cdot)) = \hat{\varphi}_i(1, \cdot)^{-1}, \\
\varphi_i(0, \cdot) &= \mathcal{E}_{\rho_1}(g_i(\cdot)).
\end{aligned}
$$

Since $M(g_1, g_2) = (m(g_1^{-1}, g_2^{-1}))^{-1}$, it follows that $\mathcal{D}$ is an isometry in the Hilbert metric. Next consider the map $\mathcal{E}_{\rho_1}$. The projective diameter of $\mathcal{E}_{\rho_1}$ is

$$\Delta(\mathcal{E}_{\rho_1}) \leq 2\sup\{d_H(\mathcal{E}_{\rho_1}(g), 1) \mid g \in \mathcal{L}_+^\infty(S_1)\}.$$

But since for any $g \in \mathcal{L}_+^\infty$

$$\alpha \int_{S_1} \rho_1(x_1)g(x_1)dx_1 \leq \mathcal{E}_{\rho_1}(g) \leq \beta \int_{S_1} \rho_1(x_1)g(x_1)dx_1,$$

$$\Delta(\mathcal{E}_{\rho_1}) \leq 2\sup\{\log(\frac{\sup \mathcal{E}_{\rho_1}(g)}{\inf \mathcal{E}_{\rho_1}(g)}) \mid g \in \mathcal{L}_+^\infty(S_1)\}$$

$$\leq 2\log(\frac{\beta}{\alpha}) < \infty.$$

Hence,

$$\kappa(\mathcal{E}_{\rho_1} \circ \mathcal{D}) = \tanh(\frac{1}{4}\Delta(\mathcal{E}_{\rho_1}))$$

$$\leq \tanh(\frac{1}{2}\log(\frac{\beta}{\alpha})) < 1.$$

Similarly we have

$$\kappa(\mathcal{E}_{\rho_0}^\dagger \circ \mathcal{D}) \leq \tanh(\frac{1}{2}\log(\frac{\beta}{\alpha})) < 1.$$

Combining the above we have that $\kappa(\mathcal{C}) \leq \tanh^2(\frac{1}{2}\log(\frac{\beta}{\alpha})) < 1$. ∎

Proof of Proposition 1: Let $\hat{\varphi}_0(1, \cdot)$ be any function in $\mathcal{L}_+^\infty(S_1)$ and consider the sequence

$$\hat{\varphi}_k(1, \cdot) = \mathcal{C}^k\hat{\varphi}_0(1, \cdot),$$

for $k = 1, 2, \ldots$. This is Cauchy with respect to the Hilbert metric in $\mathcal{L}_+^\infty(S_1)$ because $\mathcal{C}$ is strictly contractive. We normalize so as to obtain the unit-norm sequence

$$g_k(x) = \frac{\hat{\varphi}_k(1, x)}{\|\hat{\varphi}_k(1, \cdot)\|_2}, \quad k = 1, 2, \ldots,$$

which can be done since $\mathcal{L}^\infty(S_1) \subset \mathcal{L}^2(S_1)$. The sequence $\{g_k\}$ is also Cauchy with respect to the Hilbert metric in

$$E = \{f \in \mathcal{L}_+^\infty(S_1) \mid \|f\|_2 = 1\}.$$

It is easy to see from (12) that

$$\frac{\sup_x(g_k(x))}{\inf_x(g_k(x))} \leq \frac{\beta}{\alpha}. \tag{14}$$

Since $\|g_k\|_2 = 1$ we have

$$\sup_x(g_k(x)) \geq \frac{1}{\sqrt{|S_1|}} \geq \inf_x(g_k(x)), \tag{15}$$

where $|S_1|$ denotes the Lebesgue measure of S_1. Combining (14) and (15) we easily deduce that $\{g_k\}$ is uniformly bounded from both, above and below; more precisely,

$$\sup_x(g_k(x)) \leq \frac{\alpha}{\beta\sqrt{|S_1|}}, \quad \text{while} \quad \inf_x(g_k(x)) \geq \frac{\beta}{\alpha\sqrt{|S_1|}}. \tag{16}$$

For any two elements g_k, g_ℓ of the sequence,

$$m(g_k, g_\ell) \leq 1 \leq M(g_k, g_\ell),$$

and therefore

$$
\begin{aligned}
\|g_k - g_\ell\|_2 &= \left\{ \int_{S_1} (g_k(x) - g_\ell(x))^2 dx \right\}^{1/2} \\
&= \left\{ \int_{S_1} (g_k/g_\ell - 1)^2 g_\ell(x)^2 dx \right\}^{1/2} \\
&\leq \left\{ \int_{S_1} (M(g_k, g_\ell) - m(g_k, g_\ell))^2 g_\ell(x)^2 dx \right\}^{1/2} \\
&= M(g_k, g_\ell) - m(g_k, g_\ell) \\
&= \{\exp(d_H(g_k, g_\ell)) - 1\} m(g_k, g_\ell) \\
&\leq \exp(d_H(g_k, g_\ell)) - 1.
\end{aligned}
$$

It follows that $\{g_k\}$ is also Cauchy in the 2-norm, and thus, converges to a unique $g \in \mathcal{L}^2(S_1)$ which satisfies $\|g\|_2 = 1$ as well as inherits the bounds in (16). Therefore, $g \in \mathcal{L}_+^\infty(S_1)$.

We next show that $\{g_k\}$ converges to g with respect to the Hilbert metric in E as well, i.e., that $d_H(g_k, g) \to 0$ as $k \to \infty$. Since q is uniformly continuous, given any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$
|q(0, y, 1, x_1) - q(0, y, 1, x_2)| < \alpha\varepsilon\sqrt{|S_1|}
$$

for any $\|x_1 - x_2\| < \delta$. It follows that for each k,

$$
\begin{aligned}
|g_k(x_1) - g_k(x_2)| &= \frac{1}{\|\hat{\varphi}_k(1, \cdot)\|_2} |\hat{\varphi}_k(1, x_1) - \hat{\varphi}_k(1, x_2)| \\
&\leq \frac{1}{\|\hat{\varphi}_k(1, \cdot)\|_2} \int_{\mathbb{R}^n} |q(0, y, 1, x_1) - q(0, y, 1, x_2)| \hat{\varphi}_k(0, y) dy \\
&< \frac{1}{\alpha\sqrt{|S_1|}\|\hat{\varphi}_k(0, \cdot)\|_1} \|\hat{\varphi}_k(0, \cdot)\|_1 \alpha\varepsilon\sqrt{|S_1|} = \varepsilon,
\end{aligned}
$$

which shows that g_k is uniformly continuous. Moreover, because the value of δ doesn't depend on k, $\{g_k\}$ is in fact uniformly equicontinuous. Considering that this sequence is also uniformly bounded (16), we apply the Arzelà-Ascoli theorem [37] to deduce that it has a uniformly convergent subsequence $\{g_{n_k}\}$. On the other hand, we already know that $\{g_k\}$ converges to g in the 2-norm. Hence, $\{g_{n_k}\}$ converges uniformly to g and therefore g is uniformly continuous. This implies $\|g_{n_k}/g - 1\|_\infty \to 0$, from which $d_H(g_{n_k}, g) \to 0$ follows easily. Combining this with the fact that $\{g_k\}$ is Cauchy with respect to the Hilbert metric we conclude that $d_H(g_k, g) \to 0$.

Now let $\hat{\mathcal{C}}$ be the projection of $\mathcal{C}$ onto E, namely, $\hat{\mathcal{C}}(f) = \mathcal{C}(f)/\|\mathcal{C}(f)\|_2$, then $\hat{\mathcal{C}}$ is contractive in E with respect to the Hilbert metric, and it is therefore continuous. It follows that

$$
g = \lim_{k \to \infty} g_{k+1} = \lim_{k \to \infty} \hat{\mathcal{C}}(g_k) = \hat{\mathcal{C}}(\lim_{k \to \infty} g_k) = \hat{\mathcal{C}}(g). \tag{17}
$$

Here the limit is taken with respect to the Hilbert metric. Due to the contractiveness of $\mathcal{C}$, the limit g is independent of the choice of initial value $\hat{\varphi}_0(1, \cdot)$ and it is therefore unique.

We finally argue that $g = \mathcal{C}(g)$. Noting that $g = \hat{\mathcal{C}}(g)$, we have $\mathcal{C}(g) = \lambda g$ for some fix $\lambda = \|\mathcal{C}(g)\|_2 > 0$. Since $\langle \phi, \hat{\mathcal{D}}_{\rho_0}\phi \rangle = \|\rho_0\|_1 = 1$ for any $\phi \in \mathcal{L}_+^\infty(S_0)$, we have

$$
\begin{aligned}
\|\rho_0\|_1 &= \langle \mathcal{E} \circ \mathcal{D}_{\rho_1} g, \ \hat{\mathcal{D}}_{\rho_0} \circ \mathcal{E} \circ \mathcal{D}_{\rho_1} g \rangle \\
&= \langle \mathcal{D}_{\rho_1} g, \ \underbrace{\mathcal{E}^\dagger \circ \hat{\mathcal{D}}_{\rho_0} \circ \mathcal{E} \circ \mathcal{D}_{\rho_1}}_{\mathcal{C}} g \rangle \\
&= \langle \mathcal{D}_{\rho_1} g, \ \lambda g \rangle \\
&= \lambda \|\rho_1\|_1,
\end{aligned}
$$

and it follows that $\lambda = 1$. Now let

$$
\begin{aligned}
\hat{\varphi}(1, \cdot) &= g \\
\varphi(1, \cdot) &= \mathcal{D}_{\rho_1} g \\
\varphi(0, \cdot) &= \mathcal{E} \circ \mathcal{D}_{\rho_1} g \\
\hat{\varphi}(0, \cdot) &= \hat{\mathcal{D}}_{\rho_0} \circ \mathcal{E} \circ \mathcal{D}_{\rho_1} g,
\end{aligned}
$$

then $\varphi(0, \cdot), \hat{\varphi}(0, \cdot), \varphi(1, \cdot), \hat{\varphi}(1, \cdot)$ is a solution to the Schrödinger system (10). The "uniqueness" follows from the uniqueness of g.

∎

Proposition 1 can be rephrased as follows.

Proposition 2: Suppose $S_0 \subset \mathbb{R}^n, S_1 \subset \mathbb{R}^n$ are compact sets, μ_0 and μ_1 are absolutely continuous probability measures on the σ-fields Σ_0, Σ_1 of Borel sets of S_0 and S_1, respectively, and that q is an everywhere continuous, strictly positive function on $S_0 \times S_1$, then there is a unique pair π, ν of measures on $\Sigma_0 \times \Sigma_1$ for which

1) π is a probability measure and ν is a finite product measure;
2) $\pi(E_0 \times S_1) = \mu_0(E_0)$, $\pi(S_0 \times E_1) = \mu_1(E_1)$,
   for some $E_0 \in \Sigma_0$, $E_1 \in \Sigma_1$;
3) $d\pi/d\nu = q$.

Proof: Apply Proposition 1 to obtain a solution $\varphi(0, \cdot)$, $\hat{\varphi}(0, \cdot)$, $\varphi(1, \cdot)$, $\hat{\varphi}(1, \cdot)$ to the Schrödinger system (10). Then, define ν by the formula

$$
\nu(E_0 \times E_1) = \left(\int_{E_0} \hat{\varphi}(0, x) dx \right) \left(\int_{E_1} \varphi(1, x) dx \right), \tag{18}
$$

and π by

$$
\pi(E_0 \times E_1) = \int_{E_0 \times E_1} \hat{\varphi}(0, x) q(0, x, 1, y) \varphi(1, y) dx dy \tag{19}
$$

for any $E_0 \in S_0$, $E_1 \in S_1$. It is easy to see π and ν satisfy all the three properties in the statement. This proves the existence part of the proposition. On the other hand, any pair π and ν satisfying the above three requirements can be written as (18) and (19). The uniqueness of the solution π and ν follows from the uniqueness of the solution to the Schrödinger system (10) in Proposition 1. ∎

Remark 5: It is easy to see that in the above, μ_0, μ_1 don't need to be probability measures, only to have equal mass, i.e. $\mu_0(S_0) = \mu_1(S_1)$. The statements of the propositions hold verbatim.

The results in the above lemma and proposition can be extended to μ_0, μ_1 having non-compact support. In fact, Proposition 2 and then, the following extension to general measures (Theorem 6) were given by Jamison in [23]. Jamison's proof of Proposition 2 does not invoke the Hilbert metric or contractiveness. The proof of Theorem 6 is reworked in the Appendix for completeness.

Theorem 6: Suppose μ_0 and μ_1 are absolutely continuous probability measures on the σ-field Σ of Borel sets of $\mathbb{R}^n$, and q is an everywhere continuous, strictly positive function on $\mathbb{R}^n \times \mathbb{R}^n$, then there is a unique pair π, ν of measures on $\Sigma \times \Sigma$ for which

1) π is a probability measure and ν is a σ-finite product measure;
2) $\pi(E \times \mathbb{R}^n) = \mu_0(E)$, $\pi(\mathbb{R}^n \times E) = \mu_1(E)$, $E \in \Sigma$;
3) $d\pi/d\nu = q$.

IV. Computational algorithm

Given marginal probability measures $\mu_0(dx) = \rho_0(x)dx$ and $\mu_1(dx) = \rho_1(x)dx$ on $\mathbb{R}^n$, we begin by specifying a compact subset $S \subset \mathbb{R}^n$ that supports most of the two densities, i.e., such that $\mu_0(S)$ and $\mu_1(S)$ are both $\geq 1 - \delta$, for sufficiently small value $\delta > 0$. We treat the restriction to S for both, after normalization so that they integrate to 1, as the end-point marginals for which we wish to construct the corresponding entropic and displacement interpolation. Thus, for the purposes of this section and subsequent examples/applications both, ρ_0 and ρ_1 are supported on a compact subset $S \in \mathbb{R}^n$.

It is important to consider the potential spread of the mass along entropic interpolation and the need for S to support the flow without "excessive" constraints at the boundary. Thus, depending on the application (imaging, spectral analysis, etc.) a sufficiently conservative choice for S, beyond what is suggested in the previous paragraph, may be in order.

Next, we discretize in space and represent functions $\varphi(i, x)$, $\hat{\varphi}(i, x)$ ($i \in \{0, 1\}$) using (column) vectors ϕ_i, $\hat{\phi}_i$, e.g., $\phi_i(k) = \varphi(i, x_k)$ for a choice of sample points $x_k \in S$, $k = 1, \ldots, N$ and, likewise, ρ_0, ρ_1 (column) vectors representing the sampled values of the two densities. Then, we recast (11) as operations on these vectors. Accordingly,

$$\mathcal{E} : \ \phi_1 \mapsto \phi_0 = Q\phi_1 \tag{20a}$$

$$\mathcal{E}^\dagger : \ \hat{\phi}_0 \mapsto \hat{\phi}_1 = Q^\dagger \hat{\phi}_0 \tag{20b}$$

$$\hat{\mathcal{D}}_{\rho_0} : \ \phi_0 \mapsto \hat{\phi}_0 = \rho_0 \oslash \phi_0 \tag{20c}$$

$$\mathcal{D}_{\rho_1} : \ \hat{\phi}_1 \mapsto \phi_1 = \rho_1 \oslash \hat{\phi}_1, \tag{20d}$$

using the same symbols for the corresponding operators, and using $\oslash$ to denote entry-wise division between two vectors, i.e., $a \oslash b := [a_i/b_i]$. Accordingly, here, Q represents a matrix. The values of its entries depend on the diffusivity ϵ. By iterating the discretized action of $\mathcal{C}$, we obtain a fixed-point pair of vectors $(\phi_i, \hat{\phi}_i)$. From this we can readily construct the entropic interpolation between the marginals discretizing (7) for intermediate points in time. Since the contraction ratio $\kappa(\mathcal{C}) < 1$, the worst case convergence speed is linear. As noted earlier, the displacement interpolation can be approximated by the entropic interpolation for small enough ϵ (Theorem 3).

We note that, when the noise intensity ϵ is too small, numerical issues may arise due to machine precision. One way to mediate this effect, especially regarding (20c-20d), is to store and operate with the logarithms of elements in $Q, \rho_i, \phi_i, \hat{\phi}_i$, denoted by $lQ, l\rho_i, l\phi_i, l\hat{\phi}_i$ ($i \in \{0, 1\}$). More specifically, let $(lQ)_{jk} = \log Q_{jk}$ and set

$$(l\rho_i)_j = \begin{cases} \log(\rho_i)_j & \text{if } (\rho_i)_j > 0, \\ -10000 & \text{otherwise}, \end{cases}$$

(since, e.g., in double precision floating point numbers $< 10^{-308}$ are taken to be zero). Carrying out operations in (20) in logarithmic coordinates, $\hat{\mathcal{D}}_{\rho_0}$ becomes

$$(l\hat{\phi}_0)_j = (l\rho_0)_j - (l\phi_0)_j,$$

and similarly for $\mathcal{D}_{\rho_1}$. The map $\mathcal{E}^\dagger$ becomes

$$(l\hat{\phi}_1)_k = \log \sum_j \exp(lQ_{jk} + (l\hat{\phi}_0)_j)$$

$$= M_k + \log \sum_j \exp(lQ_{jk} + (l\hat{\phi}_0)_j - M_k),$$

(and similarly for $\mathcal{E}$) where $M_k = \max_j\{lQ_{jk} + (l\hat{\phi}_0)_j\}$. In this way the dominant part of the power index, which is M_k, is respected.
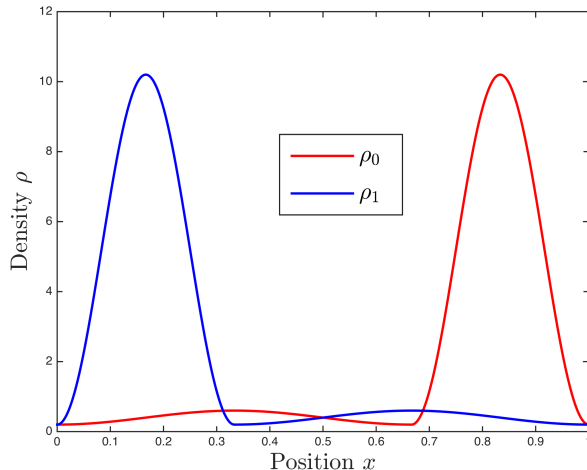
Fig. 1: Marginal distributions

V. NUMERICAL EXAMPLES

In this section we study three examples to illustrate the algorithm. The first example is a Schrödinger bridge problem on one dimensional space. Right after that is an OMT version of it with the same marginal distributions. We observe that as the diffusivity goes to zero, the solution of the Schrödinger bridge problem converges to the solution of the OMT problem. The last example highlights interpolation between images where computations have been carried out using the above algorithm.

A. One dimensional Schrödinger bridge

Consider a collection of Brownian particles in one dimension and let

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

and

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

represent their distribution at $t = 0$ and $t = 1$, respectively; see Figure 1. We seek to construct entropic interpolants of the two densities over the interval $[0, 1]$. We do so by determining Schrödinger bridges between ρ_0 and ρ_1 with diffusion kernel

$$
q(s, x, t, y) = \frac{1}{\sqrt{2\pi(t - s)\epsilon}} \exp\left[-\frac{(x - y)^2}{2(t - s)\epsilon} \right], \ s < t,
$$

for a range of values for the diffusion coefficient $\sqrt{\epsilon}$ of the underlying Brownian motion. Figures 2 (a-d) depict the corresponding interpolation flow taking $\sqrt{\epsilon} = 0.5, \ 0.2, \ 0.1, \ 0.01$, respectively. It is worth noting that, for larger diffusion coefficient, starting at $t = 0$, the flow of density spreads out before "reassembling" at the other end-point $t = 1$. We also observe an apparent time-symmetry of the evolution – a point that was central to Schrödinger's thinking already in [38].

(a) $\sqrt{\epsilon} = 0.5$

(b) $\sqrt{\epsilon} = 0.2$

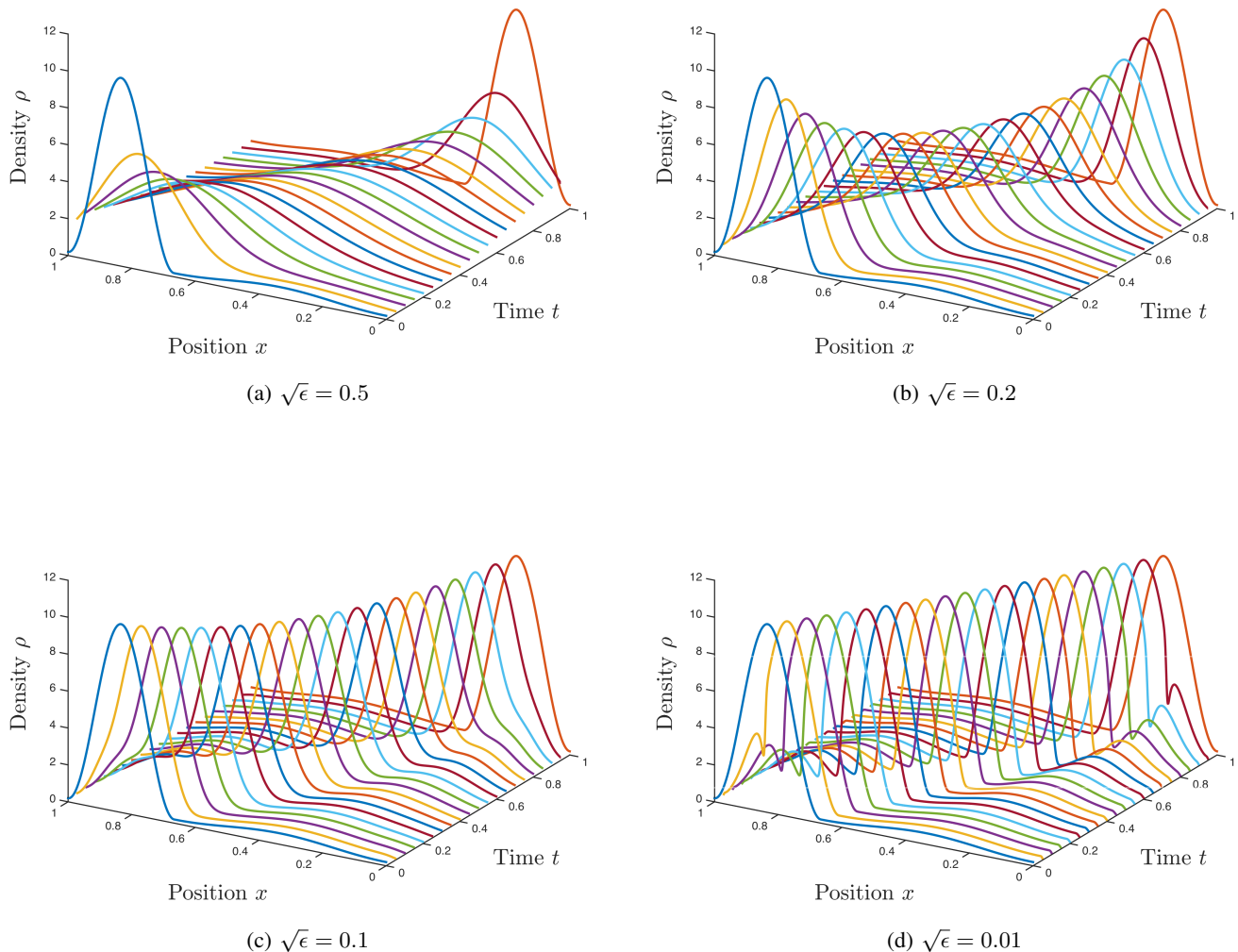(c) $\sqrt{\epsilon} = 0.1$

(d) $\sqrt{\epsilon} = 0.01$

Fig. 2: Entropic interpolation

B. One dimensional OMT

Consider a collection of particles with the same marginal distributions ρ_0 and ρ_1 as the previous example. However, instead of doing Brownian motion, we assume the particles are driven by some unknown "deterministic" forces. Under this assumption, the "displacement interpolation" based on OMT is a reasonable choice. For the monotonicity of the optimal map, a "closed form" solution is available for one dimensional OMT problem. The optimal map $y = T(x)$ satisfies

$$\int_{-\infty}^{x} \rho_0(y)dy = \int_{-\infty}^{T(x)} \rho_1(y)dy, \tag{21a}$$

and the interpolation flow ρ_t, $0 \leq t \leq 1$ satisfies

$$\int_{-\infty}^{x} \rho_0(y)dy = \int_{-\infty}^{(1-t)x+tT(x)} \rho_t(y)dy. \tag{21b}$$

Figure 3 gives the displacement interpolation by using the exact form of the solution in (21). This can be compared to the entropic interpolation with $\sqrt{\epsilon} = 0.01$ (Figure 2(d)); for more detailed comparison
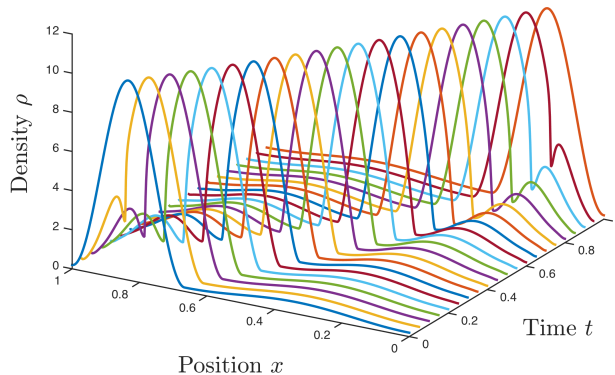
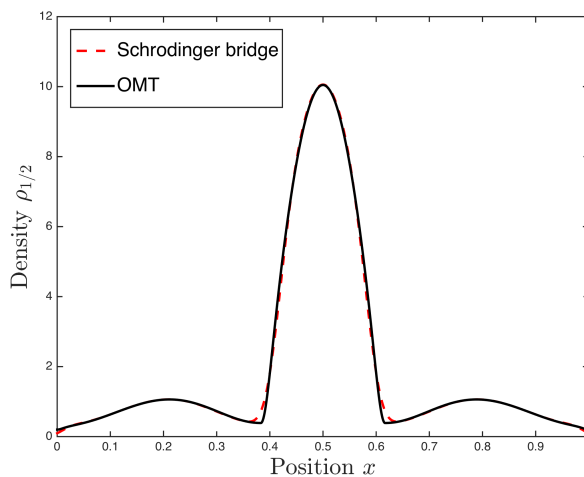Fig. 3: Dispacement interpolation (OMT)

Fig. 4: Comparison between OMT and Schrödinger bridge

we plot the corresponding densities at $t = 1/2$ in Figure 4. From an applications standpoint, it is worth noting that entropic interpolation with a "small" (but not insignificant) diffusion coefficient suppresses potentially spurious peaks as seen by comparing Figure 3 with Figure 2(c).

C. Image morphing

In this subsection, we consider interpolation/morphing of 2D images. When suitably normalized, these can be viewed as probability densities on $\mathbb{R}^2$. Interpolation is important in many applications. One such application is Magnetic Resonance Imaging (MRI) where due to cost and time limitations, a limited number of slices are scanned. Suitable interpolation between the 2D-slices may yield a better 3D reconstruction.

Figure 5 shows the two brain images that we seek to interpolate. The data is available as mri.dat in Matlab®. Figure 6 compares displacement interpolants at $t = 0.2,\ 0.4,\ 0.6,\ 0.8$, respectively, based on solving a Schrödinger bridge problem with diffusivity $\epsilon = 0.01$ using our numerical algorithm. For comparison, we display in Figure 7 another set of interpolants corresponding to larger diffusivity, namely, $\epsilon = 0.04$. As expected, we observe a more blurry set of interpolants due to the larger diffusivity.
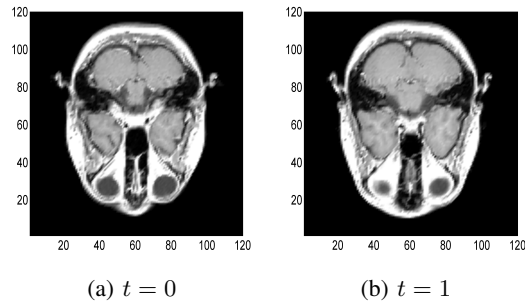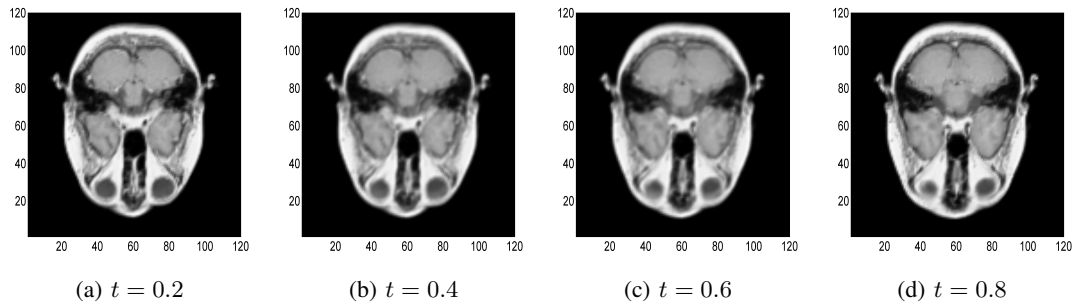
(a) $t = 0$ (b) $t = 1$

Fig. 5: MRI slices at two different points

(a) $t = 0.2$ (b) $t = 0.4$ (c) $t = 0.6$ (d) $t = 0.8$
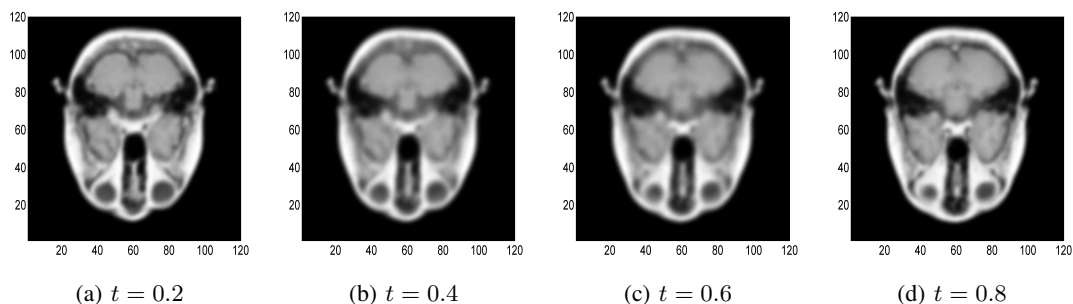
Fig. 6: Interpolation with $\epsilon = 0.01$

VI. CONCLUDING REMARKS

The present paper is concerned with the problem of interpolating distributions. Linear interpolation,

$$\mu_t(dx) = ((1 - t)\rho_0(x) + t\rho_1(x))\,dx, \text{ for } t \in [0, 1],$$

while widely used in signal analysis, is deeply pathological from many different angles. For instance, in the context of image processing and signal analysis, linear interpolation creates fade-in fade-out effects. To see this consider two Gaussian distributions with the same variance and sufficiently different means. Linear interpolation is bi-modal and the two peaks of μ_t trade-off relative significance as t changes from 0 to 1. Clearly, this is an undesirable feature. If on the other hand, μ_t represents the density of particles, linear interpolation of one-time marginals requires infinite flow velocity (suggesting teleportation rather than a physically realistic flow of the particles). Likewise, this feature is unreasonable on physical grounds.

Instead, and for those very reasons, other methods of interpolation have been pursued with the geometry

(a) $t = 0.2$ (b) $t = 0.4$ (c) $t = 0.6$ (d) $t = 0.8$

Fig. 7: Interpolation with $\epsilon = 0.04$

of OMT taking a prominent role, see e.g., [30], [43], [21], [24], [35]. In this, the displacement interpolation path for one-time densities (4),

$$\mu_t(dx) = ((1-t)x + tT(x))_\sharp \, \mu_0(dx), \text{ for } t \in [0,1],$$

requires computing the optimal transport map T which in itself is a challenging task. More recently, it was noted that the OMT problem with quadratic cost is a limiting form of the stochastic control problem to steer a controlled diffusion between two end-point marginals with minimum energy [31], [32], [28], [29], [14], [15]. The latter is equivalent to the problem posed by Erwin Schrödinger in 1931, the Schrödinger bridge problem, namely to identify the most likely flow of particles between two empirical one-time marginals. The connection between OMT and Schrödinger bridges has led to a fast developing circle of ideas with important implications in stochastic control and many other fields. In particular, the problem to steer deterministic systems with random initial and terminal conditions is akin to OMT and can be solved with similar methods [13]. Equally important, however, is the reverse implication: techniques from the theory of Schrödinger bridges can be used to solve the OMT as the entropic interpolation

$$\mu_t(dx) = \hat{\varphi}(t,x)\varphi(t,x)dx, \text{ for } t \in [0,1],$$

with $\hat{\varphi}(t,x)$, $\varphi(t,x)$ obtained by solving the Schrödinger system (10) approximates the displacement interpolation as the power of the stochastic excitation ϵ (see Theorem (3)) tends to zero [10], [13].

The purpose of the present work has been to draw attention to points of contact between OMT, the Schrödinger bridge problem, and the Hilbert metric. The Hilbert metric allows an independent direct approach to solving the Schrödinger system of equations as it renders a certain key map contractive ($\mathcal{C}$ in Section III). Fixed points of this map provide a solution to the Schrödinger bridge problem and the contractiveness ensures linear convergence. We expect that further numerical studies and specialized software will permit applying the approach to problems of substantially large scale.

VII. Appendix

[Proof of Theorem 6 - largely based on Jamison's [23] arguments] Let $A_1 \subset A_2 \subset A_3 \ldots$ be an increasing sequence of compact Borel sets and $B_1 \subset B_2 \subset B_3 \ldots$ be an increasing sequence of compact Borel sets such that

$$\mu_0(A_k) = 1 - \frac{1}{(k+1)^2},$$
$$\mu_1(B_k) = 1 - \frac{1}{(k+1)^2}.$$

Let $C_k = A_k \times B_k$, and let $\Sigma_0^k = \Sigma \cap A_k = \{E \cap A_k \mid E \in \Sigma\}$, $\Sigma_1^k = \Sigma \cap B_k$, then $\Sigma_0^k \times \Sigma_1^k$ is the class of Borel subsets of C_k. On C_k, by Theorem 2 and Remark 5, for any pair of A_k, B_k, there exists a finite product measure ν^k on $\Sigma_0^k \times \Sigma_1^k$ and a measure π^k on $\Sigma_0^k \times \Sigma_1^k$ such that

$$\pi^k(E_0 \times B_k) = \mu_0(E_0), \ \forall E_0 \in \Sigma_0^k$$
$$\pi^k(A_k \times E_1) = \mu_1(E_1), \ \forall E_1 \in \Sigma_1^k$$

and

$$\frac{d\pi^k}{d\nu^k} = q \text{ on } C_k. \tag{22}$$

The measures π^k and ν^k can be extended to the space $\mathbb{R}^n \times \mathbb{R}^n$ by setting them equal to 0 on sets disjoint from C_k. Fix m and construct a set Π_m of measures in $\Sigma_0^m \times \Sigma_1^m$ by restricting $\{\pi^k\}$ to C_m. The set Π_m

is of course tight since C_m is compact. The set Π_m is also uniformly bounded in total variation norm since

$$\pi^k(C_m) \leq \pi^k(A_m \times B_k) = \mu_0(A_m) = 1 - \frac{1}{(m+1)^2}. \tag{23}$$

By (extension of) Prokhorov's theorem [6][Theorem 8.6.2.], the set Π_m is sequential compact with respect to weak topology and therefore has a weakly convergent subsequence. This implies the existence of a measure π on $\Sigma \times \Sigma$ whose restriction to $\Sigma_0^m \times \Sigma_1^m$ is for each m the weak limit relative to $C(C_m)$ of a subsequence in Π_m. Let π^{k_ℓ} denote this subsequence, then $\int g d\pi^{k_\ell} \to \int g d\pi$ for any continuous g on $\mathbb{R}^n \times \mathbb{R}^n$ with compact support. We next show $\int g d\pi^{k_\ell} \to \int g d\pi$ for any bounded continuous g. It is enough to show π is bounded. Combine (23) and

$$
\begin{aligned}
\pi^k(C_m) &\geq \pi^k(A_k \times B_m) + \pi^k(A_m \times B_k) - \pi^k(C_k) \\
&= 1 - \frac{1}{(m+1)^2} + 1 - \frac{1}{(m+1)^2} - (1 - \frac{1}{(k+1)^2}) \\
&\geq 1 - \frac{2}{(m+1)^2}
\end{aligned}
$$

we obtain

$$1 - \frac{2}{(m+1)^2} \leq \pi(C_m) \leq 1 - \frac{1}{(m+1)^2},$$

and

$$\pi(C_m \backslash C_{m-1}) \leq -\frac{1}{(m+1)^2} + \frac{2}{m^2} \propto \frac{1}{m^2}.$$

It follows that

$$\pi(\bigcup_{m=1}^{\infty} C_m) = \pi(C_1 \cup (\bigcup_{m=2}^{\infty} C_m \backslash C_{m-1})) < \infty,$$

and therefore π is bounded (obviously $\pi(\mathbb{R}^n \backslash \bigcup_{m=1}^{\infty} C_m) = 0$ since $\pi^k(\mathbb{R}^n \backslash \bigcup_{m=1}^{\infty} C_m) = 0$). Thus, the sequence $\{\pi^{k_\ell}\}$ weakly converges to π and π is a probability measure (total mass is 1). Let μ_0^k, μ_1^k be the marginals of π^k, then it is clear $\{\mu_0^k\}$ and $\{\mu_1^k\}$ weakly converge to μ_0 and μ_1 respectively. On the other hand, $\{\mu_0^k\}$ and $\{\mu_1^k\}$ weakly converge to the marginals of π since $\{\pi^k\}$ weakly converge to π. Therefore μ_0 and μ_1 must be the marginals of π.

So far we establish 2) and the first half of 1). We now show 3) and the second half of 1). Fix m and let f be a bounded continuous function with support in C_m. The restriction of f/q to C_m is also a bounded continuous function, so by (22)

$$\int f d\nu^{k_\ell} = \int (f/q) d\pi^{k_\ell} \to \int (f/q) d\pi \tag{24}$$

as $\ell \to \infty$. This shows that the restriction of ν^{k_ℓ} to C_m converges weakly to a finite measure ν_m (in fact $d\nu_m = d\pi/q$ on C_m), from which we can construction a σ-finite measure ν whose restriction to C_m is ν_m, $m = 1, 2, \ldots$. In view of (24) we conclude that $d\nu/d\pi = 1/q$ and $d\pi/d\nu = q$ follows. Each ν_m is a product measure based on the fact ν^k is a product measure. Therefore ν must be a product measure. This completes the proof of the existence of π and ν that satisfy all the 3 properties.

To establish that ν and π are unique, assume that ν' is a product measure and π' a probability measure for which 1),2) and 3) hold. Then

$$\mu_0(E) = \int_{E \times \mathbb{R}^n} q d\nu = \int_{E \times \mathbb{R}^n} q d\nu' \tag{25}$$

and

$$\mu_1(E) = \int_{\mathbb{R}^n \times E} q \, d\nu = \int_{\mathbb{R}^n \times E} q \, d\nu'$$

for each $E \in \Sigma$. Suppose $\nu = \nu_0 \times \nu_1$, $\nu' = \nu_0' \times \nu_1'$. Let $h_0(x) = \int q(x, y) \nu_1(dy)$, $x \in \mathbb{R}^n$, $h_1(y) = \int q(x, y) \nu_0(dx)$, $y \in \mathbb{R}^n$, and let $h(x, y) = h_0(x) h_1(y)$, $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$. Let k_0, k_1 and k be similarly defined but with ν_0', ν_1' replacing ν_0, ν_1 respectively. Let g_0 and g_1 be bounded Σ-measurable functions on $\mathbb{R}^n$, and let $g(x, y) = g_0(x) g_1(y)$, $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$. By virtue of (25), we have $\int g_0 \, d\mu_0 = \int g_0 h_0 \, d\nu_0$ and $\int g_1 \, d\mu_1 = \int g_1 h_1 \, d\nu_1$. Multiplying corresponding sides of these two equations, we have

$$\int g \, d(\mu_0 \times \mu_1) = \int g h \, d\nu.$$

Since h is strictly positive, we can rewrite this as

$$\int (g/h) \, d(\mu_0 \times \mu_1) = \int g \, d\nu. \tag{26}$$

Similarly

$$\int (g/k) \, d(\mu_0 \times \mu_1) = \int g \, d\nu'. \tag{27}$$

The definition of $\Sigma \times \Sigma$ as the σ-field generated by the field of finite disjoint unions of rectangles $E \times F$ with $E, F \in \Sigma$ ensures that (26) and (27) hold for all non-negative $\Sigma \times \Sigma$-measurable functions g. Let σ_0 and σ_1 be bounded Σ-measurable functions on $\mathbb{R}^n$, and let $\sigma(x, y) = \sigma_0(x) + \sigma_1(y)$, $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$. Then

$$\begin{aligned}
\int \sigma \, d(\mu_0 \times \mu_1) &= \int \sigma_0 \, d\mu_0 + \int \sigma_1 \, d\mu_1 \\
&= \int \sigma q \, d\nu = \int (\sigma q/h) \, d(\mu_0 \times \mu_1)
\end{aligned}$$

by virtue of (26) and (27). Using ν' instead of ν, we obtain similarly $\int \sigma \, d(\mu_0 \times \mu_1) = \int (\sigma q/k) \, d(\mu_0 \times \mu_1)$. Therefore we conclude that

$$\int (\sigma q/h) \, d(\mu_0 \times \mu_1) = \int (\sigma q/k) \, d(\mu_0 \times \mu_1). \tag{28}$$

Since σ is bounded and $q \, d\nu$ is a probability measure, the common value of the two sides of (28) is finite. Thus we get

$$\int \sigma q (1/h - 1/k) \, d(\mu_0 \times \mu_1) = 0.$$

In particular, this last equation holds when we take a specific σ as

$$\sigma(x, y) = \frac{1/h_0(x)}{1/h_0(x) + 1/k_0(x)} - \frac{1/k_1(y)}{1/h_1(y) + 1/k_1(y)} \tag{29}$$

from which we deduce

$$\int q (1/h - 1/k)^2 \, d(\mu_0 \times \mu_1) = 0. \tag{30}$$

Since $q > 0$, we must have $h = k$ on the support of $\mu_0 \times \mu_1$. It now follows from (26) and (27) that $\nu = \nu'$, and it follows from $d\pi/d\nu = d\pi'/d\nu'$ that $\pi = \pi'$. This completes the proof.

References

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

[2] S. Angenent, S. Haker, and A. Tannenbaum, *Minimizing flows for the Monge–Kantorovich problem*, SIAM Journal on Math. Analysis, 35 (2003), pp. 61–97.

[3] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numerische Mathematik, 84 (2000), pp. 375–393.

[4] A. Beurling, *An automorphism of product measures*, The Annals of Mathematics, 72 (1960), pp. 189–200.

[5] G. Birkhoff, *Extensions of Jentzsch's theorem*, Transactions of the American Mathematical Society, (1957), pp. 219–227.

[6] V. I. Bogachev, *Measure Theory, vol. 2*, Springer, 2007.

[7] P. J. Bushell, *Hilbert's metric and positive contraction mappings in a Banach space*, Archive for Rational Mechanics and Analysis, 52 (1973), pp. 330–338.

[8] L. Caffarelli, M. Feldman, and R. McCann, *Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs*, Journal of the American Mathematical Society, 15 (2002), pp. 1–26.

[9] Y. Chen, T. Georgiou, and M. Pavon, *Fast cooling for a system of stochastic oscillators*, arXiv preprint arXiv:1411.1323, (2014).

[10] ——, *On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint*, arXiv preprint arXiv:1412.4430, (2014).

[11] ——, *Optimal steering of a linear stochastic system to a final probability distribution, Part I*, arXiv preprint arXiv:1408.2222, *IEEE Trans. on Aut. Control* (to appear), (2014).

[12] ——, *Optimal steering of a linear stochastic system to a final probability distribution, Part II*, arXiv preprint arXiv:1410.3447, *IEEE Trans. on Aut. Control* (to appear), (2014).

[13] ——, *Optimal transport over a linear dynamical system*, arXiv preprint arXiv:1502.01265, *IEEE Trans. on Aut. Control* (under review), (2015).

[14] P. Dai Pra, *A stochastic control approach to reciprocal diffusion processes*, Applied mathematics and Optimization, 23 (1991), pp. 313–329.

[15] 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.

[16] R. Filliger and M.-O. Hongler, *Relative entropy and efficiency measure for diffusion-mediated transport processes*, Journal of Physics A: Mathematical and General, 38 (2005), p. 1247.

[17] 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.

[18] R. Fortet, *Résolution d'un système d'équations de M. Schrödinger*, J. Math. Pures Appl., 83 (1940).

[19] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, Acta Mathematica, 177 (1996), pp. 113–161.

[20] T. T. Georgiou and M. Pavon, *Positive contraction mappings for classical and quantum Schrödinger systems*, Journal of Mathematical Physics *56, 033301 (2015); doi: 10.1063/1.4915289; arXiv preprint arXiv:1405.6650, (2014).*

[21] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, Optimal mass transport for registration and warping, *International Journal of Computer Vision, 60 (2004), pp. 225–240.*

[22] D. Hilbert, Über die gerade linie als kürzeste verbindung zweier punkte, *Mathematische Annalen, 46 (1895), pp. 91–96.*

[23] B. Jamison, Reciprocal processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete, 30 (1974), pp. 65–86.*

[24] X. Jiang, Z.-Q. Luo, and T. T. Georgiou, Power spectral geodesics and tracking, *in Decision and Control, 2008. CDC 2008. 47th IEEE Conference on, IEEE, 2008, pp. 1315–1319.*

[25] R. Jordan, D. Kinderlehrer, and F. Otto, The variational formulation of the Fokker–Planck equation, *SIAM journal on Math. Analysis, 29 (1998), pp. 1–17.*

[26] B. Lemmens and R. Nussbaum, Nonlinear Perron-Frobenius Theory, *vol. 189, Cambridge University Press, 2012.*

[27] ——, Birkhoff's version of Hilbert's metric and its applications in analysis, in: Papadopoulos, A and Troyanov, M, eds. *Handbook of Hilbert Geometry*. IRMA Lectures in Math. and Theoretical Physics, European Math. Soc., pp. 275-303; preprint arXiv:1304.7921v1, (2013).

[28] C. Léonard, *From the Schrödinger problem to the Monge–Kantorovich problem*, Journal of Functional Analysis, 262 (2012), pp. 1879–1920.

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

[30] R. J. MCCANN, *A convexity principle for interacting gases*, Advances in Mathematics, 128 (1997), pp. 153–179.

[31] T. MIKAMI, *Monge's problem with a quadratic cost by the zero-noise limit of h-path processes*, Probability theory and related fields, 129 (2004), pp. 245–260.

[32] T. MIKAMI AND M. THIEULLEN, *Optimal transportation problem by stochastic optimal control*, SIAM Journal on Control and Optimization, 47 (2008), pp. 1127–1139.

[33] G. MONGE, *Mémoire sur la théorie des déblais et des remblais*, De l'Imprimerie Royale, 1781.

[34] E. NELSON, *Derivation of the Schrödinger equation from Newtonian mechanics*, Physical review, 150 (1966), p. 1079.

[35] 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.

[36] S. T. RACHEV AND L. RÜSCHENDORF, *Mass Transportation Problems: Volume I: Theory*, vol. 1, Springer, 1998.

[37] W. RUDIN, *Principles of Mathematical Analysis*, vol. 3, McGraw-Hill New York, 1964.

[38] E. SCHRÖDINGER, *Über die Umkehrung der Naturgesetze*, Sitzungsberichte der Preuss Akad. Wissen. Phys. Math. Klasse, Sonderausgabe, IX (1931), pp. 144–153.

[39] ———, *Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique*, in Annales de l'institut Henri Poincaré, vol. 2(4), Presses universitaires de France, 1932, pp. 269–310.

[40] C. VILLANI, *Topics in optimal transportation*, vol. 58, American Mathematical Soc., 2003.

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

[42] 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.

[43] L. ZHU, Y. YANG, S. HAKER, AND A. TANNENBAUM, *An image morphing technique based on optimal mass preserving mapping*, Image Processing, IEEE Transactions on, 16 (2007), pp. 1481–1495.