ENTROPIC INTERPOLATION AND GRADIENT FLOWS ON WASSERSTEIN PRODUCT SPACES

YONGXIN CHEN*, TRYPHON GEORGIOU† AND MICHELE PAVON‡

Abstract. We show that the entropic interpolation between two given marginals provided by the Schrödinger bridge may be characterized as the curve in the Wasserstein space $\mathcal{W}_2$ which minimizes a suitable action. We also study the relative entropy evolution between an uncontrolled and a controlled random evolution as a gradient flow on $\mathcal{W}_2 \times \mathcal{W}_2$.

Key words. Optimal mass transport, Wasserstein distance, gradient flow, Schrödinger bridge.

AMS subject classifications. 47H07, 47H09, 60J25, 34A34, 49J20

1. Introduction.

In the Schrödinger bridge problem (SBP) [13], one seeks the random evolution (a probability measure on path-space) which is closest in the relative entropy sense to a prior Markov diffusion evolution and has certain prescribed initial and final marginals μ and ν. As already observed by Schrödinger [37, 38], the problem may be reduced to a *static* problem which, except for the cost, resembles the Kantorovich relaxed formulation of the optimal mass transport problem (OMT). Considering that since [2] (OMT) also has a dynamic formulation, we have two problems which admit equivalent static and dynamic versions [24]. Moreover, in both cases, the solution entails a flow of one-time marginals joining μ and ν. The OMT yields a *displacement interpolation flow* whereas the SBP provides an *entropic interpolation flow*.

Trough the work of Mikami, Mikami-Thieullen and Leonard [26, 27, 28, 23, 24], we know that the OMT may be viewed as a "zero-noise limit" of SBP when the prior is a sort of uniform measure on path space with vanishing variance. This connection has been extended to more general prior evolutions in [8, 9]. Moreover, we know that, thanks to a very useful intuition by Otto [32], the displacement interpolation flow $\{\mu_t; 0 \leq t \leq 1\}$ may be viewed as a constant-speed geodesic joining μ and ν in Wasserstein space [40]. What can be said from this geometric viewpoint of the entropic flow? It cannot be a geodesic, but can it be characterized as a curve minimizing a suitable action? In this paper, we show that this is indeed the case resorting to a time-symmetric fluid dynamic formulation of SBP. This characterization of the Schrödinger bridge answers as a byproduct a question posed by Carlen [4, pp. 130-131].

*Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota MN 55455, USA; chen2468@umn.edu

†Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, Minnesota MN 55455, USA; tryphon@umn.edu

‡Dipartimento di Matematica, Università di Padova, via Trieste 63, 35121 Padova, Italy; pavon@math.unipd.it

§Research partially supported by the NSF, the AFOSR and by the University of Padova Research Project CPDA 140897.

SBP may be also formulated as a stochastic control problem with atypical boundary constraints. It is therefore interesting to compare the flow associated to the uncontrolled evolution (prior) to the optimal one. In particular, it is interesting to study the evolution of the relative entropy on the product Wasserstein space.

The paper is outlined as follows....

2. Elements of optimal mass transport theory. The literature on this problem is by now so vast and our degree of competence is such that we shall not even attempt here to give a reasonable and/or balanced introduction to the various fascinating aspects of this theory. Fortunately, there exist excellent monographs and survey papers on this topic, see [35, 11, 40, 1, 41, 33], to which we refer the reader. We shall only briefly review some concepts and results which are relevant for the topics of this paper.

2.1. The static problem. Let μ and ν be probability measures on the measurable spaces X and Y, respectively. Let $c : X \times Y \to [0, +\infty)$ be a measurable map with $c(x, y)$ representing the cost of transporting a unit of mass from location x to location y. Let $\mathcal{T}_{\mu\nu}$ be the family of measurable maps $T : X \to Y$ such that $T\#\mu = \nu$, namely such that ν is the *push-forward* of μ under T. Then Monge's optimal mass transport problem (OMT) is

$$\inf_{T\in\mathcal{T}_{\mu\nu}} \int_{X\times Y} c(x, T(x))d\mu(x). \tag{2.1}$$

As is well known, this problem may be unfeasible, namely the family $\mathcal{T}_{\mu\nu}$ may be empty. This is never the case for the "relaxed" version of the problem studied by Kantorovich in the 1940's

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} c(x, y)d\pi(x, y) \tag{2.2}$$

where $\Pi(\mu, \nu)$ are "couplings" of μ and ν, namely probability distributions on $X \times Y$ with marginals μ and ν. Indeed, $\Pi(\mu, \nu)$ always contains the product measure $\mu \otimes \nu$. Let us specialize the Monge-Kantorovich problem (2.2) to the case $X = Y = \mathbb{R}^N$ and $c(x, y) = |x - y|^2$. Then, if μ does not give mass to sets of dimension $\leq n - 1$, by Brenier's theorem [40, p.66], there exists a unique optimal transport plan π (Kantorovich) induced by a $d\mu$ a.e. unique map T (Monge), $T = \nabla\varphi$, φ convex, and we have

$$\pi = (I \times \nabla\varphi)\#\mu, \quad \nabla\varphi\#\mu = \nu. \tag{2.3}$$

Among the extensions of this result, we mention that to strictly convex, superlinear costs c by Gangbo and McCann [15]. The optimal transport problem may be used to introduce a useful distance between probability measures. Indeed, let $\mathcal{P}_2(\mathbb{R}^N)$ be the set of probability measures μ on $\mathbb{R}^N$ with finite second moment. For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^N)$, the Kantorovich-Rubinstein distance, usually called Wasserstein (Vasershtein) quadratic distance, is defined by

$$W_2(\mu, \nu) = \left(\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^N\times\mathbb{R}^N} |x - y|^2 d\pi(x, y)\right)^{1/2}. \tag{2.4}$$

As is well known [40, Theorem 7.3], W_2 is a *bona fide* distance. Moreover, it provides a most natural way to "metrize" weak convergence[1] in $\mathcal{P}_2(\mathbb{R}^N)$ [40, Theorem 7.12], [1, Proposition 7.1.5] (the same applies to the case $p \geq 1$ replacing 2 with p everywhere). The *Wasserstein space* $\mathcal{W}_2$ is defined as the metric space $\big(\mathcal{P}_2(\mathbb{R}^N), W_2\big)$. It is a *Polish space*, namely a separable, complete metric space.

2.2. The dynamic problem. So far, we have dealt with *the static* optimal transport problem. Nevertheless, in [2, p.378] it is observed that "...a continuum mechanics formulation was already implicitly contained in the original problem addressed by Monge... Eliminating the time variable was just a clever way of reducing the dimension of the problem". Thus, a *dynamic* version of the OMT problem was already *in fieri* in Gaspar Monge's 1781 *mémoire sur la théorie des déblais et des remblais*! It was elegantly accomplished by Benamou and Brenier in [2] by showing that

$$W_2^2(\mu,\nu) = \inf_{(\mu,v)} \int_0^1 \int_{\mathbb{R}^N} \|v(x,t)\|^2 \mu_t(dx)dt, \tag{2.5a}$$

$$\frac{\partial \mu}{\partial t} + \nabla \cdot (v\mu) = 0, \tag{2.5b}$$

$$\mu_0 = \mu, \quad \mu_1 = \nu. \tag{2.5c}$$

Here the flow $\{\mu_t; 0 \leq t \leq 1\}$ varies over continuous maps from $[0,1]$ to $\mathcal{P}_2(\mathbb{R}^N)$ and v over smooth fields. In [41], Villani states at the beginning of Chapter 7 that two main motivations for the time-dependent version of OMT are

- a time-dependent model gives a more complete description of the transport;
- the richer mathematical structure will be useful later on.

We can add three further reasons:

- it opens the way to establish a connection with the *Schrödinger bridge* problem (see Section 9 below), where the latter appears as a regularization of the former [26, 27, 28, 23, 24, 8, 9];
- it allows to view the optimal transport problem as an (atypical) *optimal control* problem [6]-[9].
- In some applications such as interpolation of images [10] or spectral morphing [21], the interpolating flow is essential!

Let $\{\mu_t^*; 0 \leq t \leq 1\}$ and $\{v^*(x,t); (x,t) \in \mathbb{R}^N \times [0,1]\}$ be optimal for (2.5). Then

$$\mu_t^* = [(1-t)I + t\nabla\varphi] \,\#\mu,$$

with $T = \nabla\varphi$ solving Monge's problem, provides, in McCann's language, the *displacement interpolation* between μ and ν. Then $\{\mu_t^*; 0 \leq t \leq 1\}$ may be viwed as a constant-speed geodesic joining μ and ν in Wasserstein space (Otto). This formally endows $\mathcal{W}_2$ with a "pseudo" Riemannian structure. McCann discovered [25] that certain functionals are *displacement convex*, namely convex along Wasserstein geodesics.

[1] μ_k converges weakly to μ if $\int_{\mathbb{R}^N} f d\mu_k \to \int_{\mathbb{R}^N} f d\mu$ for every continuous, bounded function f.

This has led to a variety of applications. Following one of Otto's main discoveries [22, 32], it turns out that a large class of PDE's may be viewed as *gradient flows* on the Wasserstein space $\mathcal{W}_2$. This interpretation, because of the displacement convexity of the functionals, is well suited to establish uniqueness and to study energy dissipation and convergence to equilibrium. A rigorous setting in which to make sense of the Otto calculus has been developed by Ambrosio, Gigli and Savaré [1] for a suitable class of functionals. Convexity along geodesics in $\mathcal{W}_2$ also leads to new proofs of various geometric and functional inequalities [25], [40, Chapter 9]. Finally, we mention that, when the space is not flat, qualitative properties of optimal transport can be quantified in terms of how bounds on the Ricci-Curbastro curvature affect the displacement convexity of certain specific functionals [41, Part II].

The *tangent space* of $\mathcal{P}_2(\mathbb{R}^N)$ at a probability measure μ, denoted by $T_\mu\mathcal{P}_2(\mathbb{R}^N)$ [1] may be identified with the closure in L^2_μ of the span of $\{\nabla\varphi : \varphi \in C_c^\infty\}$, where C_c^∞ is the family of smooth functions with compact support. It is naturally equipped with the scalar product of L^2_μ.

3. The Fokker-Planck equation as a gradient flow on Wasserstein space. Let us review the variational formulation of the Fokker-Planck equation as a gradient flow on Wasserstein space [22, 40]. Consider a physical system with *phase space* $\mathbb{R}^N$ and with *Hamiltonian* $\mathcal{H} : x \mapsto H(x) = E_x$. The thermodynamic states of the system are given by the family $\mathcal{P}(\mathbb{R}^N)$ of probability distributions P on $\mathbb{R}^N$ admitting density ρ. On $\mathcal{P}(\mathbb{R}^N)$, we define the *internal energy* as the expected value of the Energy *observable* in state P

$$U(H,\rho) = \mathbb{E}_P\{\mathcal{H}\} = \int_{\mathbb{R}^N} H(x)\,\rho(x)dx = \langle H, \rho\rangle. \tag{3.1}$$

Let us also introduce the (differential) *Gibbs entropy*

$$S(p) = -k\int_{\mathbb{R}^N} \ln\rho(x)\rho(x)dx, \tag{3.2}$$

where k is Boltzmann's constant. S is strictly concave on $\mathcal{P}(\mathbb{R}^N)$. According to the Gibbsian postulate of classical statistical mechanics, the equilibrium state of a microscopic system at constant absolute temperature T and with Hamiltonian function H is necessarily given by the Boltzmann distribution law with density

$$\bar\rho(x) = Z^{-1}\exp\left[-\frac{H(x)}{kT}\right] \tag{3.3}$$

where Z is the partition function[2]. Let us introduce the *Free Energy* functional F defined by

$$F(H,\rho,T) := U(H,\rho) - TS(\rho). \tag{3.4}$$

Since S is strictly concave on $\mathcal{S}$ and $U(E,\cdot)$ is linear, it follows that F is strictly convex on the state space $\mathcal{P}(\mathbb{R}^N)$. By Gibbs' variational principle, the Boltzmann

[2]The letter Z was chosen by Boltzmann to indicate "zuständige Summe" (pertinent sum- here integral).

distribution $\bar{\rho}$ is a minimum point of the free energy F on $\mathcal{P}(\mathbb{R}^N)$. Also notice that

$$\begin{aligned}\mathbb{D}(\rho\|\bar{\rho}) &= \int_{\mathbb{R}^N} \log\frac{\rho(x)}{\bar{\rho}(x)}\,\rho(x)dx\\ &= -\frac{1}{k}S(\rho) + \log Z + \frac{1}{kT}\int_{\mathbb{R}^N} H(x)\rho(x)dx = \frac{1}{kT}F(H,\rho,T) + \log Z.\end{aligned}$$

Since Z does not depend on ρ, we conclude that Gibb's principle is a trivial consequence of the fact that $\bar{\rho}$ minimizes $\mathbb{D}(\rho\|\bar{\rho})$ on $\mathcal{D}(\mathbb{R}^N)$.

Consider now an absolutely continuous curve $\mu_t : [t_0, t_1] \to \mathcal{W}_2$. Then [1, Chapter 8], there exist "velocity field" $v_t \in L^2_{\mu_t}$ such that the following continuity equation holds on $(0, T)$

$$\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0.$$

Suppose $d\mu_t = \rho_t dx$, so that the continuity equation

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(v\rho) = 0 \tag{3.5}$$

holds. We want to study the free energy functional $F(H, \rho_t, T)$ or, equivalently, $\mathbb{D}(\rho_t\|\bar{\rho})$, along the flow $\{\rho_t; t_0 \le t \le t_1\}$. Using (3.5), we get

$$\begin{aligned}\frac{d}{dt}\mathbb{D}(\rho_t\|\bar{\rho}) &= \int_{\mathbb{R}^N}\left[1 + \log\rho_t + \frac{1}{kT}H(x)\right]\frac{\partial\rho_t}{\partial t}dx\\ &= -\int_{\mathbb{R}^N}\left[1 + \log\rho_t + \frac{1}{kT}H(x)\right]\nabla\cdot(v\rho_t)dx.\end{aligned}\tag{3.6}$$

Integrating by parts, if the boundary terms at infinity vanish, we get

$$\frac{d}{dt}\mathbb{D}(\rho_t\|\bar{\rho}) = \int_{\mathbb{R}^N}\nabla\left[\log\rho_t + \frac{1}{kT}H(x)\right]\cdot v\rho_t dx = \langle\nabla\log\rho_t + \frac{1}{kT}\nabla H(x), v\rangle_{L^2_{\rho_t}}. \tag{3.7}$$

Thus, the Wasserstein gradient of $\mathbb{D}(\rho_t\|\bar{\rho})$ is

$$\nabla_{\mathcal{W}_2}\mathbb{D}(\rho_t\|\bar{\rho}) = \nabla\log\rho_t + \frac{1}{kT}\nabla H(x).$$

The corresponding gradient flow is

$$\frac{\partial\rho_t}{\partial t} = \nabla\cdot\left[\left(\nabla\log\rho_t + \frac{1}{kT}\nabla H(x)\right)\rho_t\right] = \nabla\cdot\left[\frac{1}{kT}\nabla H(x)\rho_t\right] + \Delta\rho_t. \tag{3.8}$$

But this is precisely the Fokker-Planck equation corresponding to the diffusion process

$$dX_t = -\frac{1}{kT}\nabla H(X_t)dt + \sqrt{2}dW_t \tag{3.9}$$

where W is a standard n-dimensional Wiener process. The process (3.9) has the Boltzmann distribution (3.3) as invariant density. Recall that [1, p.220] $F(H, \rho_t, T)$ or, equivalently, $\mathbb{D}(\rho_t\|\bar{\rho})$ are *displacement convex* and have therefore a unique minimizer.

Remark 1. It seems worthwhile investigating to what extent the fundamental assumption of statistical mechanics that the variables with longer relaxation time form a *vector Markov process* having (3.3) as invariant density is equivalent to the requirement that the flow of one-time densities be a gradient flow in Wasserstein space for the free energy.

Let us finally plug the "steepest descent" (3.8) into (3.6). We get, after integrating by parts, the well known formula [16]

$$\begin{aligned}\frac{d}{dt}\mathbb{D}(\rho_t\|\bar\rho) &= \int_{\mathbb{R}^N}\left[1+\log\rho_t+\frac{1}{kT}H(x)\right]\frac{\partial\rho_t}{\partial t}dx\\ &= \int_{\mathbb{R}^N}\left[1+\log\rho_t+\frac{1}{kT}H(x)\right]\nabla\cdot\left[\frac{1}{kT}\nabla H(x)\rho_t+\nabla\rho_t\right]dx\\ &= -\int_{\mathbb{R}^N}\|\nabla\log\left(\frac{\rho_t}{\bar\rho}\right)\|^2\rho_t dx. \qquad (3.10)\end{aligned}$$

The last integral in (3.10) is sometimes called the *relative Fisher information* of ρ_t with respect to $\bar\rho$ [40, p.278].

4. Relative entropy as a functional on Wasserstein product spaces. Consider now two absolutely continuous curves $\mu_t : [t_0,t_1]\rightarrow\mathcal{W}_2$ and $\tilde\mu_t : [t_0,t_1]\rightarrow\mathcal{W}_2$ and their velocity fields $v_t\in L^2_{\mu_t}$ and $\tilde v_t\in L^2_{\tilde\mu_t}$. Then, on $(0,T)$

$$\frac{d}{dt}\mu_t+\nabla\cdot(v_t\mu_t)=0, \qquad (4.1)$$

$$\frac{d}{dt}\tilde\mu_t+\nabla\cdot(\tilde v_t\tilde\mu_t)=0. \qquad (4.2)$$

Let us suppose that $d\mu_t=\rho_t(x)dx$ and $d\tilde\mu_t=\tilde\rho_t(x)dx$, for all $t\in[t_0,t_1]$. Then (4.1)-(4.2) become

$$\frac{\partial\rho}{\partial t}+\nabla\cdot(v\rho)=0, \qquad (4.3)$$

$$\frac{\partial\tilde\rho}{\partial t}+\nabla\cdot(\tilde v\tilde\rho)=0, \qquad (4.4)$$

where the fields v and $\tilde v$ satisfy

$$\int_{\mathbb{R}^N}\|v(x,t)\|^2\rho_t(x)dx<\infty,\quad \int_{\mathbb{R}^N}\|\tilde v(x,t)\|^2\tilde\rho_t(x)dx<\infty.$$

The differentiability of the Wasserstein distance $W_2(\tilde\rho_t,\rho_t)$ has been studied [41, Theorem 23.9]. Consider instead the relative entropy functional on $\mathcal{W}_2\times\mathcal{W}_2$

$$\mathbb{D}(\tilde\rho_t\|\rho_t)=\int_{\mathbb{R}^N}h(\tilde\rho_t,\rho_t)dx=\int_{\mathbb{R}^N}\log\left(\frac{\tilde\rho_t}{\rho_t}\right)\tilde\rho_t dx,\quad h(\tilde\rho,\rho)=\log\left(\frac{\tilde\rho}{\rho}\right)\tilde\rho.$$

Relative entropy functionals $\mathbb{D}(\cdot\|\gamma)$, where γ is a fixed probability measure (density), have been studied as geodesically convex functionals on $P_2(\mathbb{R}^N)$, see [1, Section 9.4]. Our study of the evolution of $\mathbb{D}(\tilde\rho_t\|\rho_t)$ is motivated by problems on a finite time interval such as the Schrödinger bridge problem (Section 6) and stochastic control problems (Section 8) where it is important to evaluate relative entropy on *two* flows of marginals.

We get

$$
\begin{aligned}
\frac{d}{dt}\mathbb{D}(\tilde{\rho}_t\|\rho_t) &= \int_{\mathbb{R}^N}\left[\frac{\partial h}{\partial\tilde{\rho}}\frac{\partial\tilde{\rho}}{\partial t}+\frac{\partial h}{\partial\rho}\frac{\partial\rho}{\partial t}\right]dx\\
&= \int_{\mathbb{R}^N}\left[(1+\log\tilde{\rho}_t-\log\rho_t)\left(-\nabla\cdot(\tilde{v}\tilde{\rho}_t\right)+\left(-\frac{\tilde{\rho}_t}{\rho_t}\right)(-\nabla\cdot(v\rho_t)\right]dx \qquad (4.5)
\end{aligned}
$$

After an integration by parts, assuming that the boundary terms at infinity vanish, we get

$$
\begin{aligned}
\frac{d}{dt}\mathbb{D}(\tilde{\rho}_t\|\rho_t) &= \int_{\mathbb{R}^N}\left[\nabla\log\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\cdot\tilde{v}\tilde{\rho}_t-\nabla\frac{\tilde{\rho}_t}{\rho_t}\cdot v\rho_t\right]dx\\
&= \int_{\mathbb{R}^N}\left[\begin{pmatrix}\nabla\log\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\\ -\nabla\frac{\tilde{\rho}_t}{\rho_t}\end{pmatrix}\cdot\begin{pmatrix}\tilde{v}\tilde{\rho}_t\\ v\rho_t\end{pmatrix}\right]dx. \qquad (4.6)
\end{aligned}
$$

Notice that the last expression looks like

$$
\left\langle\begin{pmatrix}\nabla\log\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\\ -\nabla\frac{\tilde{\rho}_t}{\rho_t}\end{pmatrix},\begin{pmatrix}\tilde{v}\\ v\end{pmatrix}\right\rangle_{L^2_{\tilde{\rho}_t}\times L^2_{\rho_t}}.
$$

Thus, we identify the gradient of the functional $\mathbb{D}(\tilde{\rho}\|\rho)$ on $\mathcal{W}_2\times\mathcal{W}_2$ as

$$
\begin{pmatrix}\nabla^1_{\mathcal{W}_2}\mathbb{D}(\tilde{\rho}\|\rho)\\ \nabla^2_{\mathcal{W}_2}\mathbb{D}(\tilde{\rho}\|\rho)\end{pmatrix}=\begin{pmatrix}\nabla\log\left(\frac{\tilde{\rho}}{\rho}\right)\\ -\nabla\frac{\tilde{\rho}}{\rho}\end{pmatrix}. \qquad (4.7)
$$

Let us now compute the gradient flow on $\mathcal{W}_2\times\mathcal{W}_2$ corresponding to gradient (4.7). We get

$$
\frac{\partial}{\partial t}\begin{pmatrix}\tilde{\rho}_t\\ \rho_t\end{pmatrix}-\nabla\cdot\begin{pmatrix}\nabla\log\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\tilde{\rho}_t\\ -\nabla\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\rho_t\end{pmatrix}=0. \qquad (4.8)
$$

Since

$$
J_1=\nabla\log\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\tilde{\rho}_t=\nabla\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\rho_t=-J_2,
$$

we observe the remarkable property that in the "steepest descent" (4.8) on the product Wasserstein space the "fluxes" are *opposite* and, therefore, $\frac{\partial\tilde{\rho}}{\partial t}=-\frac{\partial\rho}{\partial t}$. If we plug the steepest descent (4.8) into (4.5), we get what appears to be a new formula

$$
\begin{aligned}
\frac{d}{dt}\mathbb{D}(\tilde{\rho}_t\|\rho_t) &= \int_{\mathbb{R}^N}\left[\left(1+\log\tilde{\rho}_t-\log\rho_t+\frac{\tilde{\rho}_t}{\rho_t}\right)\frac{\partial\tilde{\rho}}{\partial t}\right]dx\\
&= -\int_{\mathbb{R}^N}\left[\|\nabla\log\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\|^2\tilde{\rho}_t+\|\nabla\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\|^2\rho_t\right]dx\\
&= -\int_{\mathbb{R}^N}\left[\left(1+\frac{\tilde{\rho}_t}{\rho_t}\right)\|\nabla\log\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\|^2\tilde{\rho}_t\right]dx, \qquad (4.9)
\end{aligned}
$$

which should be compared to (3.10).

Let us return to equation (4.6). By multiplying and dividing by $\tilde{\rho}_t$ in the last term of the middle expression, we get

$$\frac{d}{dt}\mathbb{D}(\tilde{\rho}_t\|\rho_t)=\int_{\mathbb{R}^N}\left[\nabla\log\left(\frac{\tilde{\rho}_t}{\rho_t}\right)\cdot(\tilde{v}-v)\right]\tilde{\rho}_t dx \tag{4.10}$$

which is precisely the expression obtained in [34, Theorem III.1].

5. Elements of Nelson-Föllmer kinematics of finite-energy diffusion processes. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a complete probability space. A stochastic process $\{\xi(t);t_0\le t\le t_1\}$ is called a *finite-energy diffusion* with constant diffusion coefficient $\sigma^2 I_N$ if the paths $\xi(\omega)$ belong to $C\left([t_0,t_1];\mathbb{R}^N\right)$ (N-dimensional continuous functions) and

$$\xi(t)-\xi(s)=\int_s^t\beta(\tau)d\tau+\sigma[W_+(t)-W_+(s)],\quad t_0\le s<t\le t_1, \tag{5.1}$$

where $\beta(t)$ is at each time t a measurable function of the past $\{\xi(\tau);t_0\le\tau\le t\}$ and W is a standard N-dimensional Wiener process. Moreover, the drift β satisfies the finite energy condition

$$\mathbb{E}\left\{\int_{t_0}^{t_1}\|\beta\|^2 d\tau\right\}<\infty.$$

In [12], Föllmer has shown that a finite-energy diffusion also admits a reverse-time Ito differential. Namely, there exists a measurable function $\gamma(t)$ of the future $\{\xi(\tau);t\le\tau\le t_1\}$ called *backward drift* and another Wiener process W_- such that

$$\xi(t)-\xi(s)=\int_s^t\gamma(\tau)d\tau+\sigma[W_-(t)-W_-(s)],\quad t_0\le s<t\le t_1. \tag{5.2}$$

Moreover, γ satisfies

$$\mathbb{E}\left\{\int_{t_0}^{t_1}\|\gamma\|^2 d\tau\right\}<\infty.$$

Let us agree that dt always indicate a strictly positive variable. For any function $f:[t_0,t_1]\to\mathbb{R}$ let $d_+f(t)=f(t+dt)-f(t)$ be the *forward increment* at time t and let $d_-f(t)=f(t)-f(t-dt)$ be the *backward increment* at time t. For a finite-energy diffusion, Föllmer has also shown in [12] that forward and backward drifts may be obtained as Nelson's conditional derivatives [29]

$$\beta(t)=\lim_{dt\searrow 0}\mathbb{E}\left\{\frac{d_+\xi(t)}{dt}|\xi(\tau),t_0\le\tau\le t\right\},$$

$$\gamma(t)=\lim_{dt\searrow 0}\mathbb{E}\left\{\frac{d_-\xi(t)}{dt}|\xi(\tau),t\le\tau\le t_1\right\},$$

the limits being taken in $L^2_N(\Omega,\mathcal{F},\mathbb{P})$. It was finally shown in [12] that the one-time probability density $\rho_t(\cdot)$ of $\xi(t)$ (which exists for every $t.t_0$) is absolutely continuous on $\mathbb{R}^N$ and the following duality relation holds $\forall t>0$

$$\mathbb{E}\left\{\beta(t)-\gamma(t)|\xi(t)\right\}=\sigma^2\nabla\log\rho(\xi(t),t),\quad \text{a.s.}. \tag{5.3}$$

Let us introduce the fields

$$b_+(x,t) = \mathbb{E}\{\beta(t)|\xi(t) = x\}, \quad b_-(x,t) = \quad \mathbb{E}\{\gamma(t)|\xi(t) = x\}.$$

Then, Ito's rule for the forwardd and backward differential of ξ imply that ρ_t satisfies the two Fokker-Planck equations

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(b_+\rho) - \frac{\sigma^2}{2}\Delta\rho = 0, \tag{5.4}$$

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(b_-\rho) + \frac{\sigma^2}{2}\Delta\rho = 0. \tag{5.5}$$

Following Nelson, let us introduce the *current* and *osmotic* drift of ξ by

$$v(t) = \frac{\beta(t)+\gamma(t)}{2}, \quad u(t) = \frac{\beta(t)-\gamma(t)}{2}, \tag{5.6}$$

respectively. Clearly v is similar to the classical velocity, whereas u is the velocity due to the noise which tends to zero when σ^2 tends to zero. Let us also introduce

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

Then, combining (5.4) and (5.5), we get

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(v\rho) = 0, \tag{5.7}$$

which has the form of a continuity equation expressing conservation of mass. When ξ is *Markovian* with $\beta(t) = b_+(\xi(t), t)$ and $\gamma(t) = b_-(\xi(t), t)$, (5.3) reduces to Nelson's relation

$$b_+(x,t) - b_-(x,t) = \sigma^2\nabla\log\rho_t(x). \tag{5.8}$$

Then (5.7) holds with

$$v(x,t) = b_+(x,t) - \frac{\sigma^2}{2}\nabla\log\rho_t(x). \tag{5.9}$$

6. Schrödinger bridges and entropic interpolation.

Let $\Omega = C([t_0,t_1];\mathbb{R}^N)$ be the space of $\mathbb{R}^N$ valued continuous functions. Let $W_x^{\sigma^2}$ denote Wiener measure on Ω with variance $\sigma^2 I_N$ starting at the point x at time t_0. If, instead of a Dirac measure concentrated at x, we give the volume measure as initial condition, we get the unbounded measure on path space $W^{\sigma^2} = \int_{\mathbb{R}^N} W_x^{\sigma^2} dx$. It is a useful tool to introduce the family of distributions $\mathcal{P}$ on Ω which are equivalent to it. Let $P \in \mathcal{P}$ represent an "a priori" random evolution and let $Q \in \mathcal{P}$. Then, we have [12]

$$\mathbb{D}(Q\|P) = \mathbb{E}_Q\left[\log\frac{dQ}{dP}\right] = \mathbb{D}(q_0\|p_0) + E_Q\left[\int_{t_0}^{t_1}\frac{1}{2\sigma^2}\|\beta^Q - \beta^P\|^2 dt\right] \tag{6.1a}$$

$$= \mathbb{D}(q_1\|p_1) + E_Q\left[\int_{t_0}^{t_1}\frac{1}{2\sigma^2}\|\gamma^Q - \gamma^P\|^2 dt\right]. \tag{6.1b}$$

Here q_0, q_1 are the marginal densities of Q at t_0 and t_1, respectively. Similarly, p_0, p_1 are the marginal densities of P. Let ρ_0 and ρ_1 be two probability densities. Let

$\mathcal{P}(\rho_0, \rho_1)$ denote the set of distributions in $\mathcal{P}$ having the prescribed marginal densities at t_0 and t_1. Given $P \in \mathcal{P}$, we consider the following problem:

$$\text{Minimize} \quad \mathbb{D}(Q\|P) \quad \text{over} \quad Q \in \mathcal{P}(\rho_0, \rho_1). \tag{6.2}$$

Conditions for existence and uniqueness for this problem and properties of the minimizing measure have been studied by many authors, most noticeably by Fortet, Beurlin, Jamison and Föllmer [14, 3, 20, 13]. If there is at least one Q in $\mathbb{D}(\rho_0\|\rho_1)$ such that $\mathbb{D}(Q\|P) < \infty$, there exists a unique minimizer Q^* in $\mathcal{P}(\rho_0\|\rho_1)$ called *the Schrödinger bridge* from ρ_0 to ρ_1 over P [13]. Existence is guaranteed by under conditions on P, ρ_0 and ρ_1, see [24, Proposition 2.5]. We shall tacitly assume henceforth that they are satisfied so that Q^* is well defined.

In view of Sanov's theorem [36], solving the maximum entropy problem (6.1) is equivalent to a problem of large deviations of the empirical distribution as showed by Föllmer [13] recovering Schrödinger's original motivation. Using (5.6) and observing that $\mathbb{D}(q_0\|p_0)$ and $\mathbb{D}(q_1\|p_1)$ are constant for Q varying in $\mathcal{P}(\rho_0\|\rho_1)$, we get that problem (6.2) is equivalent to

$$\text{Minimize} \quad \mathbb{E}_Q\left[\int_{t_0}^{t_1} \frac{1}{2\sigma^2}\|v^Q - v^P\|^2 + \frac{1}{2\sigma^2}\|u^Q - u^P\|^2 dt\right] \quad \text{over} \quad Q \in \mathcal{P}(\rho_0, \rho_1). \tag{6.3}$$

Suppose now that the prior measure P is *Markovian*. In this case, the classical results of Jamison [20] imply that the solution of (6.2) or, equivalently, of (6.3) is also Markovian. Thus, we can restrict our search to $\mathcal{P}^M(\rho_0, \rho_1)$, namely Markovian measures in $\mathcal{P}(\rho_0, \rho_1)$. Taking (5.8) into account, we can rewrite (6.3) in the form

$$\begin{aligned}\text{Minimize} \quad \mathbb{E}_Q\Bigg[\int_{t_0}^{t_1} &\frac{1}{2\sigma^2}\|v^Q(\xi(t), t) - v^P(\xi(t), t)\|^2 \\ &+\frac{1}{2\sigma^2}\|\frac{\sigma^2}{2}\nabla \log \frac{\rho_t^Q}{\rho_t^P}(\xi(t), t)\|^2 dt\Bigg] \quad \text{over} \quad Q \in \mathcal{P}^M(\rho_0\|\rho_1).\end{aligned} \tag{6.4}$$

Finally, in view of (5.7), we get the following equivalent fluid dynamic formulation

$$\begin{aligned}\text{Minimize}_{(\rho,v)} \int_{t_0}^{t_1}\int_{\mathbb{R}^N} &\left[\frac{1}{2\sigma^2}\|v(x,t) - v^P(x,t)\|^2\right. \\ &\left.+\frac{\sigma^2}{8}\|\nabla \log \frac{\rho}{\rho^P}(x,t)\|^2\right]\rho(x,t)dxdt,\end{aligned} \tag{6.5a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \tag{6.5b}$$

$$\rho_{t_0} = \rho_0, \quad \rho_{t_1} = \rho_1. \tag{6.5c}$$

Comparing (6.5) to the OMT with prior formulated and studied in [8], see also [9] for the Gauss-Markov case, we see that the essential difference is that there is here an extra term in the action functional which has the form of a *relative Fisher information* of ρ_t with respect to the prior one-time density ρ_t^P (Dirichlet form) [40, p.278] integrated over time. Also notice that, in view of (4.7), we have

$$\nabla \log \frac{\rho_t}{\rho_t^P} = \nabla^1_{\mathcal{W}_2}\mathbb{D}(\rho_t\|\rho_t^P). \tag{6.6}$$

To find the connection to the classical OMT, let us specialize to the situation where the prior $P = W^{\sigma^2}$. In that case, $v^P = u^P = 0$ and, multiplying the criterion by σ^2, we get the problem

$$\text{Minimize}_{(\rho,v)} \int_{t_0}^{t_1} \int_{\mathbb{R}^N} \left[\frac{1}{2}\|v(x,t)\|^2 + \frac{\sigma^4}{8}\|\nabla \log \rho_t(x)\|^2\right] \rho_t(x) dx dt, \tag{6.7a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \tag{6.7b}$$

$$\rho_{t_0} = \rho_0, \quad \rho_{t_1} = \rho_1. \tag{6.7c}$$

Again, specializing (6.6) to the case $\rho_t^P(x) \equiv 1$, we get

$$-\nabla \log \rho_t = \nabla_{\mathcal{W}_2} \frac{1}{k} S(\rho_t). \tag{6.8}$$

If $\sigma^2 \searrow 0$ it appears that in the limit we get the Benamou-Brenier formulation of OMT (2.5). This is indeed the case, see [26, 27, 28, 23, 24] and [8, 9] for the case with prior.

7. Optimal transport and Nelson's stochastic mechanics. There has been some interest in connecting optimal transport with Nelson's stochastic mechanics [4], [41, p.707] or directly with the Schrödinger equation [42]. Consider the case of a free, non-relativistic particle of mass m. Then, a variational principle leading to the Schrödinger equation, can be based on the Guerra-Morato action functional [15] which, in fluid dynamic form, is

$$\begin{aligned} \mathcal{A}_{GM}(t_0,t_1) &= \int_{t_0}^{t_1} \int_{\mathbb{R}^N} \frac{m}{2} \left[\|v(x,t)\|^2 - \|u(x,t)\|^2\right] \rho_t(x) dx dt \\ &= \int_{t_0}^{t_1} \left[\int_{\mathbb{R}^N} \frac{m}{2}\|v(x,t)\|^2 \rho_t(x) dx - \frac{\hbar^2}{8m} I(\rho_t)\right] dt \end{aligned} \tag{7.1}$$

where

$$I(\rho) = \int_{\mathbb{R}^N} \frac{\|\nabla \rho\|^2}{\rho} dx \tag{7.2}$$

is the *Fisher information* of ρ since, for the Nelson process, $\sigma^2 = \hbar/m$. Instead, the Yasue action [44] in fluid-dynamic form is

$$\begin{aligned} \mathcal{A}_{Y}(t_0,t_1) &= \int_{t_0}^{t_1} \int_{\mathbb{R}^N} \frac{m}{2} \left[\|v(x,t)\|^2 + \|u(x,t)\|^2\right] \rho_t(x) dx dt \\ &= \int_{t_0}^{t_1} \left[\int_{\mathbb{R}^N} \frac{m}{2}\|v(x,t)\|^2 \rho_t(x) dx + \frac{\hbar^2}{8m} I(\rho_t)\right] dt. \end{aligned} \tag{7.3}$$

In [4, p.131], Eric Carlen poses the question of minimizing the Yasue action subject to the continuity equation (6.7b) for given initial and final marginals (6.7c)) stating that "...the Euler-Lagrange equations for it are not easy to understand". In view of Section 6, we already know the solution to this problem: It is provided by the current velocity and the flow of one-time densities of the Schrödinger bridge with (6.7c) and stationary Wiener measure as a prior.

8. Relative entropy production for controlled evolution. Consider on $[t_0, t_1]$ a finite-energy Markov process taking values in $\mathbb{R}^N$ with forward Ito differential

$$d\xi = b_+(\xi(t), t)dt + \sigma dW_+. \tag{8.1}$$

Let $\rho_t(x)$ be the probability density of $\xi(t)$. Consider also the feedback controlled process ξ^u with forward differential

$$d\xi^u = b_+(\xi^u(t), t)dt + u(\xi^u(t), t)dt + \sigma dW_+. \tag{8.2}$$

Here the control u is adapted to the past and is such that ξ^u is a finite-energy diffusion. Let $\rho_t^u(x)$ be the probability density of $\xi^u(t)$. We are interested in the evolution of $\mathbb{D}(\rho_t^u \| \rho_t)$. By (5.9)-(5.7), the densities satisfy

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (v\rho) = 0, \quad v(x,t) = b_+(x,t) - \frac{\sigma^2}{2}\nabla \log \rho_t(x) \tag{8.3}$$

$$\frac{\partial \rho^u}{\partial t} + \nabla \cdot (v^u\rho^u) = 0, \quad v^u(x,t) = b_+(x,t) + u(x,t) - \frac{\sigma^2}{2}\nabla \log \rho_t^u(x). \tag{8.4}$$

By (4.10), we now get

$$\begin{aligned} \frac{d}{dt}\mathbb{D}(\rho_t^u \| \rho_t) &= \int_{\mathbb{R}^N} \left[\nabla \log \left(\frac{\rho_t^u}{\rho_t}\right) \cdot (v^u - v)\right] \rho_t^u dx \\ &= \int_{\mathbb{R}^N} \left[\nabla \log \left(\frac{\rho_t^u}{\rho_t}\right) \cdot \left(u - \frac{\sigma^2}{2}\nabla \log \left(\frac{\rho_t^u}{\rho_t}\right)\right)\right] \rho_t^u dx. \end{aligned} \tag{8.5}$$

Suppose now $\rho_t^u = \rho_t^0$ is also uncontrolled and differs from ρ_t only because of the initial condition at $t = t_0$. Then (8.5) gives the well known formula generalizing (3.10)

$$\frac{d}{dt}\mathbb{D}(\rho_t^0 \| \rho_t) = -\frac{\sigma^2}{2}\int_{\mathbb{R}^N} \left[\nabla \log \left(\frac{\rho_t^0}{\rho_t}\right) \cdot \nabla \log \left(\frac{\rho_t^0}{\rho_t}\right)\right] \rho_t^u dx \tag{8.6}$$

which shows that two solutions of the same Fokker-Plank equation tend to get closer.

9. Schrödinger bridges as controlled evolution. Consider now the situation where $\xi(t)$ represents a "prior" evolution on $[t_0, t_1]$ and the controlled evolution $\xi^{u^*} = \xi^*$ is the solution of the Schrödinger bridge problem for a pair of initial and final marginals ρ_0 and ρ_1 [13, 43]. Then the differential of ξ^* is given by

$$d\xi^* = b_+(\xi^*(t), t)dt + \sigma^2 \nabla \log \varphi(\xi^*(t), t)dt + \sigma dW_+ \tag{9.1}$$

where φ is space-time harmonic for the prior evolution, namely it satisfies

$$\frac{\partial \varphi}{\partial t} + b_+ \cdot \nabla \varphi + \frac{\sigma^2}{2}\Delta\varphi = 0. \tag{9.2}$$

Let ρ^φ be the density of ξ^*. Let us single out a special case of the Schrödinger bridge problem where relative entropy on path space is minimised under the only constraint that the final marginal density be $\rho_1 \neq \rho_{t_1}$. In such case, we have

$$\rho_t^\varphi(x) = \rho_t(x) \cdot \varphi(x,t).$$

Then (8.5) gives

$$\frac{d}{dt}\mathbb{D}(\rho_t^{\varphi}\|\rho_t) = \frac{\sigma^2}{2}\int_{\mathbb{R}^N} \left[\nabla \log \varphi \cdot \nabla \log \varphi\right] \rho_t^{\varphi} dx. \tag{9.3}$$

This shows that $\mathbb{D}(\rho_t^{\varphi}\|\rho_t)$ increases up to time $t = t_1$. It represents the intuitive fact that the bridge evolution has to be as close as possible to the prior but the final value of the relative entropy must be the positive quantity $\mathbb{D}(\rho_1\|\rho_{t_1})$. Thus, $\mathbb{D}(\rho_t^{\varphi}\|\rho_t)$ approaches this positive quantity from below. Result (9.3) may be viewed as a reverse-time H-theorem, as the bridge and the reference evolution have the same backward drift [13].

REFERENCES

[1] L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2nd ed. 2008.

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

[3] A. Beurling, An automorphism of product measures, *Ann. Math.* **72** (1960), 189-200.

[4] E. Carlen, Stochastic mechanics: A look back and a look ahead, in *Diffusion, quantum theory and radically elementary mathematics*, W. G. Faris ed., Mathematical Notes 47, Princeton University Press, 2006, 117-139.

[5] Y. Chen and T.T. Georgiou, Stochastic bridges of linear systems, preprint, `http://arxiv.org/abs/1407.3421`, *IEEE Trans. Aut. Control*, to appear.

[6] Y. Chen, T.T. Georgiou and M. Pavon, Optimal steering of a linear stochastic system to a final probability distribution, , Part I, Aug. 2014, `http://arxiv.org/abs/1408.2222`, *IEEE Trans. Aut. Control*, to appear.

[7] Y. Chen, T.T. Georgiou and M. Pavon, Optimal steering of a linear stochastic system to a final probability distribution, part II, Oct. 2014, `http://arxiv.org/abs/1410.3447`, *IEEE Trans. Aut. Control*, to appear.

[8] Y. Chen, T.T. Georgiou and M. Pavon, "On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint" Dec. 2014, arXiv :1412.4430v1, submitted for publication.

[9] Y. Chen, T.T. Georgiou and M. Pavon, "Optimal transport over a linear dynamical system", Feb. 2015, arXiv:1502.01265v1, submitted for publication.

[10] Y. Chen, T.T. Georgiou and M. Pavon, Entropic and displacement interpolation: a computational approach using the Hilbert metric, June 2015, arXiv:1506.04255v1, submitted for publication.

[11] L. C. Evans, Partial differential equations and Monge-Kantorovich mass transfer, in *Current developments in mathematics*, 1977, Int. Press, Boston, 1999, 65-126.

[12] H. Föllmer, in *Stochastic Processes - Mathematics and Physics* , Lecture Notes in Mathematics (Springer-Verlag, New York,1986), Vol. 1158, pp. 119-129.

[13] H. Föllmer, Random fields and diffusion processes, in: *Ècole d'Ètè de Probabilitès de Saint-Flour XV-XVII*, edited by P. L. Hennequin, Lecture Notes in Mathematics, Springer-Verlag, New York, 1988, vol.1362,102-203.

[14] R. Fortet, Résolution d'un système d'equations de M. Schrödinger, *J. Math. Pure Appl.* IX (1940), 83-105.

[15] W. Gangbo and R. McCann, The geometry of optimal transportation. *Acta Mathematica*, **177** (1996), no. 2, 113?161.

[16] R. Graham, Path integral methods in nonequilibrium thermodynamics and statistics, in *Stochastic Processes in Nonequilibrium Systems*, L. Garrido, P. Seglar and P.J.Shepherd Eds., Lecture Notes in Physics 84, Springer-Verlag, New York, 1978, 82-138.

[17] F. Guerra and L. Morato, Quantization of dynamical systems and stochastic control theory, *Physical Review D*, **27** (8), 1774-1786.

[18] U.G.Haussmann and E.Pardoux, Time reversal of diffusions, *The Annals of Probability* **14**, 1986, 1188.

[19] D.B.Hernandez and M. Pavon, Equilibrium description of a particle system in a heat bath, *Acta Applicandae Mathematicae* **14** (1989), 239-256.

[20] B. Jamison, The Markov processes of Schrödinger, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **32** (1975), 323-331.

[21] X. Jiang, Z. Luo, and T. Georgiou, Geometric methods for spectral analysis, *IEEE Transactions on Signal Processing*, **60**, no. 3, 2012, 10641074.

[22] R. Jordan, D. Kinderlehrer, and F. Otto, The variational formulation of the Fokker-Planck equation, *SIAM J. Math. Anal.* **29**: 1-17 (1998).

[23] C. Léonard, A survey of the Schroedinger problem and some of its connections with optimal transport, *Discrete Contin. Dyn. Syst. A*, 2014, **34** (4): 1533-1574.

[24] C. Léonard, From the Schrödinger problem to the Monge-Kantorovich problem, *J. Funct. Anal.*, 2012, **262**, 1879-1920.

[25] R. McCann, A convexity principle for interacting gases, *Adv. Math.*, **128** (1997), 1, 153-179.

[26] T. Mikami, Monge's problem with a quadratic cost by the zero-noise limit of h-path processes, *Probab. Theory Relat. Fields*, **129**, (2004), 245-260.

[27] T. Mikami and M. Thieullen, Duality theorem for the stochastic optimal control problem., *Stoch. Proc. Appl.*, 116, 1815?1835 (2006).

[28] T. Mikami and M. Thieullen, Optimal Transportation Problem by Stochastic Optimal Control, *SIAM Journal of Control and Optimization*, **47**, N. 3, 1127-1139 (2008).

[29] E. Nelson, *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton, 1967.

[30] L. Ning, X. Jiang and T. Georgiou, On the geometry of covariance matrices, *IEEE Signal Processing Letters*, vol. 20, August 2013, 787- 790.

[31] L. Ning, T.T. Georgiou, and A. Tannenbaum, Matrix-valued Monge-Kantorovich Optimal Mass Transport, *IEEE Trans. Aut. Contr.*, to appear.

[32] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, *Comm. Partial Differential Equations* **26**, 1?2 (2001), 101?174.

[33] Y. Ollivier, H. Pajot and C. Villani, *Optimal Transportation. Theory and Applications*, London Mathematical Society Lecture Notes Series 413, Cambridge University Press, 2014.

[34] M. Pavon and F. Ticozzi, On entropy production for controlled Markovian evolution, *J. Math. Phys.*, **47**, 06330, doi:10.1063/1.2207716 (2006).

[35] T. Rachev and L. Rüschendorf, *Mass Transportation Problems, Vol. I: Theory, Vol. II: Applications*. Probability and its applications. Springer- Verlag, New York, 1998.

[36] I. S. Sanov, On the probability of large deviations of random magnitudes (in Russian), *Mat. Sb. N. S.*, **42** (84) (1957) 1144. Select. Transl. Math. Statist. Probab., 1, 213-244 (1961).

[37] E. Schrödinger, Über die Umkehrung der Naturgesetze, *Sitzungsberichte der Preuss Akad. Wissen.* Berlin, Phys. Math. Klasse (1931), 144-153.

[38] E. Schrödinger, Sur la théorie relativiste de l'electron et l'interpretation de la mécanique quantique, *Ann. Inst. H. Poincaré* 2, 269 (1932).

[39] E. Tannenbaum, T. Georgiou and A. Tannenbaum, Signals and Control Aspects of Optimal Mass Transport and the Boltzmann Entropy, *Proc. 49th IEEE-CDC conference*, Atlanta, GA, 2010, 1885-1890.

[40] C. Villani, *Topics in optimal transportation*, AMS, 2003, vol. 58.

[41] C. Villani, *Optimal transport. Old and new.* Grundlehren der Mathematischen Wissenschaften 338. Springer-Verlag, 2009.

[42] M. K. von Renesse, An optimal transport view on Schrödinger's equation, *Canad. Math. Bull*, **55** (4), 2012, 858-869.

[43] A. Wakolbinger, Schroedinger bridges from 1931 to 1991, in *Proc. of the 4th Latin American Congress in Probability and Mathematical Statistics*, Mexico City 1990, *Contribuciones en probabilidad y estadistica matematica*, **3** (1992), 61-79.

[44] K. Yasue, Stochastic calculus of variations, *J. Funct. Analysis*, **41**, 327-340 (1981).