Università degli Studi di Padova

Padua Research Archive - Institutional Repository

Extremal flows in Wasserstein space

(Article begins on next page)

Extremal flows on Wasserstein space

Giovanni Conforti[1] and Michele Pavon[2]

[1]*Centre de Mathématiques Appliquées,
École Polytechnique, Palaiseau, France*[∗]

[2]*Dipartimento di Matematica "Tullio Levi-Civita",
Università di Padova, via Trieste 63, 35121 Padova, Italy*[†]

Abstract. We develop an intrinsic geometric approach to calculus of variations on Wasserstein space. We show that the flows associated to the Schrödinger bridge with general prior, to Optimal Mass Transport and to the Madelung fluid can all be characterized as annihilating the first variation of a suitable action. We then discuss the implications of this unified framework for stochastic mechanics: It entails, in particular, a sort of fluid-dynamic reconciliation between Bohm's and Nelson's stochastic mechanics.

I. INTRODUCTION

Interpolation between two given probability distributions has important applications, for instance in image morphing. The interpolating flow of one-time marginals $\{\mu_t; 0 \leq t \leq 1\}$ may be thought of as a curve in *Wasserstein space* [79][1]. One example is the displacement interpolation flow associated to the Benamou-Brenier formulation of the *Optimal Mass Transport* (OMT) problem with quadratic cost [3]. Other examples are the flows associated to the *Schrödinger Bridge Problem* (SBP) [75, 76] and the quantum evolution of a nonrelativistic particle in *Nelson's Stochastic Mechanics* (NSM) [60, 61]. It is known that SBP may be viewed as a "regularization" of OMT , the latter problem being recovered through a "zero-noise limit" [16, 17, 46, 47, 55–57]. Recently, it was shown by von Renesse [81], for the quantum mechanical *Madelung fluid* [36], and by one of us [20] for the Schrödinger bridge, that their flows satisfy a suitable Newton-like law in Wasserstein space. In [32], the foundations of a Hamilton-Jacobi theory in Wasserstein space were laid.

In this paper, we show that the solution flows of OMT, SBP and NSM may all be seen as extremal curves in Wasserstein space of a suitable action. The actions only differ by the presence or the sign of a (relative) *Fisher information functional* besides the kinetic energy term. The solution marginals flows correspond to critical points, i.e. annihilate the first variation, of the respective functionals. The extremality condition implies indeed a generalization of the known Newton-type second laws in Wasserstein space. With respect to [20], we present an alternative approach. There, the Newton's law for SBP is proved using its dual representation as a generalized *h*-transform of the equilibrium dynamics. Here, we use a more geometrical approach which consists in taking advantage of the Riemannian metric associated with the Wasserstein distance to implement the fundamental lemma of

[1] Vershik has argued [78] that it should be called *Kantorovich space*. According to him, it was Dobrushin, unaware of Kantorovich's work, that spread this terminology.

calculus of variations on the fluid dynamic formulation of SBP. This approach has the nice feature of being more intrinsic, providing a unified framework for the results of [81] and [20] and allows to prove some new ones. For example, we are able to treat non-reversible prior laws for SBP and show that the non reversibility translates in the addition of a viscous term in Newton's law. The calculations of this article, however, are more or less formal, and more work is required to turn them into rigorous results. Let us note that for SBP rigorous results can be found in [20]. Our results also shed light on the relation between various versions of stochastic mechanics (Schrödinger, Bohm, Nelson), a controversial topic made more misterious by the presence of two different Lagrangians within Nelson's stochastic mechanics [61].

Some of the results presented here had been announced, without developing the geometric approach, in our conference article [21]. The paper is outlined as follows. In Section II, we recall some some basic concepts and result from OMT. Section III is devoted to Schrödinger bridges and entropic interpolation. In Section IV, we introduce the basic geometric tools in Wasserstein space and provide an intrinsic variational representation for the entropic interpolating flow (Theorem 1). In Section V, we show that the entropic interpolation satisfies a Newton-type law involving the covariant derivative of the velocity generalizing to the case of a general non-reversible prior the results of [20]. Section VI is devoted to the classical Madelung fluid and to stochastic mechanics. We first connect to the results of [81] on the Madelung fluid. We then consider Nelson's stochastic mechanics and compare it to Bohmian mechanics. We finally briefly discuss the differences in stochastic mechanics between the fluid-dynamic variational principles and the stochastic variational principles.

II. BACKGROUND ON OPTIMAL MASS TRANSPORT

The literature on OMT is huge and keeps growing at a fast pace. A selection of monographs and survey papers is [1, 2, 26, 64, 71, 74, 79, 80]. Let $\Pi(\mu_0, \mu_1)$ be "couplings" of μ_0 and μ_1, namely probability distributions on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ_0 and μ_1. In the Kantorovich relaxed formulation of the original Monge problem, one seeks a distribution $\pi \in \Pi(\mu_0, \mu_1)$ which minimizes the index

$$\int_{\mathbb{R}^N \times \mathbb{R}^N} c(x, y) d\pi(x, y),$$

where c represents the cost of transporting an infinitesimal "pebble" from x to y. When $c(x, y) = \|x - y\|^2$ and μ_0, μ_1 belong to $\mathcal{P}_2(\mathbb{R}^d)$, the set of probability measures μ on $\mathbb{R}^d$ with finite second moment, the Kantorovich-Rubinstein (Wasserstein) quadratic distance, is defined by

$$W_2(\mu_0, \mu_1) = \left(\inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 d\pi(x, y) \right)^{1/2}. \tag{1}$$

As is well known [79, Theorem 7.3], W_2 is a *bona fide* distance. Moreover, it provides a most natural way to "metrize" weak convergence[2] in $\mathcal{P}_2(\mathbb{R}^N)$ [79, Theorem 7.12], [1, Proposition 7.1.5]. The quadratic *Wasserstein space* $\mathcal{W}_2$ is defined as the metric space $\left(\mathcal{P}_2(\mathbb{R}^N), W_2 \right)$. It is a *Polish space*, namely a separable, complete metric space.

A *dynamic* version of the above 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

---

[2] 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.

Benamou and Brenier in [3] by showing that

$$W_2^2(\mu_0, \mu_1) = \inf_{(\rho,v)} \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 \rho_t(dx) dt, \tag{2a}$$

$$\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0, \tag{2b}$$

$$\rho_0 = \mu_0, \quad \rho_1 = \mu_1. \tag{2c}$$

Here the flow $\{\rho_t; 0 \le t \le 1\}$ varies over continuous maps from $[0,1]$ to $\mathcal{P}_2(\mathbb{R}^d)$ and v over vector fields. In [80, Chapter 7], Villani provides motivation to study the time-dependent version of OMT, namely that a time-dependent model gives a more complete description of the transport and that the richer mathematical structure is useful. Further reasons are the following. It allows to view the optimal transport problem as an (atypical) optimal control problem [13]-[17]. It provides a ground on which the connection with the Schrödinger bridge problem appears as a regularization of the former [16, 17, 46, 47, 55–57]. Similarly with Nelson's stochastic mechanics, see below. In some applications, such as interpolation of images [18] or spectral morphing [41], the interpolating flow is the object of interest.

Let $\{\mu_t; 0 \le t \le 1\}$ be optimal for (2). Then (μ_t) provides the *displacement interpolation* between μ_0 and μ_1 and may be viewed as a constant-speed geodesic joining μ_0 and μ_1 in Wasserstein space (Otto).

III. SCHRÖDINGER BRIDGES AND ENTROPIC INTERPOLATION

Let $\Omega = C([0,1]; \mathbb{R}^d)$ be the space of $\mathbb{R}^d$ valued continuous functions, P a positive measure[3] on Ω. We recall the definition of the relative entropy functional

$$\mathbb{D}(Q\|P) = \begin{cases} \mathbb{E}_Q\left(\log \frac{dQ}{dP}\right), & \text{if } Q \ll P \\ +\infty & \text{otherwise} \end{cases}$$

For given P, μ_0, μ_1, the *Schrödinger Problem* is the problem of minimizing the relative entropy with respect to P in the set $\mathcal{P}(\mu_0, \mu_1)$ of all laws whose marginals at times $t = 0, 1$ are μ_0 and μ_1, namely

$$\text{Minimize} \quad \mathbb{D}(Q\|P) \quad \text{over} \quad Q \in \mathcal{P}(\mu_0, \mu_1). \tag{3}$$

The optimal solution is called the *Schrödinger bridge* between μ_0 and μ_1 over P, and its marginal flow (μ_t) is the *entropic interpolation*. There are at least three distinct situations when this problem arises naturally:

(a) As a model for the "hot gas experiment": This was Schrödinger's original motivation.

(b) As a statistical inference problem, in the entropic approach to Bayesian statistics developed by Jaynes, see e.g. [39, 40].

(c) As a regularization of the Monge Kantorovich problem.

---

[3] It is sometimes convenient to work with infinite measures P. This is the case, for instance, when P is *stationary Wiener measure* (reversible Brownian motion) on $\mathbb{R}^d$

a.   hot gas experiment   Here, the prior distribution P is a stationary Langevin dynamics for a generator $\mathcal{L}$ of the form

$$\mathcal{L} = \frac{\sigma}{2}\Delta - \nabla U \cdot \nabla,$$

which represents an *equilibrium dynamics* for a particle system. At time $t = 0$ we are given N independent particles $(X_t^i)_{t \leq 1, i \leq N}$ whose configuration is μ.

$$\mu := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_0^i}.$$

We then let the particles travel independently following the Langevin dynamics and look at their configuration ν at $t = 1$.

$$\nu := \frac{1}{N}\sum_{i=1}^{N} \delta_{X_1^i}.$$

If the Langevin dynamics has good ergodic properties and N is very large, one expects ν to be very close to the invariant measure of P, which we denote $\mathbf{m}$. However, although very unlikely, it is still possible to observe an *unexpected configuration*, meaning that ν is significantly[4] different from $\mathbf{m}$. Schrödinger's question is to find, conditionally on this rare event, the most likely evolution of the particle system. In the limit when $N \to +\infty$, Sanov's theorem tells that the *most likely* evolution is given by the minimiser of (3) (see also [46, sec.6]). This derivation of the Scrödinger problem can be viewed as a stochastic counterpart to the "lazy gas experiment" of optimal transport [80, Ch. 16]. Indeed

- Particles choose their final destination minimizing the relative entropy instead of the mean square distance;

- Particles travel along Brownian bridges instead of geodesics.

We refer to [46, sec.6] for an extensive treatment of this analogy. Quite remarkably, the description of the hot gas experiment we gave is very similar to the original formulation of the problem that Schrödinger proposed back in 1932, (see [76]) when none of the modern tools of probability or of modern optimal transport theory was available.

b.   Maximum entropy principle and statistical inference   Suppose now that P is a prior distribution on some hypothesis space. When measurements are made and new information becomes available, Jaynes' maximum entropy principle asserts that the prior distribution should be updated by picking, among all laws compatible with the information, the one which minimizes the relative entropy w.r.t. to P^5. Loosely speaking, it is a "maximum ignorance principle", saying that the posterior distribution should make the least possible prediction about everything which is beyond the available information. When the hypothesis space is Ω, and the new information comes in the form of marginal laws, the problem of finding the posterior becomes SBP.

---

[4] Schrödinger writes in [76] *" un écart spontané et considerable par rapport à cette uniformité"*

[5] The confusion between minimization and maximization comes from the different definitions of entropy.

c. Regularization of Optimal transport Suppose now that the prior measure P is *Markovian.* In this case, the classical results of Jamison [38] imply that the solution of (3) is also Markovian. Thus, we can restrict our search to $\mathcal{P}^M(\mu_0, \mu_1)$, namely Markovian measures in $\mathcal{P}(\mu_0, \mu_1)$. Along the lines of [16, p.683], [33, Corollary 5.8], [44] we get that the entropic flow of the Schrödinger bridge can be characterized through the following fluid dynamic problem

$$\inf_{(\rho,v)} \int_{t_0}^{t_1} \int_{\mathbb{R}^d} \left[\frac{1}{2\sigma} |v_t(x) - v_t^P(x)|^2 \right. \tag{4a}$$

$$\left. + \frac{\sigma}{8} |\nabla \log \frac{\rho}{\rho^P}_t (x)|^2 \right] \rho_t(x) dx dt,$$

$$\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0, \tag{4b}$$

$$\rho_0 = \nu_0, \quad \rho_1 = \nu_1. \tag{4c}$$

Here v and v^P are the *current velocity fields* [60] of the trial and prior measure, respectively. Comparing (4) to the OMT with prior formulated and studied in [16], see also [17] 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.* To find the connection to the classical OMT, let us specialize to the situation where the prior $P = W^\varepsilon$. In that case, $v^P = u^P = 0$ and, multiplying the criterion by ε, we get the problem

$$\inf_{(\rho,v)} \int_{t_0}^{t_1} \int_{\mathbb{R}^d} \left[\frac{1}{2} |v_t(x)|^2 + \frac{\varepsilon^2}{8} |\nabla \log \rho_t(x)|^2 \right] \rho_t(x) dx dt, \tag{5a}$$

$$\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0, \tag{5b}$$

$$\mu_0 = \nu_0, \quad \mu_1 = \nu_1. \tag{5c}$$

If $\varepsilon \searrow 0$ it appears that in the limit we get the Benamou-Brenier formulation of OMT (2). This is indeed the case, see [46, 47, 55–57] and [16, 17] for the case with prior.

Alternatively, since both OMT and SBP admit a static formulation [46], the *regularization* of OMT as in (1) can be obtained by adding a term proportional to the Shannon entropy on $\mathbb{R}^d \times \mathbb{R}^d$

$$\inf_{\pi \in \Pi(\mu_0,\mu_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\pi(x,y) + 2\varepsilon \int_{\mathbb{R}^d \times \mathbb{R}^d} \pi(x,y) \log \pi(x,y) dx dy, \tag{6}$$

where we used the same notation for a measure and its density function. Problem (6) is the projection onto the marginals at $t = 0, 1$ of the Schrödinger problem for a prior P which is stationary Wiener measure, i.e. the stationary law for the Markov generator

$$\mathcal{L}^\varepsilon = \frac{\varepsilon}{2} \Delta.$$

It is well known that the solution to the path-space formulation of the Schrödinger problem (3) can be recovered by mixing the optimal coupling in (6) with Brownian bridges [30], [46, Sec.1]. For the computational relevance of (6) see ([22]). For the dynamic form of entropic regularization of OMT (5), discovered in [16][6], see instead [53]. Notice that it differs from the

---

[6] Incidentally, this characterization of the Schrödinger bridge flow answers at once a question posed by Carlen [11, pp. 130-131].

regularization of the Benamou-Brenier formulation which involves replacing the continuity equation with a Fokker-Planck equation [18].

Conditions for existence and uniqueness for SBP and properties of the minimizing measure have been studied by many authors, most noticeably by Fortet, Beurlin, Jamison and Föllmer [4, 30, 31, 38]. If there is at least one Q in $\mathcal{P}(\mu_0, \mu_1)$ such that $\mathbb{D}(Q\|P) < \infty$, there exists a unique minimizer Q^*. Existence is guaranteed under conditions on P, ρ_0 and ρ_1, see [47, Proposition 2.5]. We shall tacitly assume henceforth that they are satisfied so that Q^* is well defined. It has been observed since the early nineties that SBP can be turned, thanks to *Girsanov's theorem*, into a stochastic control problem with atypical boundary constraints, see [9, 23, 24, 28, 70].

IV.   SECOND ORDER CALCULUS IN WASSERSTEIN SPACE

It is known that the *displacement interpolation* between two measures is a constant speed geodesic in the quadratic Wasserstein space $\mathcal{W}_2$. In a seminal paper [63], Otto showed that we can go beyond this. More precisely, it is possible to define formally a kind of Riemannian metric on $\mathcal{P}_2(\mathbb{R}^d)$ for which the displacement interpolation is a constant speed geodesic. We sketch here the main steps in the construction of this metric. All what follows is only formally correct, and much work is required to turn it into rigorous statements. For this, we refer the reader to [2], [34]. The first step consists in identifying the tangent space at μ with the space of square integrable vector fields.

$$\mathbf{T}_\mu := \overline{\{\nabla\varphi; \varphi \in C_c^\infty\}}^{L^2(\mu)}$$

The second step is to define the first derivative (velocity field) $v_t \in \mathbf{T}_{\mu_t}$ of a curve (μ_t) through the continuity equation

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0, \quad v_t \in \mathbf{T}_{\mu_t}.$$

Then, one defines the Riemannian metric by means of the L^2 product

$$\langle \nabla\varphi, \nabla\psi \rangle_{\mathbf{T}_\mu} := \int_{\mathbb{R}^d} \langle \nabla\varphi(x), \nabla\psi(x) \rangle \, \mu(x)\mathrm{d}x,$$

where $\langle ., . \rangle$ stands for the standard inner product of $\mathbb{R}^d$. The Benamou-Brenier formula establishes that the displacement interpolation is a constant speed geodesic for this Riemannian structure, as it minimizes the energy functional among all curves with a given start and end. The next task is to define the covariant derivative $\nabla_{v_t}^{W_2} u_t$ of the vector field (u_t) along the curve (μ_t) in such a way that the torsion free identity and the compatibility with the metric are satisfied. It turns out that to do so, the right definition is to set (cf. [2, eq 6.7, Def 6.8])

$$\nabla_{v_t}^{W_2} u_t := \mathrm{P}_{\mu_t} \left(\partial_t u_t + \mathrm{D}_{v_t} u_t \right), \tag{7}$$

provided u_t, v_t are smooth enough. Here, $\mathrm{D}_{v_t} u_t(x)$ denotes the Jacobian of u_t at x applied to v_t and $\mathrm{P}_\mu : L^2(\mu) \to \mathbf{T}_\mu$ is the orthogonal projection. In particular, we have that the acceleration of a curve is given by

$$\nabla_{v_t}^{W_2} v_t = \partial_t v_t + \frac{1}{2}\nabla|v_t|^2. \tag{8}$$

Once the covariant derivative has been constructed, we can also define the Jacobian of a vector field. A vector field $\mathscr{U}$ on $\mathcal{P}_2(\mathbb{R}^d)$ is a map $\mu \mapsto \mathscr{U}(\mu) \in \mathbf{T}_\mu$ that associates to $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ a gradient vector field over $\mathbb{R}^d$. We define the Jacobian by

$$\mathrm{D}_v^{W_2}\mathscr{U} = \nabla_{v_t}^{W_2}\mathscr{U}(\mu_t)\Big|_{t=0},$$

where μ_t is any curve such that $\mu_0 = \mu$ and $v_0 = v$ Finally, we deonte by $^A\mathrm{D}^{W_2}$ the antisymmetric part of the Jacobian, i.e.

$$\langle {}^A\mathrm{D}_v^{W_2}\mathscr{U}, w\rangle_{\mathbf{T}_\mu} := \frac{1}{2}\left(\langle\mathrm{D}_v^{W_2}\mathscr{U}, w\rangle_{\mathbf{T}_\mu} - \langle\mathrm{D}_w^{W_2}\mathscr{U}, v\rangle_{\mathbf{T}_\mu}\right) \tag{9}$$

Some classical functionals and their gradients. Consider a functional $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$. Then, its gradient $\nabla^{W_2}\mathcal{F}$ is defined via

$$\langle\nabla^{W_2}\mathcal{F}, v\rangle_{\mathbf{T}_\mu} = \lim_{h\to 0}\frac{1}{h}\left(\mathcal{F}(\mu_h) - \mathcal{F}(\mu_0)\right), \tag{10}$$

where (μ_t) is any curve such that $\mu_0 = \mu$ and $v_0 = v$.

Let us recall some common functionals and compute their gradients. As we will see, they express the acceleration of the curves we study in this article.

- For a potential U, the corresponding energy functional is $\mathcal{E}_U(\mu)$ is

$$\mathcal{E}_U(\mu) = \int_{\mathbb{R}^d} U(x)\mu(x)\mathrm{d}x. \tag{11}$$

  Its gradient is easily computed. We have

$$\nabla^{W_2}\mathcal{E}_U(\mu) = \nabla U \tag{12}$$

  However, one has to keep in mind that the right hand side still depends on μ, because the properites of ∇U as a tangent vector on $\langle.,.\rangle_{\mathbf{T}_\mu}$ vary with μ, although, as a vector field on $\mathbb{R}^d$, ∇U doesn't change.

- The Shannon Entropy functional is

$$\mathcal{S}(\mu) = -\int_{\mathbb{R}^d}\log\mu(x)\mu(x)dx \tag{13}$$

  The gradient of $\mathcal{S}$ is known:

$$\nabla^{W_2}\mathcal{S}(\mu) = -\nabla\log\mu \tag{14}$$

- The Fisher information is the squared norm of the gradient of the Shannon entropy, i.e.

$$\mathcal{I}(\mu) := \int_{\mathbb{R}^d}|\nabla\log\mu|^2\mu(x)dx \tag{15}$$

  Its gradient is
$$\nabla^{W_2}\mathcal{I}(\mu) = -\nabla\left(|\nabla\log\mu|^2 + 2|\Delta\log\mu|\right), \tag{16}$$
  see for instance [81, pg. 12], [20, Eq. (5)]. It should be observed that this relation can already be found in [5, p.172].

A. Time-symmetric fluid-dynamic formulation of SBP

We proceed to derive below a characterization of the entropic interpolation (4) in the Riemannian-like geometry of Wasserstein space when the prior is any stationary Markov diffusion measure. Assume that the prior distribution P is the stationary dynamics for a Markov generator of the form

$$\mathcal{L} = \frac{\sigma}{2}\Delta + b \cdot \nabla, \tag{17}$$

where the vector field b may not be of gradient type. We obtain the following fluid-dynamical formulation

Theorem 1 *The entropic interpolation between μ_0, μ_1 is an optimal solution for the problem*

$$\inf_{(\rho,v)} \frac{1}{2} \int_0^1 |v_t - \frac{\sigma}{2}\nabla^{W_2}\mathcal{S}(\rho_t) - \mathscr{B}(\rho_t)|^2_{\mathbf{T}_{\rho_t}} dt \tag{18a}$$

$$\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0, \quad v_t \in \mathbf{T}_{\rho_t} \tag{18b}$$

$$\rho_0 = \mu_0, \quad \rho_1 = \mu_1, \tag{18c}$$

where

$$\mathscr{B}(\rho) := \mathrm{P}_\rho(b), \tag{19}$$

and $\mathrm{P}_\rho : L^2(\rho) \to \mathbf{T}_\rho$ is the orthogonal projection and $\mathcal{S}$ the Shannon entropy (13).

Proof. It is known that the Schrödinger bridge is the law of a diffusion process Q whose generator is of the same form as P, i.e.

$$\frac{\sigma}{2}\Delta + c_t \cdot \nabla,$$

for some time-dependent vector field $c_t(x)$. For such processes, we can rewrite the Kullback-Liebler divergence using Itô calculus as

$$\mathbb{D}(Q\|P) = \mathbb{D}(q_0\|p_0) + \frac{1}{2}\mathbb{E}_Q\Big(\int_0^1 |(c_t - b)(\omega_t)|^2 dt\Big) = \frac{1}{2}\int_0^1 \int_{\mathbb{R}^d} |c_t - b|^2(x)\rho_t(x)dxdt,$$

where ρ_t is the marginal at time t of Q. The marginal flow (ρ_t) is known to satisfy the Fokker Planck equation

$$\partial_t \rho_t = -\nabla \cdot (c_t \rho_t) + \frac{\sigma}{2}\Delta \rho_t,$$

Therefore we arrive at a first reformulation of the problem as

$$\inf_{(\rho,c)} \frac{1}{2} \int_0^1 \int_{\mathbb{R}^d} |c_t - b|^2(x)\rho_t(x)dxdt, \tag{20a}$$

$$\partial_t \rho_t + \nabla \cdot \Big((c_t - \frac{\sigma}{2}\nabla \log \rho_t)\rho_t\Big) = 0, \tag{20b}$$

$$\rho_0 = \mu_0, \quad \rho_1 = \mu_1, \tag{20c}$$

Next, we claim that the latter is equivalent to

$$\inf_{(\rho,w)} \frac{1}{2} \int_0^1 \int_{\mathbb{R}^d} |w_t - \mathrm{P}_{\rho_t}(b)|^2(x)\rho_t(x)dxdt, \tag{21a}$$

$$\partial_t \rho_t + \nabla \cdot \Big((w_t - \frac{\sigma}{2}\nabla \log \rho_t)\rho_t\Big) = 0, \quad w_t \in \mathbf{T}_{\rho_t} \tag{21b}$$

$$\rho_0 = \mu_0, \quad \rho_1 = \mu_1, \tag{21c}$$

To see this, assume that (ρ_t^*, w_t^*) is optimal for (21). Then, it is not hard to show that if we set $c_t^* := w_t^* + b - P_{\rho_t^*}(b)$, then (ρ_t^*, c_t^*) is admissible and optimal for (20). On the other hand it is known, see e.g. [23] that the optimal solution (ρ_t^*, c_t^*) in (20) is such that $c_t^* - b$ is of gradient type, i.e. it belongs to $\mathbf{T}_{\rho_t^*}$. Thus, if we set $w_t^* := c_t^* - b + P_{\rho_t^*}(b)$ we have that (ρ_t^*, w_t^*) is admissible for (21); after some standard calculations it is not hard to see that it is also optimal. Upon setting $v_t = w_t - \frac{\sigma}{2}\nabla\log\rho_t$, we obtain from (20) another equivalent formulation of (21) as

$$\inf_{(\rho,v)} \frac{1}{2}\int_0^1 \int_{\mathbb{R}^d} |v_t + \frac{\sigma}{2}\nabla\log\rho_t - P_{\rho_t}(b)|^2(x)\rho_t(x)\mathrm{d}x\mathrm{d}t, \tag{22a}$$

$$\partial_t\rho_t + \nabla\cdot(v_t\rho_t) = 0, \quad v_t \in \mathbf{T}_{\rho_t} \tag{22b}$$

$$\rho_0 = \mu_0, \quad \rho_1 = \mu_1, \tag{22c}$$

The conclusion now follows from the definitons (19),(14)

Lemma 1 *The formulation* (18) *is equivalent to*

$$\inf_{(\rho,v)} \frac{1}{2}\int_0^1 |v_t|_{\mathbf{T}_{\rho_t}}^2 + \frac{\sigma^2}{4}\mathcal{I}(\rho_t) + |\mathscr{B}(\rho_t)|_{\mathbf{T}_{\rho_t}}^2 + \mathcal{E}_{\sigma\nabla\cdot b}(\rho_t) - 2\langle\mathscr{B}(\rho_t), v_t\rangle_{\mathbf{T}_{\rho_t}}\mathrm{d}t \tag{23a}$$

$$\partial_t\rho_t + \nabla\cdot(v_t\rho_t) = 0, \quad v_t \in \mathbf{T}_{\rho_t} \tag{23b}$$

$$\rho_0 = \mu_0, \quad \rho_1 = \mu_1, \tag{23c}$$

Proof. It suffices to expand the squares, use the defintion (15), observe that the product $-\sigma\int_0^1 \langle\nabla^{W_2}\mathcal{S}(\rho_t), v_t\rangle_{\mathbf{T}_{\rho_t}}\mathrm{d}t$ is constantly equal to $\mathcal{S}(\mu_0) - \mathcal{S}(\mu_1)$ over the admissible region and can thus be neglected, and to rewrite the cross product $\sigma\int_0^1 \langle\nabla^{W_2}\mathcal{S}(\rho_t), \mathscr{B}(\rho_t)\rangle_{\mathbf{T}_{\rho_t}}\mathrm{d}t$ in the following way:

$$\sigma\int_0^1 \langle\nabla^{W_2}\mathcal{S}(\rho_t), \mathscr{B}(\rho_t)\rangle_{\mathbf{T}_{\rho_t}}\mathrm{d}t$$

$$= -\sigma\int_0^1 \int_{\mathbb{R}^d} \langle\nabla\log\rho_t(x), b(x)\rangle\rho_t(x)\mathrm{d}x\mathrm{d}t$$

$$= -\sigma\int_0^1 \int_{\mathbb{R}^d} \langle\nabla\rho_t, b\rangle(x)\mathrm{d}x\mathrm{d}t$$

$$= \sigma\int_0^1 \int_{\mathbb{R}^d} (\nabla\cdot b)\rho_t(x)\mathrm{d}x\mathrm{d}t$$

$$\stackrel{(11)}{=} \int_0^1 \mathcal{E}_{\sigma\nabla\cdot b}(\rho_t)\mathrm{d}t.$$

V. VARIATIONAL ANALYSIS

The goal of this section is to prove the following result.

Theorem 2 *Let* $(\mu_t, 0 \leq t \leq 1)$ *be the entropic interpolation between* μ *and* ν *and* (v_t) *its velocity field. Then* (μ_t) *satisfies the equation*

$$\nabla_{v_t}^{W_2}v_t = \nabla^{W_2}\left[\frac{\sigma}{8}\mathcal{I} + \frac{1}{2}|\mathscr{B}|_{\mathbf{T}.}^2 + \frac{1}{2}\mathcal{E}_{\sigma\nabla\cdot b}\right](\mu_t) + 2\,^A\mathrm{D}_{v_t}^{W_2}\mathscr{B} \tag{24}$$

In the reversible case, i.e. when $b = -\nabla U$, we obtain the following

Corollary 1 *Let* $b = -\nabla U$, (μ_t) *be the entropic interpolation between* μ *and* ν *and* (v_t) *its velocity field. Then* (μ_t) *satisfies the equation*

$$\nabla^{W_2}_{v_t} v_t = \nabla^{W_2} \left[\frac{\sigma}{8}\mathcal{I} + \frac{1}{2}\mathcal{E}_{|\nabla U|^2 - \sigma \Delta U} \right](\mu_t) \tag{25}$$

Note that, in the above corollary, the quantity $|\nabla U|^2 - \sigma \Delta U$ is the *reciprocal characteristic* associated to the potential U. Among others Levy, Krener and Thieullen [45, 49, 77] observed that this quantity should express a kind of mean acceleration for the Schrödinger bridge, or, more generally, for reciprocal processes. Equation (25) clearly shows that this is the case. Notice that the covariant derivative has also been computed in [35, p.6] without relating it to the gradient of the Fisher information functional. Prior to the proof, let us introduce some useful terminology. In analogy to the classical calculus of variations, we call a *variation* of (μ_t) a family of curves (ρ^s_t) such that

- $(\rho^0_t) = (\mu_t)$

- $\forall s \in (-\varepsilon, \varepsilon), \quad (\rho^s_0) = \mu_0$ and $\rho^s_1 = \mu_1$.

For any fixed s, v^s_t stands for the velocity field of the curve $t \mapsto \rho^s_t$. That is

$$\partial_t \rho^s_t + \nabla \cdot (v^s_t \rho^s_t) = 0, \tag{26}$$

and v^s_t is a gradient vector field. In particular, we have that $(v^0_t) = (v_t)$, where v_t is the velocity field of (μ_t). On the other hand, for fixed t, w^s_t is the velocity field of the curve $s \mapsto \rho^s_t$, i.e.

$$\partial_s \rho^s_t + \nabla \cdot (w^s_t \rho^s_t) = 0, \tag{27}$$

and w^s_t is a gradient vector field.

Proof.

Consider a variation (ρ^s_t) of (μ_t). Imposing the optimality condition of (μ_t) in (23) yields

$$\int_0^1 \partial_s \left(\frac{1}{2}|v^s_t|^2_{\mathbf{T}_{\rho^s_t}} + \frac{\sigma^2}{8}\mathcal{I}(\rho^s_t) + \frac{1}{2}|\mathscr{B}|^2_{\mathbf{T}_{\rho^s_t}} + \frac{1}{2}\mathcal{E}_{\sigma \nabla \cdot b}(\rho^s_t) - \frac{1}{2}\langle v^s_t, \mathscr{B}(\rho^s_t)\rangle_{\mathbf{T}_{\rho^s_t}} \right) \Big|_{s=0} dt \leq 0 \tag{28}$$

For convenience in the analysis, set

$$A := \int_0^1 \partial_s \left(\frac{1}{2}|v^s_t|^2_{\mathbf{T}_{\rho^s_t}} \right) \Big|_{s=0} dt$$

$$B := \int_0^1 \partial_s \left(\frac{\sigma^2}{8}\mathcal{I}(\rho^s_t) + |\mathscr{B}|^2_{\mathbf{T}_{\rho^s_t}} + \frac{1}{2}\mathcal{E}_{\sigma \nabla \cdot b}(\rho^s_t) \right) \Big|_{s=0} dt$$

$$C := -\int_0^1 \partial_s \left(\langle v^s_t, \mathscr{B}(\rho^s_t)\rangle_{\mathbf{T}_{\rho^s_t}} \right) \Big|_{s=0} dt$$

Let us work separately on each term.

Term A Using the product rule we obtain

$$\frac{1}{2}\int_0^1 \partial_s |v_t^s|^2_{\mathbf{T}_{\rho_t^s}}\mathrm{d}t = \underbrace{\int \langle v_t^s, \partial_s v_t^s\rangle \rho_t^s \mathrm{d}x\mathrm{d}t}_{:=A.1} + \underbrace{\frac{1}{2}\int |v_t^s|^2 \partial_s \rho_t^s \mathrm{d}x\mathrm{d}t}_{:=A.2}, \tag{29}$$

where use the shorthand notation $\int$ for $\int_0^1 \int_{\mathbb{R}^d}$. We keep doing this throughout the proof. Since v_t^s is a gradient vector field, we have that $v_t^s = \nabla \varphi_t^s$ for some potential φ_t^s. With a first integration by parts in space we get

$$A.1 = -\int \varphi_t^s \nabla \cdot (\partial_s v_t^s \rho_t^s)\,\mathrm{d}t\mathrm{d}x \tag{30}$$

Next, we observe that the following identity can be obtained differentiating (26) with respect to s and using (27)

$$\partial_s \partial_t \rho_t^s + \nabla \cdot (v_t^s\,\partial_s \rho_t^s + \partial_s v_t^s\,\rho_t^s) = 0. \tag{31}$$

The identity (31) allows us to rewrite (30) as

$$A.1 = \underbrace{\int \varphi_t^s\,\partial_t\partial_s\rho_t^s\mathrm{d}x\mathrm{d}t}_{A.1.1} + \underbrace{\int \varphi_t^s\,\nabla \cdot (v_t^s\partial_s\rho_t^s)\,\mathrm{d}x\mathrm{d}t}_{A.1.2} \tag{32}$$

Using subsequently the continuity equations (26),(27) and integration by parts we have

$$A.1.1 = \int \partial_t\varphi_t^s\,\nabla \cdot (w_t^s\rho_t^s)\,\mathrm{d}x\mathrm{d}t$$

$$= -\int \langle \partial_t v_t^s, w_t^s\rangle \rho_t^s \mathrm{d}x\mathrm{d}t,$$

where we used the fact that $\partial_s \rho_0^s = \partial_s \rho_1^s = 0$ to get rid of the boundary terms when integrating by parts in time. Moreover, integrating by parts in space and using (27)

$$A.1.2 = \int |v_t^s|^2 \nabla \cdot (w_t^s\rho_t^s)\,\mathrm{d}x\mathrm{d}t$$

$$= -\int \langle \nabla\left(|v_t^s|^2\right), w_t^s\rangle \rho_t^s \mathrm{d}x\mathrm{d}t$$

Plugging the newly obtained expressions for $A.1.1$ and $A.1.2$ back into (32) we obtain

$$A.1 = -\int \langle \partial_t v_t^s + \nabla\left(|v_t^s|^2\right), w_t^s\rangle \rho_t^s \mathrm{d}x\mathrm{d}t$$

Concerning the term $A.2$ in (29), using (27) and an integration by parts we get

$$A.2 = \frac{1}{2}\int \langle \nabla\left(|v_t^s|^2\right), w_t^s\rangle \rho_t^s \mathrm{d}x\mathrm{d}t$$

so that the whole expression in (29) equals

$$-\int \langle \partial_t v_t^s + \frac{1}{2}\nabla\left(|v_t^s|^2\right), w_t^s\rangle \rho_t^s \mathrm{d}x t \tag{33}$$

Evaluating at $s = 0$, and using (8), we obtain

$$A = -\int \langle \nabla_{v_t}^{W_2} v_t, w_t^0\rangle_{\mathbf{T}_{\mu_t}}\mathrm{d}t \tag{34}$$

Term B For this term, we have, using (10):

$$B = \int_0^1 \left\langle \nabla^{W_2} \left[\frac{\sigma^2}{8} \mathcal{I} + \frac{1}{2} |\mathcal{B}|_{\mathbf{T}}^2 + \mathcal{E}_{\sigma \nabla \cdot b} \right] (\mu_t), w_t^0 \right\rangle_{\mathbf{T}_{\mu_t}} \mathrm{d}t. \tag{35}$$

Term C First observe that we can rewrite it as

$$- \int \langle \partial_s v_t^s |_{s=0}, \mathcal{B}(\mu_t) \rangle \mu_t + \langle v_t, \partial_s \mathcal{B}(\rho_t^s)|_{s=0} \rangle \mu_t + \langle v_t, \mathcal{B}(\mu_t) \rangle \partial_s \rho_t^s |_{s=0} \mathrm{d}x \mathrm{d}t \tag{36}$$

Let us focus on the first summand. Since $\mathcal{B}(\mu_t)$ is a gradient vector field, we may write it in the form ∇U_t for some potential U. Thus we have, using integration by parts,

$$- \int \langle \partial_s v_t^s |_{s=0}, \mathcal{B}(\mu_t) \rangle \mu_t \mathrm{d}x \mathrm{d}t$$

$$= - \int \langle \partial_s v_t^s |_{s=0}, \nabla U_t \rangle \mu_t \mathrm{d}x \mathrm{d}t$$

$$= \int \nabla \cdot (\partial_s v_t^s |_{s=0} \mu_t) U_t \mathrm{d}x \mathrm{d}t.$$

Taking advantage of (31), we can write the last integral as

$$- \int \partial_s \partial_t \rho_t^s |_{s=0} U_t \, \mathrm{d}x \mathrm{d}t - \int \nabla \cdot (\partial_s \rho_t^s |_{s=0} v_t) U_t \, \mathrm{d}x \mathrm{d}t$$

$$\overset{(27)}{=} \int \partial_t (\nabla \cdot (\mu_t w_t^0)) U_t \, \mathrm{d}x \mathrm{d}t + \int \langle v_t, \mathcal{B}(\mu_t) \rangle \, \partial_s \rho_t^s |_{s=0} \, \mathrm{d}x \mathrm{d}t$$

$$= \int \langle w_t^0, \partial_t \mathcal{B}(\mu_t) \rangle \mu_t \mathrm{d}x \mathrm{d}t + \int \langle v_t, \mathcal{B}(\mu_t) \rangle \partial_s \rho_t^s |_{s=0} \mathrm{d}x \mathrm{d}t$$

Note now that the second term of this last expression cancels with the third term in (36). Therefore the whole expression in (36) equals

$$\int \left(\langle w_t^0, \partial_t \mathcal{B}(\mu_t) \rangle - \langle v_t, \partial_s \mathcal{B}(\rho_t^s)|_{s=0} \rangle \right) \mu_t \mathrm{d}x \mathrm{d}t. \tag{37}$$

Next, since $\mathcal{B}(\rho_t^s)$ is a gradient vector field we have

$$\int \langle w_t^0, \mathrm{D}_{v_t} \mathcal{B}(\mu_t) \rangle \mu_t \mathrm{d}x \mathrm{d}t = \int \langle v_t, \mathrm{D}_{w_t^0} \mathcal{B}(\mu_t) \rangle \mu_t \mathrm{d}x \mathrm{d}t$$

and therefore we can rewrite (37) as

$$\int \left(\left\langle w_t^0, \partial_t \mathcal{B}(\mu_t) + \mathrm{D}_{v_t} \mathcal{B}(\mu_t) \right\rangle - \left\langle v_t, \left(\partial_s \mathcal{B}(\rho_t^s) + \mathrm{D}_{w_t^s} \mathcal{B}(\rho_t^s) \right) \big|_{s=0} \right\rangle \right) \mu_t \mathrm{d}x \mathrm{d}t$$

$$= \int \left(\left\langle w_t^0, \mathrm{P}_{\mu_t} \left(\partial_t \mathcal{B}(\mu_t) + \mathrm{D}_{v_t} \mathcal{B}(\mu_t) \right) \right\rangle - \left\langle v_t, \mathrm{P}_{\mu_t} \left(\partial_s \mathcal{B}(\rho_t^s) + \mathrm{D}_{w_t^0} \mathcal{B}(\mu_t) \right) \right\rangle \right) \mu_t \mathrm{d}x \mathrm{d}t$$

$$\overset{(7)}{=} \int \langle w_t^0, \mathrm{D}_{v_t}^{W_2} \mathcal{B} \rangle_{\mathbf{T}_{\mu_t}} - \langle v_t, \mathrm{D}_{w_t^0}^{W_2} \mathcal{B} \rangle_{\mathbf{T}_{\mu_t}} \mathrm{d}t$$

$$\overset{(9)}{=} 2 \int \langle w_t^0, {}^A\mathrm{D}_{v_t}^{W_2} \mathcal{B} \rangle_{\mathbf{T}_{\mu_t}} \mathrm{d}t.$$

Thus

$$C = 2 \int \langle w_t^0, \, {}^A\mathrm{D}_{v_t}^{W_2} \mathscr{B} \rangle_{\mathbf{T}_{\mu_t}} \mathrm{d}t \tag{38}$$

Putting togehter (33),(35),(38) we can rewrite the criticality condition (28) as

$$\int_0^1 \left\langle -\nabla_{v_t}^{W_2} v_t + \nabla^{W_2} \left[\frac{\sigma^2}{8}\mathcal{I} + \frac{1}{2}|\mathscr{B}|_{\mathbf{T}.}^2 + \frac{1}{2}\mathcal{E}_{\sigma\nabla \cdot b} \right](\mu_t) + 2\,{}^A\mathrm{D}_{v_t}^{W_2}\mathscr{B}, w_t^0 \right\rangle_{\mathbf{T}_{\mu_t}} \mathrm{d}t \leq 0$$

The desired conclusion now follows from the fact that w_t^0 can be chosen arbitrarily among the gradient vector fields.

VI. MADELUNG FLUID AND STOCHASTIC MECHANICS

A. The Madelung Fluid

At the very dawn of quantum mechanics, in 1926, Erwin Madelung proposed in [54] a fluid dynamic formulation of quantum mechanics, see [36] for a beautiful account. He noticed that if $\{\psi(x,t), t_0 \leq t \leq t_1\}$ satisfies the Schrödinger equation

$$\frac{\partial \psi}{\partial t} = \frac{i\hbar}{2m}\Delta\psi - \frac{i}{\hbar}V(x)\psi, \tag{39}$$

then, writing $\psi(x,t) = \exp[R(x,t) + \frac{i}{\hbar}S(x,t)]$, $\rho_t(x) = |\psi(x,t)|^2 = \exp[2R(x,t)]$, we get for R and S the system of partial differential equations

$$\frac{\partial R}{\partial t} + \frac{1}{m}\nabla R \cdot \nabla S + \frac{1}{2m}\Delta S = 0, \tag{40a}$$

$$\frac{\partial S}{\partial t} + \frac{1}{2m}\nabla S \cdot \nabla S + V - \frac{\hbar^2}{2m}[\nabla R \cdot \nabla R + \Delta R] = 0. \tag{40b}$$

This system represents the *Madelung fluid*. If we set $\mu_t(x) = \exp(2R(x,t))$ and $v_t = \frac{1}{m}\nabla S_t$, we can rewrite the system above, after taking the spatial gradient in the second equation and performing some standard calculations, in the following way

$$\partial_t \mu_t + \nabla \cdot (v_t \rho_t) = 0, \tag{41a}$$

$$\partial_t v_t + \frac{1}{2}\nabla|v_t|^2 + \frac{1}{m}\nabla V - \frac{\hbar^2}{8m^2}\nabla\left[|\nabla \log \mu_t|^2 + 2\Delta \log \mu_t\right] = 0. \tag{41b}$$

Using (8),(16),(12) we obtain that the PDE system above has the clear interpretation of a Newton's law, when (μ_t) is seen as a curve in the Riemannian manifold of optimal transport:

$$\nabla_{v_t}^{W_2} v_t = -\frac{\hbar^2}{8m^2}\nabla^{W_2}\mathcal{I}(\mu_t) - \frac{1}{m}\nabla^{W_2}\mathcal{E}_V(\mu_t)$$

This fact was first understood by Von Renesse [81] after some early efforts to connect OMT to stochastic mechanics [11], [80, p.707]. Note that this equation has the same form of (25), with the only (remarkable) difference of the minus sign in front of the gradient of the Fisher information. This variational fluid-dynamic formulation fully clarifies the relation between Shrödinger Bridges and Nelson's stochastic mechanics (see next subsection) which

has puzzled mathematical physicists since Schrödinger in [75][7], see also [83]. Notice that the Lagrangian with the plus sign also plays a role in Nelson's stochastic mechanics: It appears in the particle (first form) of Hamilton's-like principle due to Yasue [61, 82]) which relies on stochastic calculus. The Lagrangian with the minus sign in front of the Fisher information (Guerra-Morato Lagrangian) appears instead in the second, fluid dynamic formulation [37, 61]. This has contributed in the past to the confusion on the relation between the two theories which we hope this paper has definitely resolved. Historically, equation (40a) was recognized already by Madelung in 1926 to be just a continuity equation. The second was thought to be a Hamilton-Jacobi equation featuring, besides the classical potential V, the "quantum potential" [5][8]

$$-\frac{\hbar^2}{2m}\left[\nabla R \cdot \nabla R + \Delta R\right] = -\frac{\hbar^2}{2m}\frac{\Delta\sqrt{\rho}(x,t)}{\sqrt{\rho_t(x)}}.$$

The presence of such a "nonlocal" potential, however, mystified researchers for a long time.

B. Remarks on stochastic mechanics

The first formulation of quantum mechanics through stochastic processes is due to the Hungarian theoretical physicist Imre Fenyes in a series of papers between 1946 and 1952. In the abstract of [27], we find the following impressive statement:"Die wellenmechanischen Prozesse sind spezielle MARKOFFsche Prozesse. Die Relation von Heisenberg ist (im Gegensatz zur bisherigen Interpretation) ausschliesslich eine Folge der statistischen Betrachtungsweise, und ist von den bei den Messungen auftretenden Störungen unabhängig.[9]" After some further contributions by Weizel in the fifties [60, p.135], Edward Nelson provided an elegant formulation of stochastic mechanics in 1966-67 [59, 60] by developing a new kinematics for continuous stochastic evolutions. The Schrödinger equation was originally derived from the continuity equation (40b) plus a Newton type law. The Newton-Nelson law was later shown to follow, in analogy to classical mechanics, from a Hamilton-like stochastic variational principle [37, 82]. Other versions of the variational principle have been proposed in [8, 58, 61, 65, 73]. One of the most striking differences between Nelson's and other versions of mechanics such as Bohmian mechanics [5–7] or the Levy-Krener mechanics [48, 50] is that the former features a kinematics for finite-energy diffusions with *two* velocities. Indeed, let $\{\psi(x,t); t_0 \le t \le t_1\}$, satisfy (39) be such that

$$||\nabla\psi||_2^2 \in L^1_{\text{loc}}[t_0, +\infty). \tag{42}$$

This is Carlen's *finite action condition*. Under these hypotheses, the Nelson measure P may be constructed on path space, [10],[12], [8, Chapter IV], and references therein. Namely,

---

[7] Schrödinger writes: *"Merkwürdige Analogien zur Quantenmechanik, die mir sehr des Hindenkens wert erscheinen"* (remarkable analogies to quantum mechanics which appear to me very worth of reflection).

[8] It is truly unfortunate that Bohm, who had related such quantum potential to the gradient of the Fisher information functional [5, p.172], did not realized that this made postulating such a non physical potential unnecessary. This has, among other things, led to an endless controversy between Bohmian and Nelsonian followers.

[9] "The wave mechanical Processes are special MARKOFF processes. The relation of Heisenberg (in contrast to the previous interpretation) is exclusively a consequence of the statistical approach, and is independent of the disturbances occurring during the measurements."

letting $\Omega := \mathcal{C}([t_0, t_1], \mathbb{R}^d)$ the n-dimensional continuous functions on $[t_0, t_1]$, under the probability measure P, the canonical coordinate process $x(t, \omega) = \omega(t)$ is an n-dimensional, Markov, finite-energy diffusion process $\{x(t); t_0 \leq t \leq t_1\}$, called *Nelson's process*, having forward and backward Ito differentials

$$dx = \left[\frac{\hbar}{m}\nabla\left(R + \frac{1}{\hbar}S\right)\right](x(t), t)dt + \sqrt{\frac{\hbar}{m}}dw_+(t), \tag{43a}$$

$$dx = \left[\frac{\hbar}{m}\nabla\left(-R + \frac{1}{\hbar}S\right)\right](x(t), t)dt + \sqrt{\frac{\hbar}{m}}dw_-(t), \tag{43b}$$

where w_+, w_- are standard, n-dimensional Wiener process. Moreover, the probability density $\rho(\cdot, t)$ of $x(t)$ satisfies Born's relation

$$\rho_t(x) = |\psi(x, t)|^2, \quad \forall t \in [t_0, t_1], \tag{44}$$

and

$$v_t(x) = \frac{1}{m}\nabla S(x, t), \quad u(x, t) = \frac{\hbar}{m}\nabla R(x, t)$$

are the current and osmotic velocity fields. It should be emphasized, that Nelson's stochastic mechanics, when compared to other alternative formulations, features the following advantages: There is no need to postulate any potential besides the physical potential V, but only to adopt what is the natural kinematics for finite energy diffusions [29, 30]. Notice that such a kinematics leads naturally to the complexification of the velocity and of the momentum [52, 65, 66]. The latter fact permits to develop Hamilton-like equations and provides a *kinematical* justification (in constrast to the countless *mathematical* justifications) for the need of complex quantities in quantum mechanics. For a list of further "successes", as well as failures, of Nelson's stochastic mechanics see [62]. Among the successes, it might be worthwhile quoting from [62] the little known: "A stochastic picture of the two-slit experiment, explaining how particles have trajectories going through just one slit or the other, but nevertheless produce a probability density as for interfering waves; see [61, Section 17]", see also [69].

Although both the hydrodynamical formulation and the stochastic formulation are alternative to usual quantum mechanics, they are not equivalent. Indeed, let the pair (v^*, ρ^*) be optimal for Problem 4 and consider the following random evolution

$$\dot{X}(t) = v^*(X(t), t), \quad X(0) \sim \rho_0 dx,$$

where we have assumed that v^* guarantees existence and uniqueness of the initial value problem for $t \in [0, 1]$ and any deterministic initial condition. Then the probability density $\rho_t(x)$ of $X(t)$ does satisfies (weakly) (4b) with the same initial condition and therefore coincides with $\rho^*(x, t)$. This shows that variational principles on Wasserstein space such as those of this paper do not encode the same information as those formulated in terms of stochastic processes. In particular, the information on the diffusion coefficient - smoothness of trajectories is lacking. While this might not be so relevant in some applications such as image interpolation (morphing) where the flow of one time marginals is the object of interest, it is definitely crucial in physics. In stochastic mechanics, for instance, the assumption that the evolution be described by a diffusion process with diffusion coefficient $\frac{\hbar}{m}$ is made at the onset leading to stochastic control problems and corresponding Hamilton-Jacobi-Bellman equations.

Similar considerations can be made in the case of a general Markovian prior treated in Section IV. Within Nelson's stochastic mechanics, this would correspond to the variational mechanism leading to the so-called collapse of the wavefunction after a position measurement [67, 68]. The zero-noise limit of a general Schrödinger bridge, instead, is given by an OMT problem with prior formulated and studied in [16], see also [17].

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] L. Ambrosio and N. Gigli, A user's guide to optimal transport, in *Modeling and optimisation of flows on networks*, Springer Lecture Notes in Mathematics, n. 2062, B. Piccoli and M. Rascle Eds., 2013.

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

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

[5] D. Bohm, A suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables. I, *Physical Rev.* **85** 166 (1952).

[6] D. Bohm and J.P. Vigier, Phys. Rev. **96** 208 (1954).

[7] D. Bohm and B. J. Hiley, Non-locality and locality in the stochastic interpretation of quantum mechanics, *Phys. Rep* **172** (1989), 93-122.

[8] Ph. Blanchard, Ph. Combe and W. Zheng. *Math. and Physical Aspects of Stochastic Mechanics*. Lect. Notes in Physics vol. 281, (Springer-Verlag, New York, 1987).

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

[10] E. Carlen, *Comm. Math. Phys.*, **94**, 293 (1984).

[11] 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.

[12] R. Carmona, in *Taniguchi Symp. PMMP* Katata (1985), pp. 55–81.

[13] Y. Chen, T.T. Georgiou and M. Pavon, Optimal steering of a linear stochastic system to a final probability distribution, , Part I, arXiv1408.2222v1, *IEEE Trans. Aut. Control*, **61**, Issue 5, 1158-1169, 2016.

[14] Y. Chen, T.T. Georgiou and M. Pavon, Optimal steering of a linear stochastic system to a final probability distribution, Part II, arXiv1410.3447v1, *IEEE Trans. Aut. Control*, **61**, Issue 5, 1170-1180, 2016.

[15] Y. Chen, T.T. Georgiou and M. Pavon, Fast cooling for a system of stochastic oscillators, Nov. 2014, arXiv :1411.1323v2, *J. Math. Phys.*, **56**, n.11, 113302, 2015.

[16] Y. Chen, T.T. Georgiou and M. Pavon, On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint, arXiv :1412.4430v1, *J. Optim. Theory and Applic.*, **169** (2), 671-691, 2016.

[17] Y. Chen, T.T. Georgiou and M. Pavon, Optimal transport over a linear dynamical system, Feb. 2015, arXiv:1502.01265v1, *IEEE Trans. Aut. Control*, **62**, n. 5, 2137-2152, 2017.

[18] Y. Chen, T.T. Georgiou and M. Pavon, Entropic and displacement interpolation: a computational approach using the Hilbert metric, arXiv:1506.04255v1, *SIAM Journal on Applied Mathematics*, **76** (6), 2375-2396, 2016.

[19] Y. Chen, T.T. Georgiou and M. Pavon, Stochastic Control, Entropic Interpolation and Gradient Flows on Wasserstein Product Spaces, arXiv:1601.04891, Proc. 22 MTNS Conf. Minneapolis, July 2016.

[20] G. Conforti, A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost, preprint arXiv:1704.04821v1.

[21] G. Conforti and M. Pavon, Extremal curves in Wasserstein space, Proc. of *Geometric Science of Information: Third International Conference*, GSI 2017, Paris, France, November 7-9, 2017, Frank Nielsen and Frédéric Barbaresco Eds., Springer-Verlag, 2017, pp, 92-99.

[22] M. Cuturi, Sinkhorn Distances: Lightspeed Computation of Optimal Transport, *Advances in Neural Information Processing Systems*, 2292-2300, 2013.

[23] P. Dai Pra, A stochastic control approach to reciprocal diffusion processes, *Applied Mathematics and Optimization*, **23** (1), 1991, 313-329.

[24] P.Dai Pra and M.Pavon, On the Markov processes of Schroedinger, the Feynman-Kac formula and stochastic control, in Realization and Modeling in System Theory - Proc. 1989 MTNS Conf., M.A.Kaashoek, J.H. van Schuppen, A.C.M. Ran Eds., Birkaeuser, Boston, 1990, 497-504.

[25] L. de Broglie, An Introduction to the Study of Wave Mechanics, E.P. Button and Company, Inc., New York, 1930.

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

[27] I. Fényes, Eine wahrscheinlichkeitstheoretische Begründung und Interpretation der Quantenmechanik, Zeitschrift für Physik, Bd. 132, 1952, 81-106.

[28] R. Fillieger, M.-O. Hongler and L. Streit, Connection between an exactly solvable stochastic optimal control problem and a nonlinear reaction-diffusion equation, *J. Optimiz. Theory Appl.* **137** (2008), 497-505.

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

[30] 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.

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

[32] W. Gangbo, T. Nguyen, and A. Tudorascu, Hamilton-Jacobi equations in the Wasserstein space, *Methods Appl. Anal.* **15** (2008), no. 2, 155-183.

[33] I. Gentil, C. Lonard and L. Ripani, About the analogy between optimal transport and minimal entropy, arXiv: 1510.08230, to appear in *Ann. Fac. Toulouse.*

[34] N.Gigli, Second Order Analysis on (P2(M),W2), *Memoirs of the American Mathematical Society*, 2012.

[35] N. Gigli and L. Tamanini, Second order differentiation formula on compact $\mathbf{RCD}^*(K, N)$ spaces, preprint arXiv:1701.03932v1.

[36] F.Guerra, Structural aspects of stochastic mechanics and stochastic field theory *Phys.Rep.* 77 (1981)263.

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

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

[39] E. T. Jaynes, Information Theory and Statistical Mechanics, *Physical Review Series II*, **106** (4): 62017630, 1957. doi:10.1103/PhysRev.106.620. MR87305, and Information Theory and Statistical Mechanics II, *Physical Review Series II*, **108** (2): 17117190, 1957. doi:10.1103/PhysRev.108.171. MR96414.

[40] E. T. Jaynes. On the rationale of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, Sept. 1982.

[41] X. Jiang, Z. Luo, and T. Georgiou, Geometric methods for spectral analysis, *IEEE Transactions on Signal Processing*, **60**, no. 3, 2012, 1064171074.

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

[43] P. Kosmol and M. Pavon, Lagrange Lemma and the optimal control of diffusions II: Nonlinear Lagrange functionals, *Systems control letters*, 1995, **24** (3): 215-221.

[44] F. Léger, A geometric perspective on regularized optimal transport, arXiv: 1703.10243v2.

[45] A.J. Krener, Reciprocal diffusions in flat space, *Probability Theory and Related Fields* **107**(2)(1997),243-281.

[46] 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.

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

[48] B.C.Levy and A.J.Krener, Kinematics and dynamics of reciprocal diffusions, *J.Math.Phys.* **34** (1993) 1846.

[49] B.C. Levy and A.J. Krener, Dynamics and kinematics of reciprocal diffusions, *Journal of Mathematical Physics* **5**(34)(1993),1846-1875.

[50] B.C.Levy and A.J.Krener,Stochastic mechanics of reciprocal diffusions, *J.Math.Phys.* **37** (1996), 769.

[51] W. Li, P. Yin and S. Osher, Computations of optimal transport distance with Fisher information regularization, arXiv:1704.04605v1.

[52] J. Köppe, W. Grecksch, and W. Paul, Derivation and application of quantum Hamilton equations of motion, *Annalen der Physik*, **529**, n. 3, p. 1600251, 2017.

[53] W. Li, P. Yin and S Osher, Computations of optimal transport distance with Fisher information regularization, preprint arXiv:1704.04605.

[54] E. Madelung, Quantentheorie in hydrodynamischer Form, *Z. Physik* **40** (1926), 322–326.

[55] 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.

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

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

[58] L.M. Morato, *Stochastic Quantization and Coherence*, Proc. of the Int. Conference on "Quantum-like models and Coherent effects", Erice, June 13-20 1994. R. Fedele and P.K. Shuckla, 97-112, World Scientific 1995.

[59] E. Nelson, Derivation of the Schrödinger equation from Newtonian mechanics, *Physical Review*, **150**, 1079-1085,1966.

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

[61] E. Nelson, Quantum Fluctuations, Princeton University Press, 1985.

[62] E. Nelson, Review of Stochastic mechanics, *J. Phys. Conf. Ser.*, **361**, p.012011, 2012.

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

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

[65] M. Pavon, Hamilton's principle in stochastic mechanics *J. Math. Phys.* **36**, 6774 (1995).

[66] M. Pavon, A new formulation of stochastic mechanics, *Physics Letters A* **209** (1995), 143-149, Erratum 211 (1996), 383.

[67] M. Pavon, Derivation of the wavefunction collapse in the context of Nelson's stochastic mechanics, *Journal of Mathematical Physics* **40** (1999), 5565-5577.

[68] M. Pavon, Quantum Schrödinger bridges, in *Directions in Mathematical Systems Theory and Optimization*, A. Rantzer and C. I. Byrnes Eds., Springer-Verlag, 2003, 227-238.

[69] M. Pavon, A footnote to Nelson's interpretation of the two-slit experiment, *Int. J. Modern Phys. B.*, **18** (2004), 745-753.5

[70] M.Pavon and A.Wakolbinger, On free energy, stochastic control, and Schroedinger processes, Modeling, Estimation and Control of Systems with Uncertainty, G.B. Di Masi, A.Gombani, A.Kurzhanski Eds., Birkauser, Boston, 1991, 334-348.

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

[72] I. S. Sanov, On the probability of large deviations of random magnitudes (in Russian), *Mat.*

Sb. N. S., **42** (84) (1957) 111744. Select. Transl. Math. Statist. Probab., 1, 213-244 (1961).

[73] H. H. Rosenbrock, Doing quantum mechanics with control theory, *IEEE Trans. Aut. Contr.* **54**, 73-77 (2000).

[74] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Birkhäuser, 2015.

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

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

[77] M.Thieullen, Second order stochastic differential equations and non-Gaussian reciprocal diffusions, *Probability Theory and Related Fields* **97**(1)(1993),231-257.

[78] A. M. Vershik, Long history of the Monge-Kantorovich problem, *The Mathematical Intellicencer*, DOI 10.1007/s00283-013-9380-x, puplished online May 30, 2013.

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

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

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

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

[83] J. C. Zambrini, Variational processes and stochastic versions of mechanics, *J. Math. Phys.*, **27** (9) (1986), 2307-2330.