Constrained Steepest Descent in the 2–Wasserstein Metric

E. A. Carlen⋆ and W. Gangbo⋆⋆

School of Mathematics
Georgia Institute of Technology
Atlanta, GA, 30332 U.S.A.

Abstract We study several constrained variational problem in the 2–Wasserstein metric for which the set of probability densities satisfying the constraint is not closed. For example, given a probability density F_0 on $I\!R^d$ and a time–step $h > 0$, we seek to minimize $I(F) = hS(F) + W_2^2(F_0, F)$ over all of the probability densities F that have the same mean and variance as F_0, where $S(F)$ is the entropy of F. We prove existence of minimizers. We also analyze the induced geometry of the set of densities satisfying the constraint on the variance and means, and we determine all of the geodesics on it. From this, we determine a criterion for convexity of functionals in the induced geometry. It turns out, for example, that the entropy is uniformly strictly convex on the constrained manifold, though not uniformly convex without the constraint. The problems solved here arose in a study of a variational approach to constructing and studying solutions of the non–linear kinetic Fokker–Planck equation, which is briefly described here and fully developed in a companion paper.

Key words: mass transfer, Wasserstein metric, constrained optimization

Contents

Introduction

Recently there has been considerable progress in understanding a wide range of dissipative evolution equations in terms of variational problems involving the Wasserstein metric. In particular, Jordan, Kinderlehrer and Otto, have shown in [12] that the heat equation is gradient flow for the entropy functional in the 2–Wasserstein metric. We can arrive most rapidly to the point of departure for our own problem, which concerns constrained gradient flow, by reviewing this result.

Let $\mathcal{P}$ denote the set of probability densities on $I\!\!R^d$ with finite second moments; i.e., the set of all non–negative measurable functions F on $I\!\!R^d$ such that $\int_{I\!\!R^d} F(v)\mathrm{d}v = 1$ and $\int_{I\!\!R^d} |v|^2 F(v)\mathrm{d}v < \infty$. We use v and w to denote points in $I\!\!R^d$ since in the problem to be described below they represent velocities. Equip $\mathcal{P}$ with the 2–Wasserstein metric, $W_2(F_0, F_1)$, where

$$W_2^2(F_0, F_1) = \inf_{\gamma \in \mathcal{C}(F_0, F_1)} \int_{I\!\!R^d \times I\!\!R^d} \frac{1}{2}|v - w|^2 \gamma(\mathrm{d}v, \mathrm{d}w) \ . \tag{1.1}$$

Here, $\mathcal{C}(F_0, F_1)$ consists of all *couplings* of F_0 and F_1; i.e., all probability measures γ on $I\!\!R^d \times I\!\!R^d$ such that for all test functions η on $I\!\!R^d$

$$\int_{I\!\!R^d \times I\!\!R^d} \eta(v)\gamma(\mathrm{d}v, \mathrm{d}w) = \int_{I\!\!R^d} \eta(v)F_0(v)\mathrm{d}v$$

and

$$\int_{I\!\!R^d \times I\!\!R^d} \eta(w)\gamma(\mathrm{d}v, \mathrm{d}w) = \int_{I\!\!R^d} \eta(w)F_1(w)\mathrm{d}v \ .$$

The infimum in (1.1) is actually a minimum, and it is attained at a unique point γ_{F_0, F_1} in $\mathcal{C}(F_0, F_1)$. Brenier [3] was able to characterize this unique minimizer, and then further results of Caffarelli [4], Gangbo [10] and McCann [16] have shed considerable light on the nature of this minimizer.

Next, let the entropy $S(F)$ be defined by

$$S(F) = \int_{I\!\!R^d} F(v) \ln F(v)\mathrm{d}v \ . \tag{1.2}$$

This is well defined, with ∞ as a possible value, since $\int_{I\!\!R^d} |v|^2 F(v)\mathrm{d}v < \infty$.

The following scheme for solving the linear heat equation was in introduced in [12]: Fix an initial density F_0 with $\int_{I\!\!R^d} |v|^2 F_0(v)\mathrm{d}v$ finite, and also fix a time step $h > 0$. Then inductively define F_k in terms of F_{k-1} by choosing F_k to minimize the functional

$$F \to \left[W_2^2(F_{k-1}, F) + hS(F) \right] \tag{1.3}$$

on $\mathcal{P}$. It is shown in [12] that there is a unique minimizer $F_k \in \mathcal{P}$, so that each F_k is well defined. Then the time dependent probability density $F^{(h)}(v, t)$ is defined by putting

$F^{(h)}(v, kh) = F_k$ and interpolating when t is not an integer multiple of h. Finally, it is shown that for each t $F(\cdot, t) = \lim_{h \to 0} F^{(h)}(\cdot, t)$ exists weakly in L^1, and that the resulting time dependent probability density solves the heat equation $\partial/\partial t F(v, t) = \Delta F(v, t)$ with $\lim_{t \to 0} F(\cdot, t) = F_0$.

This variational approach is particularly useful when the functional being minimized with each time step is convex in the geometry associated to the 2-Wasserstein metric. It makes sense to speak of convexity in this context since, as McCann showed [16], when $\mathcal{P}$ is equipped with the 2–Wasserstein metric, every pair of elements F_0 and F_1 is connected by a unique continuous path $t \mapsto F_t$, $0 \le t \le 1$ such that $W_2(F_0, F_t) + W_2(F_t, F_1) = W_2(F_0, F_1)$ for all such t. It is natural to refer to this path as the geodesic connecting F_0 and F_1, and we shall do so. A functional Φ on $\mathcal{P}$ is *displacement convex* in McCann's sense if $t \mapsto \Phi(F_t)$ is convex on $[0, 1]$ for every F_0 and F_1 in $\mathcal{P}$. It turns out that the entropy $S(F)$ is a convex function of F in this sense.

Gradient flows of convex functions in Euclidean space are well known to have strong contractive properties, and Otto [18] showed that the same is true in $\mathcal{P}$, and applied this to obtain strong new results on rate of relaxation of certain solutions of the porous medium equation.

Our aim is to extend this line of analysis to a range of problems that are not purely dissipative, but which also satisfy certain *conservation laws*. An important example of such an evolution is given by the Boltzmann equation

$$\frac{\partial}{\partial t} f(x, v, t) + \nabla_x \cdot (v f(x, v, t)) = \mathcal{Q}(f)(x, v, t)$$

where for each t, $f(\cdot, \cdot, t)$ is a probability density on the phase space $\Lambda \times I\!\!R^d$ of a molecule in a region $\Lambda \subset I\!\!R^d$, and $\mathcal{Q}$ is a nonlinear operator representing the effects of collisions to the evolution of molecular velocities. This evolution is dissipative and decreases the entropy while formally conserving the energy $\int_{\Lambda \times I\!\!R^d} |v|^2 f(x, v, t) \mathrm{d}x \mathrm{d}v$ and the momentum $\int_{\Lambda \times I\!\!R^d} v f(x, v, t) \mathrm{d}x \mathrm{d}v$. A good deal is known about this equation [7], but there is not yet an existence theorem for solutions that conserve the energy, nor is there any general uniqueness result.

The investigation in this paper arose in the study of a related equation, the non–linear kinetic Fokker–Planck equation to which we have applied an analog of the scheme in [12] to the evolution of the conditional probability densities $F(v; x)$ for the velocities of the molecules at x; i.e., for the contributions of the collisions to the evolution of the distribution of velocities of particles in a gas. These collisions are supposed to conserve both the "bulk velocity" u and "temperature" θ, of the distribution where

$$u(F) = \int_{I\!\!R^d} v F(v) \mathrm{d}v \qquad \text{and} \qquad \theta(F) = \frac{1}{d} \int_{I\!\!R^d} |v|^2 F(v) \mathrm{d}v, \qquad (1.4)$$

and for this reason we add a constraint to the variational problem in [12]. Let $u \in I\!\!R^d$ and $\theta > 0$ be given. Define the subset $\mathcal{E}_{u,\theta}$ of $\mathcal{P}$ specified by

$$\mathcal{E}_{u,\theta} = \left\{ F \in \mathcal{P} \;\middle|\; \frac{1}{d} \int_{I\!\!R^d} |v - u|^2 F(v) \mathrm{d}v = \theta \quad \text{and} \quad \int_{I\!\!R^d} v F(v) \mathrm{d}v = u \right\} \qquad (1.5)$$

This is the set of all probability densities with a mean u and a variance $d\theta$, and we use $\mathcal{E}$ to denote it because the constraint on the variance is interpreted as an internal energy constraint in the context discussed above.

Then given $F_0 \in \mathcal{E}_{u,\theta}$, define the functional $I(F)$ on $\mathcal{E}_{u,\theta}$ by

$$I(F) = \left[\frac{W_2^2(F_0, F)}{\theta} + hS(F) \right] \tag{1.6}$$

Our main goal is to study the minimization problem associated with determining

$$\inf \left\{ I(F) \mid F \in \mathcal{E}_{u,\theta} \right\} . \tag{1.7}$$

Note that this problem is scale invariant in that if F_0 is rescaled, the minimizer F will be rescaled in the same way, and in any case, this normalization, with θ in the denominator, is dimensionally natural.

Since the constraint is *not weakly closed*, existence of minimizers does not follow as easily as in the unconstrained case. The same difficulty arises in the determination of the geodesics in $\mathcal{E}_{u,\theta}$.

We build on previous work on the geometry of $\mathcal{P}$ in the 2–Wasserstein metric, and section 2 contains a brief exposition of the relevant results. While this section is largely review, several of the simple proofs given here do not seem to be in the literature, and are more readily adapted to the constrained setting.

In section 3, we analyze the geometry of $\mathcal{E}$, and determine its geodesics. As mentioned above, since $\mathcal{E}$ is not weakly closed, direct methods do not yield the geodesics. The characterization of the geodesics is quite explicit, and from it we deduce a criterion for convexity in $\mathcal{E}$, and show that the entropy is uniformly strictly convex, in contrast with the unconstrained case.

In section 4, we turn to the variational problem (1.7), and determine the Euler–Lagrange equation associated with it, and several consequences of the Euler–Lagrange equation.

In section 5 we introduce variational problem that is dual to (1.7), and by analyzing it, we produce a minimizer for $I(F)$. We conclude the paper in section 6 by discussing some open problems and possible applications.

We would like to thank Robert McCann and Cedric Villani for many enlightening discussions on the subject of mass transport. We would also like to thank the referee, whose questions and suggestions have lead us to significantly clarify the exposition.

2: Riemannian Geometry of the 2–Wasserstein Metric

The purpose of this section is to collect a number of facts concerning the 2–Wasserstein metric and its associated Riemannian geometry. The Riemannian point of view has been developed by several authors, prominently including McCann, Otto, and Villani. Though for the most part the facts presented in this section are known, there is no single convenient reference for all of them. Moreover, it seems that some of the proofs and formulae that we use do not appear elsewhere in the literature.

We begin by recalling the identification of the geodesics in $\mathcal{P}$ equipped with the 2–Wasserstein metric. The fundamental facts from which we start are these: The infimum in (1.1) is actually a minimum, and it is attained at a unique point γ_{F_0,F_1} in $\mathcal{C}(F_0, F_1)$, and this measure is such that there exist a pair of dual convex functions ϕ and ψ such that for all bounded measurable functions η on $I\!R^d \times I\!R^d$,

$$\int_{I\!R^d \times I\!R^d} \eta(v, w) \gamma_{F_0,F_1}(\mathrm{d}v, \mathrm{d}w) = \int_{I\!R^d} \eta(v, \nabla\phi(v)) F_0 \mathrm{d}v = \int_{I\!R^d} \eta(\nabla\psi(w), w) F_1 \mathrm{d}w . \quad (2.1)$$

In particular, for all bounded measurable functions η on $I\!R^d$,

$$\int_{I\!R^d} \eta(\nabla\phi(v)) F_0 \mathrm{d}v = \int_{I\!R^d} \eta(w) F_1 \mathrm{d}w , \quad (2.2)$$

and $\nabla\phi$ is the *unique* gradient of a convex function defined on the convex hull of the support of F_0 so that (2.2) holds for all such η.

Recall that for any convex function ψ on $I\!R^d$, ψ^* denotes its Legendre transform; i.e., the dual convex function, which is defined through

$$\psi^*(w) = \sup_{v \in I\!R^d} \{ w \cdot v - \psi(v) \} . \quad (2.3)$$

The convex functions ψ arising as optimizers in (2.1) have the further property that $(\psi^*)^* = \psi$. Being convex, both ψ and ψ^* are locally Lipschitz and differentiable on the complement of a set of Haussdorff dimension $d - 1$. (It is for this reason that we work with densities instead of measures; $\nabla\psi\#\mu$ might not be well defined if μ charged sets Haussdorff dimension $d - 1$.) In our quotation of Brenier's result concerning in (2.1), the statement that the convex functions ψ and ϕ in (2.1) are a dual pair simply means that $\phi = \psi^*$ and $\psi = \phi^*$. It follows from (2.3) that $\nabla\psi$ and $\nabla\psi^*$ are inverse transormations in that

$$\nabla\psi(\nabla\psi^*(w)) = w \qquad \text{and} \qquad \nabla\psi^*(\nabla\psi(v)) = v \quad (2.4)$$

for $F_1(w)\mathrm{d}w$ almost every w and $F_0(v)\mathrm{d}v$ almost every v respectively.

Given a map $T : I\!R^d \to I\!R^d$ and $F \in \mathcal{P}$, define $T\#F \in \mathcal{P}$ by

$$\int_{I\!R^d} \eta(v) \left(T\#F(v) \right) \mathrm{d}v = \int_{I\!R^d} \eta(T(v)) F(v) \mathrm{d}v$$

for all test functions η on $I\!R^d$. Then we can express (2.2) more briefly by writing $\nabla\phi\#F_0 = F_1$. The uniqueness of the gradient of the convex potential ϕ is very useful for computing $W_2^2(F_0, F_1)$ since if one can find *some* convex function $\tilde{\phi}$ such that $\nabla\tilde{\phi}\#F_0 = F_1$, then $\nabla\tilde{\phi}$ is the potential for the minimizing map and

$$W_2^2(F_0, F_1) = \int_{I\!R^d} \frac{1}{2}|v - \nabla\tilde{\phi}(v)|^2 F_0(v)\mathrm{d}v . \quad (2.5)$$

Now it is easy to determine the Geodesics. These are given in terms of a natural interpolation between two densities F_0 and F_1 that was introduced and applied by McCann in his thesis [15] and in [16].

Fix two densities F_0 and F_1 in $\mathcal{P}$. Let ψ be the convex function on $I\!\!R^d$ such that $(\nabla\psi)\#F_0 = F_1$. Then for any t with $0 < t < 1$, define the convex function ψ_t by

$$\psi_t(v) = (1-t)\frac{|v|^2}{2} + t\psi(v) \tag{2.6}$$

and define the density F_t by

$$F_t = \nabla\psi_t \# F_0 \ . \tag{2.7}$$

At $t = 0$, $\nabla\psi_t$ is the identity, while at $t = 1$, it is $\nabla\psi$.

Cleary for each $0 \leq t \leq 1$, ψ_t is convex, and so the map $\nabla\psi_t$ gives the optimal transport from F_0 to F_t. What map gives the optimal transport from F_t on to F_1?

By definition $\nabla\psi_t\#F_0 = F_t$. It follows from (2.4) that $\nabla(\psi_t)^*\#F_t = F_0$, and therefore that $\nabla\psi \circ \nabla(\psi_t)^*\#F_t = F_1$. It turns out that $\nabla\psi \circ \nabla(\psi_t)^*$ is the *optimal* transport from F_t on to F_1. This composition property of the optimal transport maps along a McCann interpolation path provides the key to several of the theorems in the next section, and is the basis of short proofs of other known results. It is the essential observation made in this section.

To see that $\nabla\psi \circ \nabla(\psi_t)^*$ is the optimal transport map from F_t on to F_1, it suffices to show that it is a convex function. From (2.6), $\nabla\psi_t(v) = (1-t)v + t\nabla\psi(v)$, which is the same as $t\nabla\psi(v) = (\nabla\psi_t(v) - (1-t)v)$. Then using (2.4),

$$\nabla\psi \circ \nabla(\psi_t)^*(w) = \frac{1}{t}\left(w - (1-t)\nabla(\psi_t)^*(w)\right) \ . \tag{2.8}$$

Thus, $\nabla\psi \circ \nabla(\psi_t)^*(w)$ is a gradient. There are at least two ways to proceed from here. Assuming sufficient regularity of ψ and ψ^*, one can differentiate (2.4) and see that $\mathrm{Hess}\psi(\nabla\psi^*(w))\mathrm{Hess}\psi^*(w) = I$. That is, the Hessians of ψ and ψ^* are inverse to one another. Since $\mathrm{Hess}\psi_t(v) \geq (1-t)I$, this provides an upper bound on the Hessian of $(\psi_t)^*$ which can be used to show that the right side of (2.8) is the gradient of a convex function. This can be made rigorous in our setting, but the argument is somewhat technical, and involves the definition of the Hessian in the sense of Alexandroff.

There is a much simpler way to proceed. As McCann showed [15], if $\tilde{F}_t$ is the path one gets interpolating between F_0 and F_1 but starting at F_1, then $F_t = \tilde{F}_{1-t}$. So $\nabla\left((\psi^*)_{1-t}\right)^*$ is the optimal transport map from F_t on to F_1. This tells us which convex function should have $\nabla\psi \circ \nabla(\psi_t)^*(w)$ as its gradient, and this is easily checked using the mini–max theorem.

Lemma 2.1 (Interpolation and Legendre Transforms) *Let ψ be a convex function such that $\psi = \psi^{**}$. Then using the interpolation in (2.6),*

$$\left((\psi^*)_{1-t}\right)^*(w) = \frac{1}{t}\left(\frac{|w|^2}{2} - (1-t)(\psi_t)^*(w)\right) \ . \tag{2.9}$$

Proof: Calculating, with use of the the mini–max theorem, one has

$$
((\psi^*)_{1-t})^* (w) = \sup_z \left\{ z \cdot w - \left(t\frac{|z|^2}{2} + (1-t)\psi^*(z) \right) \right\}
$$

$$
= \sup_z \left\{ z \cdot w - t\frac{|z|^2}{2} - (1-t) \sup_v \{v \cdot z - \psi(v)\} \right\}
$$

$$
= \sup_z \inf_v \left\{ z \cdot (w - (1-t)v) - t\frac{|z|^2}{2} + (1-t)\psi(v) \right\}
$$

$$
= \inf_v \sup_z \left\{ z \cdot (w - (1-t)v) - t\frac{|z|^2}{2} + (1-t)\psi(v) \right\}
$$

$$
= \frac{1}{t} \left(\frac{|w|^2}{2} - (1-t)(\psi_t)^*(w) \right) \ .
$$

∎

As an immediate consequence,

$$
\nabla \left((\psi^*)_{1-t} \right)^* = \nabla\psi \circ \nabla(\psi_t)^* \tag{2.10}
$$

is the optimal transport from F_t to F_1. This also implies that $\nabla\psi_t \# F_0 = \nabla(\psi^*)_{1-t} \# F_1$, as shown by McCann in [15] using a "cyclic monotonicity" argument. Lemma 2.1 leads to a simple proof of another result of McCann, again from [15]:

Theorem 2.2 (Geodesics for the 2–Wasserstein Metric): *Fix two densities F_0 and F_1 in $\mathcal{P}$. Let ψ be the convex function on $\mathrm{I\!R}^d$ such that $(\nabla\psi)\#F_0 = F_1$. Then for any t with $0 < t < 1$, define the convex function ψ_t by (2.6) and define the density F_t by (2.7). Then for all $0 < t < 1$,*

$$
W_2(F_0, F_t) = tW_2(F_0, F_1) \qquad \text{and} \qquad W_2(F_t, F_1) = (1-t)W_2(F_0, F_1) \tag{2.11}
$$

and $t \mapsto F_t$ is the unique path from F_0 to F_1 for the 2–Wasserstein metric that has this property. In particular, there is exactly one geodesic for the 2–Wasserstein metric connecting any two densities in $\mathcal{P}$

Proof: It follows from (2.5) that

$$
W_2^2(F_0, F_t) = \frac{1}{2} \int_{\mathrm{I\!R}^d} |v - ((1-t)v + t\nabla\psi(v))|^2 \, F_0(v)\mathrm{d}v
$$

$$
= t^2 \frac{1}{2} \int_{\mathrm{I\!R}^d} |v - \nabla\psi(v)|^2 F_0(v)\mathrm{d}v = t^2 W_2^2(F_0, F_1) \ .
$$

Next, since $\nabla\left((\psi^*)_{1-t}\right)^*$ is the optimal transport from F_t to F_1, using (2.9),

$$W_2^2(F_t, F_1) = \frac{1}{2}\int_{\mathbb{R}^d}\left|w - \frac{1}{t}\left(w - (1-t)\nabla(\psi_t)^*(w)\right)\right|^2 F_t(v)\mathrm{d}v$$

$$= \left(\frac{1-t}{t}\right)^2\frac{1}{2}\int_{\mathbb{R}^d}|v - \nabla\psi_t(v)|^2 F_0(v)\mathrm{d}v = (1-t)^2 W_2^2(F_0, F_1) \ .$$

Together, the last two computations give us (2.11).

The uniqueness follows from a strict convexity property of the distance: For any probability density G_0, the function $G \mapsto W_2^2(G_0, G)$ is strictly convex on $\mathcal{P}$ in that for any pair G_1, G_2 in $\mathcal{P}$ and any t with $0 < t < 1$,

$$W_2^2(G_0, (1-t)G_1 + tG_2) \leq (1-t)W_2^2(G_0, G_1) + tW_2^2(G_0, G_2) \tag{2.12}$$

and there is equality if and only if $G_1 = G_2$. This follows easily from the uniqueness of the optimal coupling specified in (2.1); non–trivial convex combinations of such couplings are not of the form (2.1), and therefore cannot be optimal.

Now suppose that there were two geodesics $t \mapsto F_t$ and $t \mapsto \tilde{F}_t$. Pick some t_0 with $F_{t_0} \neq \tilde{F}_{t_0}$. Then the path consisting of a geodesic from F_0 to $(F_{t_0} + \tilde{F}_{t_0})/2$, and from there on to F_1 would have a strictly shorter length than the geodesic from F_0 to F_1, which cannot be. ∎

To obtain an Eulerian description of these geodesics, let f be any smooth function on $\mathbb{R}^d$, and compute:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{R^d} f(v)F_t(v)\mathrm{d}v = \frac{\mathrm{d}}{\mathrm{d}t}\int_{R^d} f(\nabla\psi_t(v))F_0(v)\mathrm{d}v$$

$$= \int_{R^d}\nabla f(\nabla\psi_t(v))\left[v - \nabla\psi(v)\right]F_0(v)\mathrm{d}v$$

$$= \int_{R^d}\nabla f(w)\left[\nabla(\psi_t)^*(w) - \nabla\psi(\nabla(\psi_t)^*(w))\right]F_t(w)\mathrm{d}w \tag{2.13}$$

$$= \int_{R^d}\nabla f(w)\left[\frac{w - \nabla(\psi_t)^*(w)}{t}\right]F_t(w)\mathrm{d}w$$

In other words, when F_t is defined in terms of F_0 and ψ as in (2.6) and (2.7), F_t is a weak solution to

$$\frac{\partial}{\partial t}F_t(w) + \nabla\cdot(W(w,t)F_t(w)) = 0 \tag{2.14}$$

where, according to Lemma 2.1,

$$W(w,t) = \frac{w - \nabla(\psi_t)^*(w)}{t} = \nabla\left(\frac{|w|^2}{2t} - \frac{1}{t}(\psi_t)^*(w)\right) \ . \tag{2.15}$$

In light of the first two equalities in (2.13),

$$W(w, 0) = \nabla \left(\frac{|w|^2}{2} - \psi(w) \right) = w - \nabla \psi(w) \ . \tag{2.16}$$

This gradient vector field can be viewed as giving the "tangent direction" to the geodesic $t \mapsto F_t$ at $t = 0$.

We would like to identify some subspace of the space of gradient vector fields as the tangent space T_{F_0} to $\mathcal{P}$ at F_0. Towards this end we ask: Given a smooth, rapidly decaying function η on $I\!R^d$, is there a geodesic $t \mapsto F_t$ passing through F_0 at $t = 0$ so that, in the weak sense,

$$\left. \left(\frac{\partial}{\partial t} F_t + \nabla \cdot (\nabla \eta F_t) \right) \right|_{t=0} = 0 \ . \tag{2.17}$$

The next theorem says that this is the case, and provides us with a geodesic that (2.17) holds with η sufficiently small. But then by changing the time parameterization, we obtain a geodesic, possibly quite short, that has any multiple of $\nabla \eta$ as its initial "tangent vector".

Theorem 2.3: (Tangents to Geodesics): *Let η be any smooth, rapidly decaying function η on $I\!R^d$ such that for all v,*

$$\psi(v) = \frac{|v|^2}{2} + \eta(v) \tag{2.18}$$

is strictly convex. For any density F_0 in $\mathcal{P}$, and t with $0 \le t \le 1$, define

$$\nabla \psi_t(v) = (1 - t)v + t \nabla \psi(v) = v + t \nabla \eta(v) \ . \tag{2.19}$$

Then for all t with $0 \le t \le 1$, $F_t = \nabla \psi_t \# F_0$ is absolutely continuous, and is a weak solution of

$$\frac{\partial}{\partial t} F_t(v) + \nabla \cdot (\nabla \eta_t(v) F_t(v)) = 0 \ , \tag{2.20}$$

where

$$\eta_t(v) = \frac{1}{t} \left(\frac{|v|^2}{2} - (\psi_t)^*(v) \right) \tag{2.21}$$

Moreover,

$$\nabla \eta_t(v) = \nabla \eta(v) - \frac{t}{2} \nabla |\nabla \eta(v)|^2 + t^2 \nabla R_t(v) \ , \tag{2.22}$$

where the remainder term $\nabla R_t(v)$ satisfies $\|\nabla R_t\|_\infty \le \|\text{Hess}(\eta)\|_\infty^2$ uniformly in t.

Proof: First, the fact that $\nabla \psi_t \# F_0$ is absolutely continuous follows from the fact that $\nabla(\psi_t)^*$ is Lipschitz. Formulas (2.20) and (2.21) follow directly from (2.14) and (2.15).

To obtain (2.22), use (2.4) to see that $\nabla(\psi_t)^*(v) = \Phi(\nabla(\psi_t)^*(v))$ where $\Phi(w) = v - t \nabla \eta(w)$. Iterating this fixed point equation three times yields (2.22). ■

In light of Theorems 2.2 and 2.3, we now know that every geodesic $t \mapsto F_t$ through F_0 at $t = 0$ satisfies (2.17), and conversely, for every smooth rapidly decaying gradient vector field, there is a geodesic $t \mapsto F_t$ through F_0 at $t = 0$ satisfying (2.17) for that function η. Moreover, along this geodesic

$$W_2^2(F_0, F_t) = \int_0^t \left(\int_{\mathbb{R}^d} |\nabla \eta_s(v)|^2 F_s(v) \mathrm{d}v \right) \mathrm{d}s = t \int_{\mathbb{R}^d} |\nabla \eta(v)|^2 F_0(v) \mathrm{d}v \ , \qquad (2.23)$$

where η_s is related to η as in Theorem 2.3.

Furthermore if $t \mapsto F_t$ is a path in $\mathcal{P}$ satisfying (2.17) for some gradient vector field $\nabla \eta$, then this vector field is unique. For suppose that $t \mapsto F_t$ also satisfies

$$\left(\frac{\partial}{\partial t} F_t + \nabla \cdot (\nabla \xi F_t) \right) \bigg|_{t=0} = 0 \ . \qquad (2.24)$$

Then, $\nabla \cdot (\nabla (\eta - \xi) F_0) = 0$. Integrating against $\eta - \xi$, we obtain that

$$\int_{\mathbb{R}^d} |\nabla \eta - \nabla \xi|^2 F_0(v) \mathrm{d}v = 0 \ .$$

Careful consideration of this well known argument, inserting a cut off function before integrating by parts, reveals that all it requires is that both $\nabla \eta$ and $\nabla \xi$ are square integrable with respect to F_0. This justifies the indentification of the tangent vector $\partial F / \partial t$ with $\nabla \eta$ when (2.17) holds and $\nabla \eta$ is square integrable with respect to F_0.

This identifies the "tangent vector" $\partial F_t / \partial t$ with $\nabla \eta$, and gives us the Riemannian metric, first introduced by Otto [18],

$$g \left(\frac{\partial F}{\partial t}, \frac{\partial F}{\partial t} \right) = \frac{1}{2} \int |\nabla \eta(v)|^2 F_0(v) \mathrm{d}v \ . \qquad (2.25)$$

By (2.23), the distance on $\mathcal{P}$ induced by this metric is the 2–Wasserstein distance.

Interestingly, Theorem 2.2 provides a global description of the geodesics without having to first determine and study the Riemannian metric. Theorem 2.3 gives an Eulerian characterization of the geodesics which provides a complement to McCann's original Lagrangian characterization. Another Eulerian analysis of the geodesics in terms of the Hamilton–Jacobi equation seems to be folklore in the subject. A clear account can be found in recent lecture notes of Villani [22].

We now turn to the notion of convexity on $\mathcal{P}$ with respect to the 2–Wasserstein metric. A functional Φ on $\mathcal{P}$ is said to be *displacement convex at F_0* in case $t \mapsto \Phi(F_t)$ is convex on some neighborhood of 0 for all geodesics $t \mapsto F_t$ passing through F_0 at $t = 0$. A functional Φ on $\mathcal{P}$ is said to be *displacement convex* if it is displacement convex at all points F_0 of $\mathcal{P}$.

If moreover $t \mapsto \Phi(F_t)$ is twice differentiable, we can check for displacement convexity by computing the Hessian:

$$\mathrm{Hess}\Phi(F_0)\langle \nabla \eta, \nabla \eta \rangle = \frac{\mathrm{d}^2}{\mathrm{d}t^2} \Phi(F_t) \bigg|_{t=0} \ , \qquad (2.26)$$

where $\nabla\eta$ is the tangent to the geodesic at $t = 0$.

Theorem 2.4 (Displacement Convexity): *If the functional Φ on $\mathcal{P}$ is given by*

$$\Phi(F) = \int_{\mathbb{R}^d} g(F(v)) \mathrm{d}v \qquad (2.27)$$

where g is a twice differentiable convex function on $\mathbb{R}_+$, then Φ is displacement convex if

$$tg'(t) - g(t) \geq 0 \qquad \text{and} \qquad t^2 g''(t) - tg'(t) + g(t) \geq 0 \qquad (2.28)$$

for all $t > 0$, where the primes denote derivatives.

Proof: We check for convexity at a density F_0 in the domain of Φ. By a standard mollification, we can find a sequence of smooth densities $F_0^{(n)}$ with $\lim_{n\to\infty} F_0^{(n)} = F_0$ and $\lim_{n\to\infty} \Phi(F_0^{(n)}) = \Phi(F_0)$. Fix any smooth rapidly decaying function η, such that (taking a small multiple if need be) $|v|^2 + \eta(v)$ is strictly convex. Then with $\nabla\psi_t$ defined as in (2.19),

$$t \mapsto \nabla\psi_t \# F_0^{(n)} = F_t^{(n)}$$

gives a geodesic passing through $F_0^{(n)}$ at $t = 0$ with the tangent direction $\nabla\eta$, and defined for $0 \leq t \leq 1$ uniformly in n. Also, $\lim_{n\to\infty} \Phi(F_t^{(n)}) = \Phi(F_t)$ for all such t. Therefore, it suffices to show that for each n, $t \mapsto \Phi(F_t^{(n)})$ is convex. In other words, we may assume that F_0 is smooth. Then so is each F_t, since $F_t(w) = F_0(\nabla(\psi_t)^*(w))\det\left(\mathrm{Hess}(\psi_t)^*\right)(w))$ is a composition of smooth functions. We may now check convexity by differentiating.

By (2.20),

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} g(F_t(v)) \mathrm{d}v = -\int_{\mathbb{R}^d} g'(F_t(v)) \nabla \cdot (\nabla\eta_t(v) F_t(v)) \, \mathrm{d}v$$

$$= \int_{\mathbb{R}^d} (g''(F_t(v)) \nabla F_t(v)) \cdot (\nabla\eta_t(v) F_t(v)) \, \mathrm{d}v$$

Defining $h(t) = tg'(t) - g(t)$ so that $h'(t) = tg''(t)$, one has from (2.20) that

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi(F_t) = \int_{\mathbb{R}^d} \nabla h(F_t(v)) \cdot \nabla\eta_t(v) \mathrm{d}v \ . \qquad (2.29)$$

To differentiate a second time, use (2.22) to obtain

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \Phi(F_t) \bigg|_{t=0} = \int_{\mathbb{R}^d} \nabla h(F_0) \cdot \nabla\left(-\frac{1}{2}|\nabla\eta|^2\right) \mathrm{d}v - \int_{\mathbb{R}^d} \frac{\partial}{\partial t} h(F_t) \bigg|_{t=0} (\Delta\eta) \, \mathrm{d}v \ .$$

But

$$\left.\frac{\partial}{\partial t} h(F_t)\right|_{t=0} = -F_0^2 g''(F_0) (\Delta \eta) - \nabla h(F_0) \cdot \nabla \eta$$

and hence

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2} \Phi(F_t)\right|_{t=0} =$$

$$\int_{\mathrm{I\!R}^d} \nabla h(F_0) \cdot \left(-\frac{1}{2} \nabla |\nabla \eta|^2 + (\Delta \eta) \nabla \eta\right) \mathrm{d}v + \int_{\mathrm{I\!R}^d} F_0^2 g''(F_0) (\Delta \eta)^2 \, \mathrm{d}v = \qquad (2.30)$$

$$\int_{\mathrm{I\!R}^d} h(F_0) \|\mathrm{Hess}\eta\|^2 \mathrm{d}v + \int_{\mathrm{I\!R}^d} \left(F_0^2 g''(F_0) - h(F_0)\right) (\Delta \eta)^2 \, \mathrm{d}v \ .$$

Here, $\|\mathrm{Hess}\eta\|^2$ denotes the square of the Hilbert-Schmidt norm of the Hessian of η. This quantity is positive whenever $h(F) = Fg'(F) - g(F)$ and $F^2 g''(F) - h(F) = F^2 g''(F) - Fg'(F) + g(F)$ are positive. ∎

The case of greatest interest here is the entropy functional $S(F)$, defined in (1.2). In this case, $g(t) = t \ln t$, so that $tg'(t) - g(t) = t$ and $tg''(t) - tg'(t) + g(t) = 0$. Hence from (2.30),

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2} S(F_t)\right|_{t=0} = \int_{\mathrm{I\!R}^d} \|\mathrm{Hess}\eta\|^2 F_0(v) \mathrm{d}v \ . \qquad (2.31)$$

This shows that the entropy is convex, as proved in [18], though not strictly convex. Consider the following example * in one dimension: Let

$$\psi(v) = \frac{|v|^2}{2} + |v|$$

For any F_0, define $F_t = \nabla \psi_t$ and then it is easy to see that

$$F_t(v) = 1_{\{v < -t\}} F_0(v + t) + 1_{\{v > t\}} F_0(v - t) \ . \qquad (2.32)$$

The geodesic $t \mapsto F_t$ can be continued indefinitely for positive t, but unless F_0 vanishes in some strip $-\epsilon < v < \epsilon$, it cannot be continued at all for negative t. Clearly, with F_t defined as in (2.32), $S(F_t) = S(F_0)$ for all t.

There are however interesting cases in which the entropy is strictly convex along a geodesic, and even uniformly so: Suppose that the "center of mass" $\int_{\mathrm{I\!R}^d} v F_t(v) \mathrm{d}v$ is constant along the geodesic $t \mapsto F_t$, which means that

$$\int_{\mathrm{I\!R}^d} \nabla \eta(v) F_0(v) \mathrm{d}v = 0 \qquad (2.33)$$

---

* We thank the referee for this example, which has clarified the formulation of Corollary 2.5 below.

where as above, $\nabla\eta$ is the tangent vector generating the geodesic.

The Poincare constant $\alpha(F)$ of a density F in $\mathcal{P}$ is defined by

$$\alpha(F) = \inf_{\varphi \in \mathcal{C}_0^\infty} \frac{\int |\nabla\varphi(v)|^2 F(v)\mathrm{d}v}{\int \left|\varphi(v) - \int \varphi(v)F(v)\mathrm{d}v\right|^2 F(v)\mathrm{d}v} \ . \tag{2.34}$$

Thus, when (2.33) holds, taking $\varphi = \partial\eta/\partial v_i$ for $i = 1\ldots d$ and then summing yields

$$\int_{\mathbb{R}^d} \|\mathrm{Hess}\eta\|^2 F_0(v)\mathrm{d}v \geq \alpha(F_0)\int_{\mathbb{R}^d} |\nabla\eta(v)|^2 F_0(v)\mathrm{d}v \ , \tag{2.35}$$

which provides a lower bound to the right side of (2.31) in terms of the Riemannian metric.

Now consider an "smooth" geodesic through a smooth density F_0, as in the previous proof, and such that (2.33) is satisfied. Then by (2.30) and (2.35), for any t and $h > 0$ such that F_{t-h} and F_{t+h} are both on the geodesic,

$$\frac{1}{h^2}\left(S(F_{t+h}) + S(F_{t-h}) - 2S(F_t)\right) \geq \alpha(F_t)\int_{\mathbb{R}^d} |\nabla\eta(v)|^2 F_0(v)\mathrm{d}v \ .$$

If the geodesic is parameterized by arclength, then the last factor on the right is one.

Summarizing the last paragraphs, we have the following corollary:

Corollary 2.5 (Strict Convexity of Entropy): *Consider a geodesic $s \mapsto F_s$ parameterized by arc length s, and defined for some interval $a < s < b$ such that $s \mapsto \int vF_s(v)\mathrm{d}v$ is constant, and such that each F_s is bounded and continuously differentiable. Then for all s and h so that $a < s + h, s - h < b$,*

$$S(F_{s+h}) + S(F_{s-h}) - 2S(F_s) \geq h^2\alpha(F_s) \ , \tag{2.36}$$

where $\alpha(F_s)$ is the Poincare constant of the density F_s.

(Notice that for the geodesic (2.32), $\alpha(F_t) = 0$ for all $t > 0$, as long as F_0 has positive mass on both sides of the origin, in addition to the fact that F_t will not in general be smooth).

We remark that Caffarelli has recently shown [6] that if F_0 is a Gaussian density, and $F_1 = e^{-V}F_0$ where V is convex, then there is an *upper* bound on the Hessian of the potential ψ for which $\nabla\psi\#F_0 = F_1$. This upper bound is inherited by ψ_t for all t. Since as Caffarelli shows, an upper bound on the Hessian of ψ and a lower bound on the Poincare constant for F_0 implies a lower bound on the Poincare constant of F_t, one obtains a *uniform* lower bound on the Poincare constant for F_t, $0 < t < 1$. Hence $S(F_t)$ is uniformly strictly convex along such a geodesic.

3: Geometry of the Constraint Manifold

Let $u \in \mathbb{R}^d$ and $\theta > 0$ be given. Consider the subset $\mathcal{E}_{u,\theta}$ of $\mathcal{P}$ specified by

$$\mathcal{E}_{u,\theta} = \left\{ F \in \mathcal{P} \ \middle| \ \frac{1}{d} \int_{\mathbb{R}^d} |v - u|^2 F(v) \mathrm{d}v = \theta \quad \text{and} \quad \int_{\mathbb{R}^d} v F(v) \mathrm{d}v = u \right\} \tag{3.1}$$

This is the set of all probability densities with a mean u and a variance $d\theta$. We will often write $\mathcal{E}$ in place of $\mathcal{E}_{u,\theta}$ when u and θ are clear from the context or simply irrelevant.

We give a fairly complete description of the geometry of $\mathcal{E}$, both locally and globally. In particular, we obtain a closed form expression for the distance between any two points on $\mathcal{E}$ in the metric induced by the 2–Wasserstein metric, and a global description of the geodesics in $\mathcal{E}$.

Notice that

$$\mathcal{E}_{u,\theta} \subset \left\{ F \mid W_2^2(F, \delta_u) = \frac{d\theta}{2} \right\} \tag{3.2}$$

where δ_u is the unit mass at u. This is quite clear from the transport point of view: If our target distribution is a point mass, there are no choices to make; everything is simply transported to the point u. Hence $\mathcal{E}_{u,\theta}$ is a part of a sphere in the 2-Wasserstein metric, centered on δ_u, and with a radius of $\sqrt{d\theta/2}$.

Our first theorem shows that for any F_0 in $\mathcal{P}$, there is a unique closest F in $\mathcal{E}$, and this is obtained by dilatation and translation. This is the first of two related variational problems solved in this section.

Theorem 3.1 (Projection onto $\mathcal{E}$): *Let F_0 be any probability density on $\mathbb{R}^d$ such that*

$$\int_{\mathbb{R}^d} v F_0(v) \mathrm{d}v = u_0 \qquad \text{and} \qquad \int_{\mathbb{R}^d} |v - u_0|^2 F_0(v) \mathrm{d}v = d\theta_0 \ .$$

Let $\theta > 0$ and u be given, and set $a = \sqrt{\theta_0/\theta}$. Then

$$\inf \left\{ W_2^2(G, F_0) \mid G \in \mathcal{E}_{\theta,u} \right\}$$

is attained at

$$\tilde{F}(v) = a^d F_0 \left(a(v - u) + u_0 \right) \ ,$$

and the minimum value is

$$W_2^2(F_0, \tilde{F}) = \frac{\left(\sqrt{\theta} - \sqrt{\theta_0} \right)^2}{2} + \frac{|u - u_0|^2}{2} \ . \tag{3.3}$$

Proof: There is no loss of generality in fixing $u = 0$ in the proof since if u_0 is arbitrary, a translation of both $\tilde{F}$ and F_0 yields the general result.

Let ϕ be defined by $\phi(v) = |v - u_0|^2/(2a)$ so that $(\nabla\phi)\#F_0 = \tilde{F}$. Let $\psi(w) = a|w|^2/2 + w \cdot u_0$ be the dual convex function so that

$$\phi(v) + \psi(w) \geq v \cdot w \ ,$$

and hence

$$\frac{1}{2}|v - w|^2 \geq \frac{a|v|^2 - |v - u_0|^2}{2a} + \frac{(1-a)|w|^2 - w \cdot u_0}{2} \tag{3.4}$$

for all v and w.

Next, given any G in $\mathcal{E}$, let γ be the optimal coupling of F_0 and G so that

$$W_2^2(F_0, G) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}|v - w|^2 \gamma(\mathrm{d}v, \mathrm{d}w) \ .$$

Then by (3.4),

$$W_2^2(F_0, G) \geq \left(\frac{a-1}{2a}\right) \int_{\mathbb{R}^d} |v|^2 F_0(v)\mathrm{d}v + \frac{|u_0|^2}{2} + \left(\frac{1-a}{2}\right) \int_{\mathbb{R}^d} |w|^2 G(w)\mathrm{d}w$$

$$= \frac{(a-1)^2 d\theta}{2} + \frac{|u_0|^2}{2} \ .$$

On the other hand, since $(\nabla\phi)\#F_0 = \tilde{F}$,

$$W_2^2(F_0, \tilde{F}) = \int_{\mathbb{R}^d} \frac{1}{2}|v - \nabla\phi(v)|^2 F_0(v)\mathrm{d}v$$

$$= \left(\frac{1}{a} - 1\right)^2 \int_{\mathbb{R}^d} \frac{1}{2}|v - u_0|^2 F_0(v)\mathrm{d}v + \frac{|u_0|^2}{2}$$

$$= \frac{(a-1)^2 d\theta}{2} + \frac{|u_0|^2}{2} \ .$$

∎

Remark: (Exact Solution for the JKO time discretization of the heat equation for Gaussian initial data) Theorem 3.1 allows us to exactly solve the Jordan–Kinderleher-Otto time discretization of the heat equation for Gaussian initial data. Take as initial data $F_0(v) = (4\pi t_0)^{-d/2} e^{-|v|^2/4t_0}$. We can now find $\inf\{W_2^2(F, F_0) + hS(F)\}$ in two steps. First, consider

$$\inf\{W_2^2(F, F_0) + hS(F) \mid F \in \mathcal{E}_{0,2td}\} \tag{3.5}$$

Now on $\mathcal{E}_{0,2td}$, S has a global minimum at $G_t = (4\pi t)^{-d/2}e^{-|v|^2/4t}$, as is well known. By Theorem 3.1, $W_2^2(F, F_0)$ also has a global minimum on on $\mathcal{E}_{0,2td}$ at G_t, since G_t is just a rescaling of F_0. Therefore, by (3.3), the infimum in (3.5) is

$$W_2^2(G_t, F_0) + hS(G_t) = d\left(\sqrt{t} - \sqrt{t_0}\right)^2 + -h\frac{d}{2}\left(\ln(4\pi t) + 1\right) .$$

In the second step, we simply compute the minimizing value of t, which amounts to finding the value of t that minimizes

$$\left(\sqrt{t} - \sqrt{t_0}\right)^2 - \frac{h}{2}\ln t .$$

Simple computations lead to the value $t = f(t_0)$ where

$$f(s) = \frac{1}{2}\left(s + h + s\sqrt{1 + \frac{2h}{s}}\right) . \tag{3.6}$$

Note that $t_0 < f(t_0) < t_0 + h$, but $f(t_0) = t_0 + h + O(h^2)$. If we then inductively define $t_n = f(t_{n-1})$, we see that the exact solution of the Jordan–Kinderleher-Otto time discretization of the heat equation is given at time step n by $F_n = (4\pi t_n)^{-d/2}e^{-|v|^2/4t_n}$ where $t_n = t_0 + nh + \mathcal{O}(h^2)$. Note that in the discrete time approximation, the variance increases more slowly than in continuous time, since the $\mathcal{O}(h^2)$ term is negative, though of course the difference in the rates vanishes as h tends to zero.

Returning to the main focus of this section, fix two densities F_0 and F_1 in $\mathcal{E}$. Let ψ be the convex function on $I\!\!R^d$ such that $(\nabla\psi)\#F_0 = F_1$. Then by Theorem 2.2, the geodesic that runs from F_0 to F_1 through the ambient space $\mathcal{P}$ is given by

$$F_t = ((1 - t)v + t\nabla\psi)\#F_0 .$$

Thinking of $\mathcal{E}$ as a subset of a sphere, and this geodesic as the chord connecting two points on the sphere, we refer to it as the *chordal geodesic* F_0 to F_1.

Lemma 3.2 (Variance Along a Chordal Geodesic): *Let F_0 and F_1 be any two densities in $\mathcal{E}$. Let $t \mapsto F_t$ be the chordal geodesic joining them. Then for all t with $0 \leq t \leq 1$,*

$$\frac{1}{2}\int_{I\!\!R^d}|v - u|^2 F_t(v)\mathrm{d}v = \frac{d\theta}{2}\left[1 - 4t(1 - t)\frac{W_2^2(F_0, F_1)}{2d\theta}\right]$$

$$= R_\theta^2\left[1 - t(1 - t)\frac{W_2^2(F_0, F_1)}{R_\theta^2}\right] \tag{3.7}$$

where $R_\theta = \sqrt{d\theta/2}$.

16

Proof: Notice first that with $F_1 = \nabla\psi\#F_0$, we have from Theorem 2.2 that

$$\int_{{\rm I\!R}^d} \frac{1}{2}|v-u|^2 F_t(v)\mathrm{d}v = \int_{{\rm I\!R}^d} \frac{1}{2}| \left((1-t)v + t\nabla\psi(v)\right) - u|^2 F_0(v)\mathrm{d}v$$

$$= \int_{{\rm I\!R}^d} \frac{1}{2}|(1-t)\,(v-u) + t\,(\nabla\psi(v) - u)\,|^2 F_0(v)\mathrm{d}v$$

$$= (1-t)^2 \int_{{\rm I\!R}^d} \frac{1}{2}|v-u|^2 F_0(v)\mathrm{d}v + t^2 \int_{{\rm I\!R}^d} \frac{1}{2}|w-u|^2 F_1(y)\mathrm{d}v$$

$$+ t(1-t) \int_{{\rm I\!R}^d} (v-u)\cdot(\nabla\psi(v) - u)\, F_0(v)\mathrm{d}v$$

$$= \frac{d\theta}{2}(1-t)^2 + \frac{d\theta}{2}t^2 + t(1-t) \int_{{\rm I\!R}^d} (v-u)\cdot(\nabla\psi(v) - u)\, F_0(v)\mathrm{d}v$$

$$\tag{3.8}$$

Next,

$$W_2^2(F_0, F_1) = \frac{1}{2} \int_{{\rm I\!R}^d} |v - \nabla\psi(v)|^2 F_0(v)\mathrm{d}v$$

$$= \frac{1}{2} \int_{{\rm I\!R}^d} |v-u|^2 F_0(v)\mathrm{d}v + \frac{1}{2} \int_{{\rm I\!R}^d} |\nabla\psi(v) - u|^2 F_0(v)\mathrm{d}v$$

$$- \int_{{\rm I\!R}^d} (v-u)\cdot(\nabla\psi(v) - u)\, F_0(v)\mathrm{d}v$$

$$= d\theta - \int_{{\rm I\!R}^d} (v-u)\cdot(\nabla\psi(v) - u)\, F_0(v)\mathrm{d}v$$

by the definition of $\mathcal{E}$, and hence

$$\int_{{\rm I\!R}^d} (v-u)\cdot(\nabla\psi(v) - u)\, F_0(v)\mathrm{d}v = d\theta - W_2^2(F_0, F_1)\;. \tag{3.9}$$

Combining (3.9) and (3.8), one has the result. ∎

We note that since $\int_{{\rm I\!R}^d}(v-u)F_0(v)\mathrm{d}v = 0$,

$$\int_{{\rm I\!R}^d} (v-u)\cdot(\nabla\psi(v) - u)\, F_0(v)\mathrm{d}v = \int_{{\rm I\!R}^d} (v-u)\cdot(\nabla\psi(v) - \nabla\psi(u))\, F_0(v)\mathrm{d}v \geq 0$$

by the convexity of ψ. It follows from this and (3.9) that

$$W_2^2(F_0, F_1) \leq d\theta = 2R_\theta^2\;, \tag{3.10}$$

where $R_\theta = \sqrt{d\theta/2}$ is the radius of $\mathcal{E}$ as in (3.2). Hence the variance in (3.7) is never smaller than R_θ^2.

The next result is the second of the variational problems solved in this section, and it the key to the determination of the geodesics in $\mathcal{E}$.

Theorem 3.3 (Midpoint Theorem): *Let F_0 and F_1 be any two densities in $\mathcal{E}$. Then*

$$\inf_{G \in \mathcal{E}} \left\{ W_2^2(F_0, G) + W_2^2(G, F_1) \right\} \tag{3.11}$$

is attained uniquely at $a^d F_{1/2}(a(v - u) + u)$ where $F_{1/2}$ is the midpoint of the chordal geodesic, and a is chosen to rescale the midpoint onto $\mathcal{E}$; i.e.,

$$a = \sqrt{1 - \frac{W_2^2(F_0, F_1)}{2d\theta}} = \sqrt{1 - \frac{W_2^2(F_0, F_1)}{(2R_\theta)^2}} \ , \tag{3.12}$$

where $R_\theta = \sqrt{d\theta/2}$ is the radius of $\mathcal{E}$ as in (3.2). Moreover, the minimal value attained in (3.11) is $f\left(W_2^2(F_0, F_1)\right)$ where

$$f(x) = 2d\theta \left(1 - \sqrt{1 - x/(2d\theta)} \right) \ . \tag{3.13}$$

The function f is convex and increasing on $[0, 2d\theta]$.

Before giving the proof itself, we first consider some formal arguments that serve to identify the minimizer and motivate the proof.

Let $\Phi(G)$ denote the functional being minimized in (3.11). This functional is strictly convex with respect to the usual convex structure on $\mathcal{E}$; that is, for all λ with $0 < \lambda < 1$, and all G_0 and G_1 in $\mathcal{E}$,

$$\Phi(\lambda G_0 + (1 - \lambda)G_1) \leq \lambda\Phi(G_0) + (1 - \lambda)\Phi(G_1)$$

with equality only if $G_0 = G_1$. The strict convexity suggests that there is a minimizer G_0, and that if we can find any critical point G of Φ, then G is the minimizer G_0.

To make variations in G, seeking a critical point, let η be a smooth, rapidly decaying function on $I\!\!R^d$, and define the map $T_t : I\!\!R^d \to I\!\!R^d$ by $T_t(v) = v + t\nabla\eta(v)$. Let $G_t = T_t \# G_0$. We want the curve $t \mapsto G_t$ to be tangent to $\mathcal{E}$ at $t = 0$, so we require in particular that

$$\int_{I\!\!R^d} v \cdot \nabla\eta(v)G_0(v)\mathrm{d}v = 0 \tag{3.14}$$

which guarantees that $\int |v|^2 G(t)\mathrm{d}v = \int |v|^2 G_0 \mathrm{d}v + \mathcal{O}(t^2)$.

Let ϕ be the convex function such that $\nabla\phi \# G_0 = F_0$, and let $\tilde{\phi}$ be the convex function such that $\nabla\tilde{\phi} \# G_0 = F_1$. The variation in $\Phi(G_t)$ can be expressed in terms of ϕ, $\tilde{\phi}$ and η as follows: Formally, assuming enough regularity,

$$\lim_{t \to 0^+} \frac{\Phi(G_t) - \Phi(G_0)}{t} = \int_{I\!\!R^d} \left(\nabla\phi(v) + \nabla\tilde{\phi}(v) - 2v \right) \cdot \nabla\eta(v)G_0(v)\mathrm{d}v \ . \tag{3.15}$$

18

(A more precise statement and explanation is provided in section 4 where we make actual use of such variations. For the present heuristic purposes it suffices to be formal.)

Combining (3.14) and (3.15), the formal condition for G_0 to be a critical point is seen to be that

$$\nabla\phi(v) + \nabla\tilde{\phi}(v) = Cv \tag{3.16}$$

for some constant C.

The formal argument tells us what to look for, namely a G_0 such that (3.16) holds. It is easy to see if G_0 is the midpoint of the chordal geodesic from F_0 to F_1 projected onto $\mathcal{E}$ by rescaling as in Theorem 3.1, then G_0 satisfies (3.16). The actual proof of the theorem consists of two steps: First we verify the assertion we just made about G_0 so defined. Then we prove, using (3.16) that G_0 is indeed the minimizer using a duality argument very much like the one used to prove Theorem 3.1.

Proof of Theorem 3.3: First, we may assume that $u = 0$. Next, let ψ be the convex function such that $\nabla\psi \# F_0 = F_1$. We may suppose initially that both F_0 and F_1 are strictly positive so that ψ will be convex on all of $I\!\!R^d$. Recall that $\nabla\left(\psi_{1/2}\right)^* \# F_{1/2} = F_0$, and that by (2.10), $\nabla\left((\psi^*)_{1/2}\right)^* \# F_{1/2} = F_1$. Then immediately from (2.9) we have

$$\left(\psi_{1/2}\right)^*(v) + \left((\psi^*)_{1/2}\right)^*(v) = |v|^2 \ . \tag{3.17}$$

Now let a be given by (3.12), and define

$$\phi(v) = \frac{1}{a}\left(\psi_{1/2}\right)^*(av) \qquad \text{and} \qquad \tilde{\phi} = \frac{1}{a}\left((\psi^*)_{1/2}\right)^*(av) \ .$$

Then, $\nabla\phi \# G_0 = F_0$ and $\nabla\tilde{\phi} \# G_0 = F_1$, and from (3.17),

$$\phi(v) + \tilde{\phi}(v) = a|v|^2 \ . \tag{3.18}$$

To use this observe that for any dual pair of convex functions η and η^*, Young's inequality say that $\eta(v) + \eta^*(w) \geq v \cdot w$. Hence for all v and w,

$$\frac{1}{2}|v - w|^2 \geq \frac{1}{2}|v|^2 + \frac{1}{2}|w|^2 - \eta(v) - \eta^*(w)$$

Now if G is any element of $\mathcal{E}$, and γ_0 is the optimal coupling between G and F_0, we have

$$W_2^2(G, F_0) = \int_{I\!\!R^d \times I\!\!R^d} \frac{1}{2}|v - w|^2 \gamma_0(\mathrm{d}v, \mathrm{d}w)$$

$$\geq d\theta - \int_{I\!\!R^d} \eta(v)G(v)\mathrm{d}v - \int_{I\!\!R^d} \eta^*(w)F_0(w)\mathrm{d}w \tag{3.19}$$

In the same way, we deduce that for any other dual pair of convex functions ζ and ζ^*,

$$
\begin{aligned}
W_2^2(G, F_1) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2} |v - w|^2 \gamma_1(\mathrm{d}v, \mathrm{d}w) \\
&\geq d\theta - \int_{\mathbb{R}^d} \zeta(v) G(v) \mathrm{d}v - \int_{\mathbb{R}^d} \zeta^*(w) F_1(w) \mathrm{d}w \ .
\end{aligned}
\tag{3.20}
$$

We now choose $\eta = \phi$ and $\zeta = \tilde{\phi}$. Then adding (3.19) and (3.20), and on account of (3.18),

$$
\begin{aligned}
\Phi(G) &= W_2^2(G, F_0) + W_2^2(G, F_1) \\
&\geq 2d\theta - \int_{\mathbb{R}^d} \left(\phi(v) + \tilde{\phi}(v) \right) G(v) \mathrm{d}v - \int_{\mathbb{R}^d} \phi^*(w) F_0(w) \mathrm{d}w - \int_{\mathbb{R}^d} \tilde{\phi}^*(w) F_1(v) \mathrm{d}w \\
&= (2 - a)\, d\theta - \int_{\mathbb{R}^d} \phi^*(w) F_0(w) \mathrm{d}w - \int_{\mathbb{R}^d} \tilde{\phi}^*(w) F_1(v) \mathrm{d}w \ .
\end{aligned}
\tag{3.21}
$$

Now suppose that $G = G_0$. Then for γ_0–almost every (v, w), we have that $v \cdot w = \phi(v) + \phi^*(w)$ so that

$$
\frac{1}{2} |v - w|^2 = \frac{1}{2} |v|^2 + \frac{1}{2} |w|^2 - \phi(v) - \phi^*(w)
$$

and hence there is equality in (3.19) when $G = G_0$ and $\eta = \phi$. In the same way, there is equality in (3.20) when $G = G_0$ and $\zeta = \tilde{\phi}$. Thus, the lower bound in (3.21) is saturated for $G = G_0$, and is in any case independent of G. This proves that G_0 is the minimizer.

It is now easy to compute the minimizing value. Theorem 3.1 tells us that that $G_0(v) = a^d F_{1/2}(av)$ where a depends only on $W_2^2(F_0, F_1)$, and is given explicitly by (3.12). Then, with this choice of a,

$$
\frac{1}{a} \nabla \psi_{1/2} \# F_0 = G_0 \ .
$$

Expressing this directly in terms of ψ and computing in the familiar way, one finds

$$
W_2^2(F_0, G_0) = \frac{d\theta}{a} \left[(a - 1) + \frac{W_2^2(F_0, F_1)}{2d\theta} \right] = d\theta(1 - a) \ .
\tag{3.22}
$$

Clearly, $W_2^2(F_0, G_0) = W_2^2(G_0, F_1)$, and so doubling the right hand side of (3.22) and inserting our formula for a, we obtain (3.13). Finally simple calculations confirm that f is increasing and convex on $[0, 1]$. $\blacksquare$

We are now prepared to consider discrete approximations to geodesics in $\mathcal{E}$. Let $\mathcal{G}$ be the set of continuous maps $t \mapsto G_t$ from $[0, 1]$ to $\mathcal{E}$ with $G_0 = F_0$ and $G_1 = F_1$.

For each natural number k, let $\mathcal{G}_k(F_0, F_1)$ denote the set of sequences

$$\{G_0, G_1, \ldots, G_{2^k}\} \tag{3.23}$$

where each G_j is in $\mathcal{E}$, $G_0 = F_0$, $G_{2^k} = F_1$, and finally

$$W_2^2(G_{j+2}, G_{j+1}) = W_2^2(G_{j+1}, G_j) \tag{3.24}$$

for all $j = 0, 1, \ldots, 2^k - 2$.

For any path $t \mapsto G_t$ in $\mathcal{G}$ and any k, we obtain a sequence in $\mathcal{G}_k(F_0, F_1)$ by an appropriate selection of times t_j and setting $G_j = G(t_j)$.

We next obtain a particular element $\{F_0^{(k)}, F_1^{(k)}, \ldots, F_{2^k}^{(k)}\}$ of $\mathcal{G}_k(F_0, F_1)$ by successive midpoint projection onto $\mathcal{E}$ as follows: For $k = 1$, let $F_0^{(1)} = F_0$ and $F_2^{(1)} = F_1$ as we must. Define $F_1^{(1)}$ to be the midpoint of the chordal geodesic from F_0 to F_1, projected onto $\mathcal{E}$ as in Theorem 3.3. Then, supposing $\{F_0^{(k)}, F_1^{(k)}, \ldots, F_{2^k}^{(k)}\}$ to be defined, put $F_{2j}^{(k+1)} = F_j^{(k)}$ for $j = 0, 1, \ldots, 2^k$. Also, for $j = 0, 1, \ldots, 2^k - 1$, set $F_{2j+1}^{(k+1)}$ to be the midpoint of the chordal geodesic from $F_j^{(k)}$ to $F_{j+1}^{(k)}$, projected onto $\mathcal{E}$ as in Theorem 3.3.

Lemma 3.4 (Discrete Geodesics): *For all $k \geq 1$,*

$$\sum_{j=0}^{2^k-1} W_2(F_j^{(k)}, F_{j+1}^{(k)}) \leq \sum_{j=0}^{2^k-1} W_2(G_j, G_{j+1})$$

for any $\{G_0, G_1, \ldots, G_{2^k}\}$ in $\mathcal{G}_k(F_0, F_1)$, and there is equality when and only when

$$\{G_0, G_1, \ldots, G_{2^k}\} = \{F_0^{(k)}, F_1^{(k)}, \ldots, F_{2^k}^{(k)}\} \ .$$

Proof By the condition (3.24),

$$\sum_{j=0}^{2^k-1} W_2(G_j, G_{j+1}) = \left(\sum_{j=0}^{2^k-1} \frac{W_2^2(G_j, G_{j+1})}{2^{-k}} \right)^{1/2} . \tag{3.25}$$

We now claim that

$$\sum_{j=0}^{2^k-1} W_2^2(F_j^{(k)}, F_{j+1}^{(k)}) \leq \sum_{j=0}^{2^k-1} W_2^2(G_j, G_{j+1})$$

and there is equality exactly when $\{G_0, G_1, \ldots, G_{2^k}\} = \{F_0^{(k)}, F_1^{(k)}, \ldots, F_{2^k}^{(k)}\}$. Clearly, on account of (3.25), once this is established, the proof is complete.

For $k = 1$, this is implied by Theorem 3.3. For $k > 1$, consider any $2^k + 1$–tuple $\{G_0, G_1, \ldots, G_{2^k}\}$ of elements of $\mathcal{E}$. We *are not* requiring $\{G_0, G_1, \ldots, G_{2^k}\} \in \mathcal{G}_k$. The point is that we are going to reduce to the case $k = 0$ by successively erasing every other element. Even if $W_2(G_j, G_{j+1}) = W_2(G_{j+1}, G_{j+2})$ for all j, it is not necessarily the case that $W_2(G_j, G_{j+2}) = W_2(G_{j+2}, G_{j+4})$ for all j, so the procedure of "erasing midpoints" does not take us from $\mathcal{G}_k$ to $\mathcal{G}_{k-1}$

Nonetheless, without assuming that $\{G_0, G_1, \ldots, G_{2^k}\} \in \mathcal{G}_k$, we have from Theorem 3.3, with f given by (3.13), that

$$
\begin{aligned}
\sum_{j=0}^{2^k-1} W_2^2(G_j, G_{j+1}) &= \sum_{\ell=0}^{2^{k-1}-1} \left(W_2^2(G_{2\ell}, G_{2\ell+1}) + W_2^2(G_{2\ell+1}, G_{2\ell+2}) \right) \\
&\geq \sum_{\ell=0}^{2^{k-1}-1} f(W_2^2(G_{2\ell}, G_{2\ell+2})) \\
&= 2^{k-1} \left(\frac{1}{2^{k-1}} \sum_{\ell=0}^{2^{k-1}-1} f(W_2^2(G_{2\ell}, G_{2\ell+2})) \right) \\
&\geq 2^{k-1} f \left(\frac{1}{2^{k-1}} \sum_{\ell=0}^{2^{k-1}-1} W_2^2(G_{2\ell}, G_{2\ell+2}) \right)
\end{aligned}
\tag{3.26}
$$

where the last inequality is the convexity of f.

Notice that both inequalities are saturated if and only if for each ℓ, $G_{2\ell+1}$ is the projected midpoint of the chordal geodesic connecting $G_{2\ell}$ and $G_{2\ell+2}$.

The proof is now easy to complete. Define a sequence $\{A_j\}$ inductively by $A_0 = W_2^2(F_0, F_1)$ and

$$
A_{j+1} = 2^j f \left(2^{-j} A_j \right) .
\tag{3.27}
$$

Because these inequalities are saturated for $\{G_0, G_1, \ldots, G_{2^k}\} = \{F_0^{(k)}, F_1^{(k)}, \ldots, F_{2^k}^{(k)}\}$,

$$
A_k = \sum_{j=0}^{2^k} W_2^2(F_j^{(k)}, F_{j+1}^{(k)}) .
$$

But a simple induction argument based on (3.26) shows that $\sum_{j=0}^{2^k} W_2^2(G_j, G_{j+1}) \geq A_k$ with equality only in the stated case. ∎

We can now define the distance $\mathcal{W}_2(F_0, F_1)$ on $\mathcal{E}$ induced by the 2–Wasserstein metric:

$$
\mathcal{W}_2(F_0, F_1) = \lim_{k \to \infty} \sum_{j=1}^{2^k-1} W_2(F_j^{(k)}, F_{j+1}^{(k)})
\tag{3.28}
$$

where clearly the sequence on the right in (3.28) is increasing. In fact, Lemma 3.4 tells us that the geodesic from F_0 to F_1 on $\mathcal{E}$ is obtained by the following simple rule: Take the chordal geodesic $t \mapsto F_t$ from F_0 to F_1 in $\mathcal{P}$, and rescale each F_t onto $\mathcal{E}$ as in Theorem 3.1. Then reparameterize this path in $\mathcal{E}$ so that it runs at constant speed. This is the geodesic. Note that this same procedure produces geodesics on the sphere S^{d-1} in $\mathbb{R}^d$.

It is now an easy matter to compute the distance $\mathcal{W}_2(F_0, F_1)$. One way is to compute $\lim_{k \to \infty} A_k$ for the sequence given by $A_0 = W_2^2(F_0, F_1)$ and (3.27). This is straightforward; it is easy to recognize the iteration as same iteration one gets by dyadically rectifying an arc of the circle.

We find it more enlightening to obtain an explicit parameterization of the corresponding geodesic, and use the Riemannian metric for the 2–Wasserstein distance.

To begin the computation, let ψ be the convex function such that $\nabla\psi \# F_0 = F_1$. We may assume without loss of generality that $u = 0$; this will simplify the computation. Then define F_t as in (2.6) and (2.7), and let $\tilde{F}_t$ be the projection of F_t onto $\mathcal{E}$ as in Theorem 3.1. Since $u = 0$,

$$\tilde{F}_t = \left(\frac{1}{a(t)} \nabla\psi_t \right) \# F_0$$

where ψ_t is defined in terms of ψ as usual and where

$$a(t) = \sqrt{1 - 4t(1-t)\frac{W_2^2(F_0, F_1)}{2d\theta}} \ .$$

Notice that the gradient vector field on $\mathbb{R}^d$ that represents the tangent vector $\partial\tilde{F}_t/\partial t$ has two terms: One is a rescaling of the gradient vector field on $\mathbb{R}^d$ that represents $\partial F_t/\partial t$, and the other generates a dilation to keep the path on $\mathcal{E}$.

Next, we have from Theorem 2.3 that for any test function χ on $\mathbb{R}^d$, after some computation,

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} \chi(v)\tilde{F}_t(v)\mathrm{d}v = \frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}^d} \chi\left(\frac{v}{a(t)}\right) F_t(v)\mathrm{d}v$$

$$= \int_{\mathbb{R}^d} \nabla\chi(v) \cdot \left(\frac{1}{a(t)}\nabla\eta_t(a(t)v) - \frac{\dot{a}(t)}{a(t)}v \right) \tilde{F}_t(v)\mathrm{d}v \ ,$$

where η_t is given by (2.21). Hence, from (2.25), we have that

$$g\left(\frac{\partial\tilde{F}_t}{\partial t}, \frac{\partial\tilde{F}_t}{\partial t} \right) = \frac{1}{2} \int_{\mathbb{R}^d} \left| \frac{1}{a(t)}\nabla\eta_t(a(t)v) - \frac{\dot{a}(t)}{a(t)}v \right|^2 \tilde{F}_t(v)\mathrm{d}v$$

$$= \frac{1}{2a^2(t)} \int_{\mathbb{R}^d} \left| \nabla\eta_t(v) - \frac{\dot{a}(t)}{a(t)}v \right|^2 F_t(v)\mathrm{d}v \ .$$

23

By (2.23), $\int_{\mathbb{R}^d} |\nabla \eta_t(v)|^2 F_t(v)\mathrm{d}v = 2W_2^2(F_0, F_1)$, and clearly $\int_{\mathbb{R}^d} |v|^2 F_t(v)\mathrm{d}v = a^2(t)d\theta$. Finally, by Theorem 2.3 and familiar computations,

$$\int_{\mathbb{R}^d} (\nabla \eta_t(v) \cdot v) F_t(v)\mathrm{d}v = \frac{1}{2t} \int_{\mathbb{R}^d} \left(|\nabla(\psi_t)^*(v) - v|^2 + |v|^2 - |\nabla(\psi_t)^*(v)|^2 \right) F_t(v)\mathrm{d}v$$

$$= \frac{1}{2t} \left(2W_2^2(F_0, F_t) + (a^2(t) - 1)d\theta \right) = (2t - 1)W_2^2(F_0, F_1) \ .$$

Putting all of this together, one has, after some algebra,

$$g\left(\frac{\partial \tilde{F}_t}{\partial t}, \frac{\partial \tilde{F}_t}{\partial t}\right) = \frac{1}{2a^2(t)} \left[2W_2^2(F_0, F_1) + \left(\frac{\dot{a}(t)}{a(t)}\right)^2 a^2(t)d\theta - 2\frac{\dot{a}(t)}{a(t)}(2t - 1)W_2^2(F_0, F_1) \right]$$

$$= W_2^2(F_0, F_1)\frac{1}{a^4(t)}\left[1 - \frac{W_2^2(F_0, F_1)}{2d\theta}\right] \ .$$

Now we reparameterize to achieve constant unit speed. We take the map $t \mapsto \tau(t)$ to be differentiable and increasing. Then with $\tilde{F}_\tau = \tilde{F}_{\tau(t)}$ we have

$$1 = g\left(\frac{\partial \tilde{F}_\tau}{\partial \tau}, \frac{\partial \tilde{F}_\tau}{\partial \tau}\right) = g\left(\frac{\partial \tilde{F}_t}{\partial t}, \frac{\partial \tilde{F}_t}{\partial t}\right)\left|\frac{\mathrm{d}t}{\mathrm{d}\tau}\right|^2 \tag{3.29}$$

provided

$$\frac{\mathrm{d}\tau(t)}{\mathrm{d}t} = W_2(F_0, F_1)\frac{1}{a^2(t)}\sqrt{1 - \frac{W_2^2(F_0, F_1)}{2d\theta}} \ .$$

This is solved by

$$\tau(t) = \sqrt{\frac{d\theta}{2}}\arctan\left((2t - 1)\sqrt{\frac{W_2^2(F_0, F_1)}{2d\theta - W_2^2(F_0, F_1)}}\right)$$

for which $\tau(1/2) = 0$ and

$$\mathcal{W}_2(F_0, F_1) = \tau(1) - \tau(0) = 2\sqrt{\frac{d\theta}{2}}\arctan\left(\sqrt{\frac{W_2^2(F_0, F_1)}{2d\theta - W_2^2(F_0, F_1)}}\right) \tag{3.30}$$

This has a very simple interpretation: Consider two points on a circle of radius R, and let D be the length of the chord that they terminate. The arc joining them subtends an angle 2ϕ where

$$\tan(\phi) = \sqrt{\frac{D^2}{4R^2 - D^2}} \ ,$$

and hence the length of the arc joining them is

$$2R\arctan\left(\sqrt{\frac{D^2}{4R^2 - D^2}}\right) \; . \tag{3.31}$$

Since $\sqrt{(d\theta)/2}$ is the radius R_θ of $\mathcal{E}$, in that this is the 2–Wasserstein distance from any point in $\mathcal{E}$ to the unit mass at u, and since $W_2(F_0, F_1)$ is the chordal separation of F_0 from F_1 in the 2–Wasserstein distance, we have that (3.31), with $R = \sqrt{(d\theta)/2}$ and $D = W_2(F_0, F_1)$, gives us $\mathcal{W}_2(F_0, F_1)$. It is somewhat simpler to express this in terms of sines instead of tangents. From (3.31) it is easy to deduce that

$$W_2(F_0, F_1) = 2R_\theta \sin\left(\frac{\mathcal{W}_2(F_0, F_1)}{2R_\theta}\right) \; , \tag{3.32}$$

$$\mathcal{W}_2(F_0, F_1) = 2R_\theta \arcsin\left(\frac{W_2(F_0, F_1)}{2R_\theta}\right) \; . \tag{3.33}$$

We summarize this in the following theorem:

Theorem 3.5 (Geometry of $\mathcal{E}$): *Let $\mathcal{W}_2(F_0, F_1)$ denote the distance between any two points F_0 and F_1 of $\mathcal{E}$ in the metric induced on $\mathcal{E}$ by the 2–Wasserstein metric. Then $\mathcal{W}_2(F_0, F_1)$ is related to $W_2(F_0, F_1)$ through (3.32) and (3.33). Moreover, the geodesic on $\mathcal{E}$ between F_0 and F_1 is obtained from the chordal geodesic in $\mathcal{P}$ between F_0 and F_1 by the following procedure: Let $t \mapsto F_t$, $t \in [0, 1]$, denote the chordal geodesic. Then, for each such t, let $\tilde{F}_t$ denote the unique point in $\mathcal{E}$ that is closest to F_t, which is simply obtained from F_t by dilatating about the mean u. This path, reparameterized to run at constant speed, is the geodesic on $\mathcal{E}$ between F_0 and F_1.*

This theorem strongly encourages one to think of $\mathcal{E}$ in spherical terms, though we see from (3.10) that the chordal distance between any two points on $\mathcal{E}$ is no more than $\sqrt{2}$ times the radius of $\mathcal{E}$, as given by (3.2), as on the spherical cap with the azimuthal angle ϕ ranging over $0 \le \phi \le \pi/4$.

We apply this to deduce a criterion for displacement convexity on the constrained manifold $\mathcal{E}$. We say that a functional Φ is displacement convex on $\mathcal{E}$ in case for all geodesics $t \mapsto G_t$ in $\mathcal{E}$, the function $t \mapsto \Phi(G_t)$ is convex. If the gradient vector field $\nabla\eta$ on $I\!R^d$ is the tangent vector at $t = 0$ to a geodesic $t \mapsto G_t$ in $\mathcal{E}$, we define

$$\mathcal{H}ess\Phi(G_0)(\nabla\eta, \nabla\eta) = \left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\Phi(G_t)\right|_{t=0} \; . \tag{3.34}$$

This should be compared with (2.26). The differences lie in the different classes of geodesics being considered in the two cases, as well as the fact that

$$\int_{I\!R^d} v \cdot \nabla\eta(v)G_0(v)\mathrm{d}v = 0 \qquad \text{and} \qquad \int_{I\!R^d} \nabla\eta(v)G_0(v)\mathrm{d}v = 0 \tag{3.35}$$

must hold for $\nabla \eta$ to represent a tangent vector to $\mathcal{E}$ at G_0.

Since we have determined the geodesics in $\mathcal{E}$, it is now a simple matter to determine a criterion for displacement convexity in $\mathcal{E}$.

Theorem 3.6 (Displacement convexity in $\mathcal{E}$) *Let $G \mapsto \Phi(G)$ be any functional of the form*

$$\Phi(G) = \int_{\mathbb{R}^d} g(G(v)) \mathrm{d}v$$

where g twice continuously differentiable on $\mathbb{R}_+$. Define the function h by $h(t) = tg'(t) - g(t)$. Suppose that $F \in \mathcal{E}_{u,\theta}$ is such that $h(F)$ is integrable, and that at F,

$$G \mapsto \mathrm{Hess}\Phi(G)(\nabla \eta, \nabla \eta)$$

is continuous in the 2–Wasserstein metric for all test functions η. Then

$$
\begin{aligned}
\mathcal{H}ess\Phi(F)(\nabla \eta, \nabla \eta) &= \mathrm{Hess}\Phi(F)(\nabla \eta, \nabla \eta) \\
&+ \frac{d}{2R_\theta^2} \left(\int_{\mathbb{R}^d} h(F) \mathrm{d}v \right) \int_{\mathbb{R}^d} |\nabla \eta|^2 F \mathrm{d}v \ ,
\end{aligned}
\tag{3.36}
$$

where $R_\theta = \sqrt{d\theta/2}$ is the radius of $\mathcal{E}_{u,\theta}$, and $\nabla \eta$ is any gradient vectorfield satisfying (3.35) In particular, if $\Phi(F) = S(F)$ is the entropy $\int_{\mathbb{R}^d} \ln(F(v))F(v) \mathrm{d}v$ of F,

$$\mathcal{H}essS(F)(\nabla \eta, \nabla \eta) = \mathrm{Hess}S(F)(\nabla \eta, \nabla \eta) + \frac{d}{2R_\theta^2} \int_{\mathbb{R}^d} |\nabla \eta|^2 F \mathrm{d}v \ , \tag{3.37}$$

and thus the entropy is uniformly convex on the constrained manifold $\mathcal{E}_{u,\theta}$.

Proof: Without loss og generality, suppose $u = 0$. For any $F \in \mathcal{E}$, let $t \mapsto \tilde{G}_t$ be a geodesic in $\mathcal{E}$ passing through F with unit speed at $t = 0$. Pick $\delta > 0$ but sufficiently small that $\tilde{G}_\delta$ and $\tilde{G}_{-\delta}$ are both defined. By definition $\mathcal{W}_2^2(\tilde{G}_{-\delta}, \tilde{G}_\delta) = 4\delta^2$. Define $h > 0$ by

$$W_2^2(\tilde{G}_{-\delta}, \tilde{G}_\delta) = 4 \, h^2 \ .$$

By Theorem 3.5,

$$h = R_\theta \sin\left(\frac{\delta}{R_\theta} \right) = \delta + \mathcal{O}(\delta^3) \ . \tag{3.38}$$

Now let $t \mapsto G_t$ be the *chordal* geodesic, in $\mathcal{P}$, from $\tilde{G}_{-\delta}$ to $\tilde{G}_\delta$ parameterized so that $\tilde{G}_{-\delta} = G_{-h}$ and $\tilde{G}_\delta = G_h$. By Theorem 3.3, $\tilde{G}_0 = F$ is obtained from G_0 by dilation:

$$\tilde{G}_0(v) = a^d G_0(av) \tag{3.39}$$

where

$$a = \sqrt{1 - \frac{h^2}{R_\theta^2}} \ . \tag{3.40}$$

Now

$$\frac{1}{\delta^2} \left[\frac{1}{2} \left(\Phi(\tilde{G}_\delta) + \Phi(\tilde{G}_{-\delta}) \right) - \Phi(\tilde{G}_0) \right] = \frac{\Phi(G_0) - \Phi(\tilde{G}_0)}{\delta^2}$$
$$+ \frac{1}{h^2} \left[\frac{1}{2} \left(\Phi(G_h) + \Phi(G_{-h}) \right) - \Phi(G_0) \right] \frac{h^2}{\delta^2} \tag{3.41}$$

Next, since

$$\Phi(G_0) - \Phi(\tilde{G}_0) = a^d \int_{I\!\!R^d} g(a^{-d} F(v)) \mathrm{d}v - \int_{I\!\!R^d} g(F(v)) \mathrm{d}v \ ,$$

it follows from (3.40) and the definition of h that

$$\lim_{\delta \to 0} \frac{\Phi(G_0) - \Phi(\tilde{G}_0)}{\delta^2} = \frac{d}{2R_\theta^2} \int_{I\!\!R^d} h(F(v)) \mathrm{d}v \ . \tag{3.42}$$

By (3.38), the continuity of $\mathrm{Hess}\Phi$ at F and our previous definitions

$$\lim_{\delta \to 0} \frac{1}{h^2} \left[\frac{1}{2} \left(\Phi(G_h) + \Phi(G_{-h}) \right) - \Phi(G_0) \right] \frac{h^2}{\delta^2} = \mathrm{Hess}\Phi(F) \ .$$

Combining this, (3.42) and (3.41), we obtain (3.36) from which the rest of the result easily follows. ∎

As an application, we deduce a strengthened form of an inequality due to Talagrand [21]. Let G_0 be a Gaussian Density in $\mathcal{E}_{\theta,u}$. Let F be any other density in $\mathcal{E}_{\theta,u}$. Let F_s be the geodesic in $\mathcal{E}_{\theta,u}$, parameterized by arclength, starting at F and going to G_0. Then by (3.37),

$$S(F) - S(G_0) = \int_0^{\mathcal{W}_2(G_0,F)} S'(F_s) \mathrm{d}s$$
$$= \int_0^{\mathcal{W}_2(G_0,F)} \left(S'(G_0) + \int_0^s S''(F_r) \mathrm{d}r \right) \mathrm{d}s \geq \frac{1}{2} \frac{d}{2R_\theta^2} \mathcal{W}_2^2(G_0,F) \ .$$

We have used the fact that $S'(G_0) = 0$ since $S(F) \geq S(G_0)$ by the entropy minimizing property of Gaussians. Also, since both F and G_0 lie in $\mathcal{E}_{\theta,u}$, $S(F) - S(G_0) = H(F|G_0)$, the relative entropy of F with respect to G_0. Therefore, since $R^2 = 2/(d\theta)$,

$$H(F|G_0) \geq \frac{1}{2\theta} \mathcal{W}_2^2(G_0,F) \ ,$$

which is Talagrand's inequality, except that here $\mathcal{W}_2^2(G_0, F)$ replaces the smaller quantity $W_2^2(G_0, F)$.

4: The Euler-Lagrange Equation

Recall that for fixed $h > 0$, and a given density $F_0 \in \mathcal{E}_{\theta,u}$, we seek to mimimize the functional

$$I(F) = \left[\frac{W_2^2(F_0, F)}{\theta} + hS(F) \right] \tag{4.1}$$

subject to the constraint that $F \in \mathcal{E}_{\theta,u}$.

This functional is strictly convex and our constraints are convex, and hence if any minimizer does exist, it would also be unique. The existence issue will be settled in the next section. Here we shall derive the Euler Lagrange equation that would be satisfied by any minimizer in our variational problem, and derive some consequences of satisfying this equation.

Theorem 4.1 *Suppose that F_1 is a minimizer of the functional given in (4.1) subject to the constraint that F_1 has the same mean and variance as F_0. Let ψ be the convex function on $I\!\!R^d$ such that*

$$\nabla\psi \# F_1 = F_0 \ . \tag{4.2}$$

Then

$$\int_{I\!\!R^d} |\nabla \ln F_1|^2 F_1(v) \mathrm{d}v < \infty \tag{4.3}$$

and

$$\nabla\psi(v) = v + h\theta \nabla_v \left(\ln \frac{F_1}{M_{F_1}} \right) + (u - v) \left[\frac{W_2^2(F_1, F_0)}{d\theta} \right] \ . \tag{4.4}$$

where for any $F \in \mathcal{P}$, M_F denotes the isotropic Gaussian density with the same mean and variance as F.

Proof: Consider a function $\xi : I\!\!R^d \to I\!\!R^d$ satisfying

$$\int_{I\!\!R^d} \xi(v) F_1(v) \mathrm{d}v = 0 \qquad \text{and} \qquad \int_{I\!\!R^d} (\xi(v) \cdot v) F_1(v) \mathrm{d}v = 0 \ . \tag{4.5}$$

Then define the flow $T_t(v) = v + t\xi(v)$ and the curve of densities $G(t) = T_t \# F_1$. Finally, let $\tilde{G}(t)$ be the projection of $G(t)$ onto $\mathcal{E}$ as in Theorem 3.1. Let u_1 and $d\theta_1$ be the mean and variance of F_1. Then by Theorem 3.1,

$$\tilde{G}(t, v) = a(t)^d G(t, a(t)(v - u(t)) + u_1) \ , \tag{4.6}$$

where, by (4.5)

$$a(t) = 1 + \mathcal{O}(t^2) \qquad \text{and} \qquad u(t) = u_1 + \mathcal{O}(t^2) \ . \tag{4.7}$$

We can also write $\tilde{G}(t) = \tilde{T}_t \# F_1$ where

$$\tilde{T}_t(v) = \frac{1}{a(t)}v + t\xi\left(\frac{1}{a(t)}v\right) .$$

The argument here is adapted from the corresponding argument in [12]. First, consider the entropy. By direct calculation and (4.7),

$$S(\tilde{G}(t)) - S(\tilde{G}(0)) = -t\int_{I\!\!R^d} F_1(v)\nabla \cdot \xi(v)\mathrm{d}v + \mathcal{O}(t^2)$$

and so

$$\lim_{t\to 0^+} \frac{S(\tilde{G}(t)) - S(F_1)}{t} = -\int_{I\!\!R^d} F_1(v)\nabla \cdot \xi(v)\mathrm{d}v .$$

To compute the variation in the 2–Wasserstein distance, note that since $\tilde{T}_t \# F_1 = \tilde{G}(t)$, $\nabla\psi \circ \tilde{T}_t^{-1} \# \tilde{G}(t) = F_0$. Thus

$$W_2^2(\tilde{G}(t), F_0) \leq \frac{1}{2}\int_{I\!\!R^d} |\nabla\psi \circ \tilde{T}_t^{-1}(v) - v|^2 \tilde{G}(t,v)\mathrm{d}v$$

$$= \frac{1}{2}\int_{I\!\!R^d} |\nabla\psi - \tilde{T}_t(v)|^2 F_1(v)\mathrm{d}v$$

$$\leq W_2^2(F_1, F_0) - t\int_{I\!\!R^d} (\nabla\psi - v)\cdot \xi F_1(v)\mathrm{d}v + o(t) .$$

Now it easily follows that

$$\limsup_{t\to 0^+} \frac{W_2^2(G(t), F_0) - W_2^2(F_1, F_0)}{t} \leq \int_{I\!\!R^d} (v - \nabla\psi(v))\, F_1(v)\cdot \xi(v)\mathrm{d}v . \qquad (4.8)$$

We deduce that

$$\int_{I\!\!R^d} \left((\nabla\psi(v) - v)\frac{F_1(v)}{\theta}\right)\cdot \xi(v)\mathrm{d}v \leq -h\int_{I\!\!R^d} F_1(v)\nabla \cdot \xi(v)$$

for all smooth and compactly supported ξ satisfying (4.5). Since these conditions are still satisfied if ξ is replaced by $-\xi$, we have that

$$\int_{I\!\!R^d} \left((\nabla\psi(v) - v)\frac{F_1(v)}{\theta}\right)\cdot \xi(v)\mathrm{d}v = -h\int_{I\!\!R^d} F_1(v)\nabla \cdot \xi(v)$$

for all smooth and compactly supported ξ satisfying (4.5). Hence

$$\left((\nabla\psi(v) - v)\frac{F_1(v)}{\theta} - h\nabla F_1(v)\right) = (A + B(u - v))F_1(v) \qquad (4.9)$$

for some vector A and scalar B. It follows from this that (4.3) holds.

Now integrating both sides of (4.9) in v, one learns that $A = 0$, and if one takes the inner product of both sides with $(u - v)$, and then integrates in v, one learns that

$$d\theta B = \frac{W_2^2(F_1, F_0)}{\theta} - dh$$

since

$$\int_{\mathbb{R}^d} (\nabla \psi(v) - v) \cdot v F_1(v) \mathrm{d}v = W_2^2(F_1, F_0) \ .$$

Combining this and (4.9), we obtain (4.4) ∎

Now still assuming that the minimizer F_1 exists, we ask *what properties does F_1 inherit from F_0?* We shall show, using the fact that F_1 satisfies the Euler–Lagrange equation (4.4) and (4.2), that F_1 inherets some localization properties from F_0. Specifically, let ζ be a non-negative, increasing convex function on $\mathbb{R}_+$ with the property that $\lim_{t \to \infty} \zeta(t)/t = \infty$ and that $\zeta(0) = 0$. Suppose that

$$\int_{\mathbb{R}^d} \zeta(|v|^2) F_0(v) \mathrm{d}v = C < \infty \ . \tag{4.10}$$

This quantity provides a quantitative measure of the localization of $|v|^2 F_0(v)$ in that

$$\int_{|v|^2 > t} |v|^2 F_0(v) \mathrm{d}v \leq \frac{t}{\zeta(t)} C \ ,$$

and the right hand side tends to zero as t increases. Here, we have used that $t \to \zeta(t)/t$ is nondecreasing. If we knew that F_1 satisfied the same inequality, we would have a quantitative localization estimate on F_1. We shall see below that this is almost the case: The function ζ is modified slightly in passing from F_0 to F_1.

First, we need to explain where the original ζ comes from. We could take $\zeta(t) = (1+t)^{1+\epsilon}$ if we assumed that F_0 possessed more than second moments. Since we wish to make a statement about generic elements F_0 of $\mathcal{E}_{u,\theta}$, we use a minor variant of a lemma of de la Vallée–Poussin, which says that for any probability density F_0 with $\int_{\mathbb{R}^d} |v|^2 F_0(v) \mathrm{d}v < \infty$, there is a a non-negative, increasing convex function on $\mathbb{R}_+$ with the property that $\lim_{t \to \infty} \zeta(t)/t = \infty$ such that (4.10) holds, and finally, that $\|\zeta''\|_\infty \leq 1$. Everything up to the last condition is standard, though the usual construction of ζ is such that ζ'' is a series of Dirac masses. We therefore sketch a short proof. Without loss of generality, we may suppose that $u = 0$ and $\theta = 1/d$.

Let

$$\lambda(t) = \int_{|v|^2 > t} F_0(v) \mathrm{d}v \qquad \text{and} \qquad \mu(t) = \int_{|v|^2 > t} |v|^2 F_0(v) \mathrm{d}v$$

so that $1 = \int_{I\!\!R^d} |v|^2 F_0(v) \mathrm{d}v = \int_0^\infty \lambda(t) \mathrm{d}t$, and that

$$\mu(t) = \int_t^\infty \lambda(u) \mathrm{d}u + \lambda(t) \geq \sum_{n>t} \lambda(n) \ . \tag{4.11}$$

Here, we have used the Layer cake representation theorem. Now define t_k by $t_0 = 0$ and for $k \geq 1$, $t_k = \inf\{t \mid \mu(t) < 2^{-k}\}$. Clearly, since $F_0(v) \mathrm{d}v$ is absolutely continuous, $\mu(t_k) = 2^{-k}$. Then by (4.11),

$$1 = \sum_{k=1}^\infty \mu(t_k) = \sum_{k=1}^\infty \sum_{n>t_k} \lambda(n) = \sum_{n=1}^\infty g(n)\lambda(n) \tag{4.12}$$

where $g(0) = 0$ and for all $n \geq 1$,

$$g(n) = \max\{k \mid t_k < n\} \ .$$

Clearly, $\lim_{n\to\infty} g(n) = \infty$ and $g(n+1) \geq g(n)$. Next, set $h(0) = 0$ and for $n \geq 1$, define $h(n)$ recursively by $h(n) - h(n-1) = 1$ if $g(n) - g(n-1) > 0$, and $h(n) - h(n-1) = 0$ otherwise. Then clearly

$$h(n) = \sum_{k=1}^n (h(k) - h(k-1)) \leq \sum_{k=1}^n (g(k) - g(k-1)) = g(n)$$

but also clearly $\lim_{n\to\infty} h(n) = \infty$ since $g(n)$ must increase infinitely often.

Now define $h(t)$ for all $t > 0$ by linear interpolation of $h(n)$, and then define $\zeta(t) = \int_0^t h(s) \mathrm{d}s$. Note that $\zeta(t)$ is a continuously differentiable convex increasing function with $\|\zeta''\|_\infty \leq 1$, and $\lim_{t\to\infty} \zeta(t)/t = \infty$. Also, since $\zeta(t)$ is increasing and $\lambda(t)$ is decreasing,

$$\int_0^\infty \zeta'(t)\lambda(t) \mathrm{d}t \leq \sum_{n=0}^\infty h(n+1)\lambda(n) \leq \sum_{n=0}^\infty (1 + g(n))\lambda(n) \leq 3 \ ,$$

where the last inequality follows from (4.12). Since

$$\int_{I\!\!R^d} \zeta(|v|^2) F_0(v) \mathrm{d}v = \int_0^\infty \zeta'(t)\lambda(t) \mathrm{d}t \ , \tag{4.13}$$

(4.10) holds.

We are now ready to prove the following:

Theorem 4.2: *Suppose F_0 is any element of $\mathcal{E}_{\theta,u}$, and suppose ψ is a convex potential with $\nabla\psi\#F_1 = F_0$ such that ψ and F_1 satisfy (4.4). Then there are a non–negative, increasing*

convex function $\zeta(t)$ such that $\lim_{t\to\infty} \zeta(t)/t = \infty$ and $\|\zeta''\|_\infty \leq 1$, and a finite constant C, both depending only on F_0, so that

$$\int_{I\!R^d} \zeta(\alpha|w-u|^2)F_1(w)\mathrm{d}w < C$$

for some α depending only on h, $W_2(F_0, F_1)$, and θ.

Proof: Without loss of generality, we continue to assume that $u = 0$ and $\theta = 1$, and thus

$$\nabla\psi(w) = \alpha w + h\nabla\ln(F_1(w))$$

for some constant $\alpha > 0$ that is readily computed from (4.4). Now let $\zeta(t)$ be the increasing convex function provided by the variant of the de la Vallée–Poussin lemma. Then, $v \to \zeta(|v|)$ is convex and so,

$$\begin{aligned}
\int_{I\!R^d} \zeta(|v|^2)F_0(v)\mathrm{d}v &= \int_{I\!R^d} \zeta(|\nabla\psi(w)|^2)F_1(w)\mathrm{d}w \\
&= \int_{I\!R^d} \zeta(|\alpha w + h\nabla\ln(F_1(w))|^2)F_1(w)\mathrm{d}w \\
&\geq \int_{I\!R^d} \zeta(\alpha|w|^2)F_1(w)\mathrm{d}w + 2h\alpha\int_{I\!R^d} \zeta'(\alpha|w|^2)w\cdot\nabla F_1(w)\mathrm{d}w \\
&= \int_{I\!R^d} \zeta(\alpha|w|^2)F_1(w)\mathrm{d}w - 2h\alpha^2\int_{I\!R^d} \zeta''(\alpha|w|^2)|w|^2\cdot F_1(w)\mathrm{d}w \\
&\quad - 2h\alpha d\int_{I\!R^d} \zeta'(\alpha|w|^2)\cdot F_1(w)\mathrm{d}w
\end{aligned}$$

Since $\int_{I\!R^d} |w|^2\cdot F_1(w)\mathrm{d}w = 1$,

$$\int_{I\!R^d} \zeta(\alpha|w|^2)F_1(w)\mathrm{d}w \leq \int_{I\!R^d} \zeta(|v|^2)F_0(v)\mathrm{d}v + 2h\alpha^2(1+d)\ , \qquad (4.14)$$

where we are using the fact that $\|\zeta''\|_\infty \leq 1$ and $\zeta'(t) \leq t$ when ζ is the function provided by the above variant of the de la Vallée–Poussin lemma. ∎

5: Existence of Minimizers

To simplify the notation, we fix $u = 0$ and $\theta = 1$ throughout this section. The main goal in this section is to prove that a minimizer exists for (4.1). As explained in the introduction, it suffices to find a density $F_1 \in \mathcal{E}$ and a convex potential ψ with $\nabla\psi\#F_1 = F_0$ such that the Euler–Lagrange equation (4.4) is satisfied.

In this, we make essential use of the dual version of the variational characterization of the 2–Wasserstein metric. This says that for all F_0 and F in $\mathcal{E}$,

$$d - W_2^2(F_0, F) =$$

$$\inf \left\{ \int_{\mathbb{R}^d} \phi(v) F_0(v) \mathrm{d}v + \int_{\mathbb{R}^d} \psi(w) F(w) \mathrm{d}w \;\middle|\; \phi(v) + \psi(w) \geq v \cdot w \text{ a.e.} \right\},$$

(5.1)

where almost everywhere refers to the measure $F_0(v) F_1(w) \mathrm{d}v \mathrm{d}w$. Furthermore, the minimizing pair, which exists, consists of a dual pair of convex functions. That is, we may assume that ϕ and ψ are Legendre transforms of one another. The gradients of the minimizing pair provide the optimal transport plans; i.e., $\nabla \phi \# F_0 = F$ and $\nabla \psi \# F = F_0$. A good reference for this is [3] or [8].

We shall assume strong assumptions on $F_0 \in \mathcal{E}$, which we shall later remove; namely we suppose that F_0 is supported in B_R, the centered ball of radius R, and that on B_R it is bounded below by some strictly positive number α. Then for any other density F in $\mathcal{P}$, these hypotheses impose some regularity on the optimal map $\nabla \psi \# F = F_0$. In particular,

$$|\nabla \psi(v)| \leq R$$

(5.2)

for all v, which means that ψ is Lipschitz.

Now define $\eta(t)$ by

$$\eta(t) = \begin{cases} +\infty & \text{if } t < 0, \\ t \ln t & \text{if } t \geq 0. \end{cases}$$

Then the Legendre transform $\eta^*(s)$ of $\eta(t)$ is

$$\eta^*(s) = e^{s-1} .$$

We shall use use the notation η^* throughout this section to emphasize the fact that we do not make much use of the specific form of η in our analysis; this point is discussed further at the end of the section. Then

$$S(F) = \int_{\mathbb{R}^d} \eta(F) \mathrm{d}v ,$$

and for any dual convex pair of functions ϕ and ψ,

$$I(F) \geq h S(F) + d - \left(\int_{\mathbb{R}^d} \phi(v) F_0(v) \mathrm{d}v + \int_{\mathbb{R}^d} \psi(w) F(w) \mathrm{d}w \right) ,$$

(5.3),

where $I(F)$ is given by (4.1). Moreover, by Young's inequality, $\eta(t) + \eta^*(s) \geq st$, and thus we have that for any $a \in R^d$ and any $b \in \mathbb{R}$,

$$\eta(F) + \eta^* \left(\frac{a \cdot w + b|w|^2/2 + \psi(w)}{h} \right) \geq \frac{a \cdot w + b|w|^2/2 + \psi(w)}{h} F .$$

(5.4)

Integrating yields

$$hS(F) - \int_{\mathbb{R}^d} \psi(w)F(w)\mathrm{d}w \geq \frac{d}{2}b - h\int_{\mathbb{R}^d} \eta^* \left(\frac{a \cdot w + b|w|^2/2 + \psi(w)}{h} \right) \mathrm{d}w . \qquad (5.5)$$

Therefore, introduce the functional

$$J(a,b,\phi,\psi) = d - \int_{\mathbb{R}^d} \phi(v)F_0(v)\mathrm{d}v + \frac{d}{2}b - h\int_{\mathbb{R}^d} \eta^* \left(\frac{a \cdot w + b|w|^2/2 + \psi(w)}{h} \right) \mathrm{d}w . \qquad (5.6)$$

Note that ϕ is bounded below and η^* is positive, and hence $J(a,b,\phi,\psi)$ is well–defined. It then follows from (5.3), (5.5) and (5.6) that or any dual convex pair of functions ϕ and ψ, $a \in R^d$ and any $b \in \mathbb{R}$,

$$I(F) \geq J(a,b,\phi,\psi) , \qquad (5.7)$$

We let $\mathcal{U}$ denote the set of all quadruplets (a,b,ϕ,ψ) where $a \in \mathbb{R}^d$, $b \in \mathbb{R}$, and ϕ and ψ are a pair of dual convex functions with

$$\phi(v) = \infty \qquad \text{for} \quad |v| > R . \qquad (5.8)$$

The reason for this last condition is that increasing ϕ off of the support of F_0 can only decrease ψ and hence increase J, so we may freely restrict our attention to such dual pairs; see [8] or [3]. This guarantees that (5.2) holds whenever $(a,b,\phi,\psi) \in \mathcal{U}$. Indeed, since ψ is determined by ϕ through the Legendre transform, J can be regarded as a functional of a, b and ϕ alone. However, the notation with ϕ included as a variable is convenient for the exposition.

As we will see below,

$$\min\{I(F) \,|F \in \mathcal{E}\} = \max\{J(a,b,\phi,\psi) \mid (a,b,\phi,\psi) \in \mathcal{U}\} . \qquad (5.9)$$

The parameters a and b will be seen to function as Lagrange multipliers guaranteeing that at the maximum on the right, $F_1 = \nabla\phi\#F_0$ does belong to $\mathcal{E}$.

Theorem 5.1 *There exists $(a_0,b_0,\phi_0,\psi_0) \in \mathcal{U}$ such that*

$$J(a_0,b_0,\phi_0,\psi_0) \geq J(a,b,\phi,\psi) \qquad (5.10)$$

for all $(a,b,\phi,\psi) \in \mathcal{U}$. Furthermore, if we define

$$F_1(w) = (\eta^*)' \left(\frac{a_0 \cdot w + b_0|w|^2/2 + \psi_0(w)}{h} \right) \qquad (5.11)$$

then $F_1 \in \mathcal{E}$,

$$\nabla\psi_0\#F_1 = F_0 \qquad (5.12)$$

and

$$\nabla \psi_0(w) = w + h\nabla \ln(F_1) + hdw - W_2^2(F_0, F_1) \ . \tag{5.13}$$

Note that this gives us a solution of the Euler–Lagrange equation for the minimum of $I(F)$ that we derived in the last section. And indeed, since $\eta(t) + \eta^*(s) = st$ with

$$t = F_1 \quad \text{and} \quad s = \frac{a_0 \cdot w + b_0|w|^2/2 + \psi_0(w)}{h}$$

with $F = F_1$, $\psi = \psi_0$, there is equality in (5.4). By (5.12), there is equality in (5.3) when $F = F_1$, $\psi = \psi_0$ and $\phi = \phi_0$. It follows that $I(F_1) = J(a_0, b_0, \phi_0, \psi_0)$. Together with (5.7), this proves that F_1 minimizes I on $\mathcal{E}$. Thus Theorem 5.1 provides us with our minimizer of the original problem. The advantage of the J functional lies in the compactness properties of the dual convex pairs.

Proof: First, suppose that the maximizer $(a_0, b_0, \phi_0, \psi_0)$ does exist. Observe that for any real number λ, $(a_0, b_0, \phi_0 + \lambda, \psi_0 - \lambda) \in \mathcal{U}$. Then by (5.10)

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} J(a_0, b_0, \phi_0 + \lambda, \psi_0 - \lambda)\Big|_{\lambda=0} = 0$$

and this clearly leads to

$$1 = \int_{I\!\!R^d} (\eta^*)' \left(\frac{a_0 \cdot w + b_0|w|^2/2 + \psi_0(w)}{h} \right) \mathrm{d}w \ . \tag{5.14}$$

Hence we see that (5.11) does define a probability density.

Next, we shall see below that for some $\epsilon > 0$,

$$\int_{I\!\!R^d} e^{\epsilon|w|^2} F_1(w)\mathrm{d}w < \infty \ . \tag{5.15}$$

This implies that

$$(a, b) \mapsto \int_{I\!\!R^d} (\eta^*)' \left(\frac{a \cdot w + b|w|^2/2 + \psi_0(w)}{h} \right) \mathrm{d}w$$

is a differentiable function of a and b in some neighborhood of (a_0, b_0). Assuming this for the moment,

$$\frac{\mathrm{d}}{\mathrm{d}b} J(a_0, b, \phi_0, \psi_0)\Big|_{b=b_0} = 0 \ ,$$

and from this we have that

$$\frac{d}{2} = \int_{I\!R^d} \frac{|w|^2}{2} \, (\eta^*)' \left(\frac{a_0 \cdot w + b_0 |w|^2/2 + \psi_0(w)}{h} \right) \mathrm{d}w$$

which means that F_1 does indeed satisfy the variance constraint. In the same way, differentiating in a shows that F_1 does satisfy the mean constraint. Thus, $F_1 \in \mathcal{E}$.

So far, the only variation that we have made in ϕ_0, and hence ψ_0, is a shift by an additive constant. We now let ζ be any smooth function supported in the interior of B_R, and define $\phi_t = \phi_0 + t\zeta$, and let ψ_t be the Legendre transform of ϕ_t. While these are not a dual pair of convex functions since ϕ_t may fail to be convex, it is nonetheless clear that for all sufficiently small t, $J(a_0, b_0, \phi_0, \psi_0) \geq J(a_0, b_0, \phi_t, \psi_t)$ and thus

$$\frac{\mathrm{d}}{\mathrm{d}t} J(a_0, b_0, \phi_t, \psi_t) \bigg|_{t=0} = 0 \ .$$

But as in [10] $\lim_{t \to 0} (\psi_t(w) - \psi_0(w))/t = -\zeta(\nabla \psi_0(w))$. From this it follows that

$$\int_{I\!R^d} \zeta(v) F_0(v) \mathrm{d}v = \int_{I\!R^d} \zeta(\nabla \psi_0(w)) F_1(w)) \mathrm{d}w$$

which means that $\nabla \psi_0 \# F_1 = F_0$.

The remaining part of the Euler–Lagrange equation follows from (5.11) by simple differentiation:

$$h \nabla F_1(w) = (a_0 + b_0 w + \nabla \psi_0(w)) F_1(w) \ . \tag{5.16}$$

Hence

$$hw \cdot \nabla F_1(w) = \left(a_0 \cdot w + b_0 |w|^2 + w \cdot \nabla \psi_0(w) \right) F_1(w) \ ,$$

and integrating both sides we obtain that

$$b_0 = -(1+h) + \frac{W_2^2(F_0, F_1)}{d} \ .$$

Even more simply, one sees by integrating (5.16) that $a_0 = 0$. Thus, provided the maximizer exists, and that $(a, b) \mapsto J(a, b, \phi_0, \psi_0)$ is differentiable in a neighborhood of (a_0, b_0), we have that $F_1 \in \mathcal{E}$, $\nabla \psi_0 \# F_1 = F_0$, and that the Euler–Lagrange equation (5.13) is satisfied.

To show the existence of an optimizer, we begin by considering any (a, b, ϕ, ψ) We now seek an *a–priori* lower bound on $\phi(v)$. Fix any $v \in B_R$ at which ϕ is differentiable. Then let $w_0 = \nabla \phi(v)$. Since ψ and ϕ are dual to one another, v belongs to the subgradient of ψ at w_0, and then by the convexity of ψ, for any $w \in I\!R^d$, $\psi(w) \geq \psi(w_0) + v \cdot (w - w_0)$.

Then since ψ is convex, and because of the mononicity of $(\eta^*)'$ and its specific form, we have that

$$(\eta^*)' \left(\frac{a \cdot w + b|w|^2/2 + \psi(w)}{h} \right) \geq \exp\left((\psi(w_0) - v \cdot w_0)/h \right) (\eta^*)' \left(\frac{(a+v) \cdot w + b|w|^2/2}{h} \right)$$

Integrating, and using (5.14), we see that b is negative, and we obtain

$$1 \geq \exp\left((\psi(w_0) - v \cdot w_0)/h\right) e^{-1} \exp\left(\frac{|a+v|^2}{2h|b|}\right) \left(\frac{2\pi h}{|b|}\right)^{d/2} . \tag{5.17}$$

But $\phi(v) = -(\psi(w_0) - w_0 \cdot v)$ and so

$$\phi(v) \geq \frac{|a+v|^2}{2|b|} - h\left(1 + \frac{d}{2}\ln\left(\frac{|b|}{2\pi h}\right)\right) . \tag{5.18}$$

Integrating against $F_0(v)$, we obtain that

$$\int_{\mathbb{R}^d} \phi(v)F_0(v)\mathrm{d}v \geq \frac{|a|^2}{2|b|} + \frac{1}{2|b|} - h\left(1 + \frac{d}{2}\ln\left(\frac{|b|}{2\pi h}\right)\right) . \tag{5.19}$$

Now consider $\tilde{\psi}$ where

$$\tilde{\psi}(w) = (1-h)\frac{|w|^2}{2} + h\left[1 - \frac{d}{2}\ln(2\pi)\right] ,$$

so that

$$\int_{\mathbb{R}^d} (\eta^*)' \left((-|w|^2/2 + \tilde{\psi}(w))/h\right)\mathrm{d}w = \left(\frac{1}{2\pi}\right)^{d/2} \int_{\mathbb{R}^d} e^{-|w|^2/2} = 1 .$$

The dual convex function of $\tilde{\psi}$ is $\tilde{\phi}$ where

$$\tilde{\phi} = \frac{1}{1-h}\frac{|w|^2}{2} - h\left[1 - \frac{d}{2}\ln(2\pi)\right] .$$

This does not satisfy (5.8), and hence $(0, -1, \tilde{\phi}, \tilde{\psi})$ is not in $\mathcal{U}$. However, define $\tilde{\phi}_R$ by $\tilde{\phi}_R(v) = \tilde{\phi}(v)$ for $|v| < R$, and $\tilde{\phi}_R$ by $\tilde{\phi}_R(v) = \infty$ otherwise, and define $\tilde{\psi}_R$ to be the dual convex function. Then $(0, -1, \tilde{\phi}_R, \tilde{\psi}_R)$ *is* in $\mathcal{U}$ and $J(0, -1, \tilde{\phi}_R, \tilde{\psi}_R) \geq J(0, -1, \tilde{\phi}, \tilde{\psi})$ since, as we have noted before, increasing ψ off the support of F_0 can only decrease the dual ψ, and hence increase J. We denote by $J_d(h)$ the finite real number $J(0, -1, \tilde{\phi}, \tilde{\psi})$, depending only on d and h. Since it is clear that

$$\sup\{J(a, b, \phi, \psi) \mid (a, b, \phi, \psi) \in \mathcal{U}\} \geq J_d(h) ,$$

and we seek a maximizer of J, we need only consider $(a, b, \phi, \psi) \in \mathcal{U}$ such that

$$J(a, b, \phi, \psi) \geq J_d(h) . \tag{5.20}$$

Furthermore, we may suppose that we have already optimized over $\phi + \lambda$ and $\psi - \lambda$ so that (5.14) holds. Then from the fact that $(\eta^*)' = \eta^*$,

$$J(a, b, \phi, \psi) = d - \int_{\mathbb{R}^d} \phi(v) F_0(v) \mathrm{d}v + b\frac{d}{2} - h \ .$$

In light of this, and (5.20),

$$\int_{\mathbb{R}^d} \phi(v) F_0(v) \mathrm{d}v \leq -J_d(h) + d(1 + \frac{b}{2}) - h \ . \tag{5.21}$$

Combining (5.19) and (5.21) we obtain after simplification that

$$-J_d(h) + d(1 + \frac{b}{2}) \geq \frac{1}{2|b|} + \frac{a^2}{2|b|} + \frac{d}{2}h\ln(2\pi h) - h\frac{d}{2}\ln|b| \ . \tag{5.22}$$

Recalling that b is negative, it is clear that $|b|$ cannot be too close to zero, for then the right hand side becomes greater than 2. Also, $|b|$ cannot be too large, since as $|b|$ increases, the left hand side tends linearly to $-\infty$, while the right hand side only does so logarithmically. Even more evidently, $|a|$ cannot be too large.

It follows that there is a constant $c > 0$, depending on h, so that

$$c \leq |b| \leq 1/c \qquad \text{and} \qquad |a| < c \ . \tag{5.23}$$

Next, use (5.11) to define F_1; that is,

$$F_1(w) = (\eta^*)' \left(\frac{a \cdot w + b|w|^2/2 + \psi(w)}{h} \right) \tag{5.24}$$

We may suppose without loss of generality that a and b have been chosen optimally so that $F_1 \in \mathcal{E}$. Since $\int_{\mathbb{R}^d} |v|^2 F_1(v) \mathrm{d}v = 1$,

$$\frac{1}{2} \leq \int_{|w| \leq \sqrt{2}} F_1(w) \mathrm{d}w \leq 1 \ .$$

This together with (5.2) and (5.24) means that for another finite constant C we have

$$|\psi(w)| \leq C + R|w| \tag{5.25}$$

for all w. In particular, with F_1 defined as in (5.24), (5.15) holds, as claimed.

This gives us all of the *a–priori* estimates that we need. Consider a sequence $\{(a_n, b_n, \phi_n, \psi_n) \in \mathcal{U}$, each of which satisfies (5.20). First we may optimize in a_n and b_n and carry out the variation over $\phi_n + \lambda$ and $\psi_n - \lambda$. With these chosen optimally, (5.14) holds.

Then by the previous paragraphs, a_n and b_n satisfy (5.23) for all n. Passing to a subsequence, we may assume that $\{a_n\}$ and $\{b_n\}$ converge to the limits a_0 and b_0 respectively.

Now for each n, define $F_1^{(n)}$ in terms of a_n, b_n and ψ_n using (5.24) Our optimizing sequence is such that for each n, $F_1^{(n)} \in \mathcal{E}$, since, as we have seen, this is what is guaranteed by optimality in a and b. Moreover, since a_n and b_n satisfy (5.23) for all n, it follows that (5.15) holds for all n for some fixed $\epsilon > 0$.

Passing to a further subsequence, we have that $\psi_0 = \lim_{n \to \infty} \psi_n$ exists uniformly on compact sets due to (5.25) and the Lipschitz bound. Since for each n, $F_1^{(n)}$ satisfies (5.15), $\lim_{n \to \infty} F_1^{(n)}$ converges strongly in L^1.

It is plain that on B_R, passing to a further subsequence if need be, we have $\lim_{n \to \infty} \phi_n = \phi$ almost everywhere and

$$\lim_{n \to \infty} \int_{B_R} \phi_n(w) F_0(w) \mathrm{d}w = \int_{B_R} \phi_0(w) F_0(w) \mathrm{d}w$$

Thus $J(a_0, b_0, \phi_0, \psi_0) = \lim_{n \to \infty} J(a_n, b_n, \phi_n, \psi_n)$. Since $\{(a_n, b_n, \phi_n, \psi_n)\}$ was a maximizing sequence, $(a_0, b_0, \phi_0, \psi_0) \in \mathcal{U}$ is the desired maximizer, and all of the properties of F_1 and ψ_0 claimed in the theorem have already been shown to be consequences of the corresponding Euler–Lagrange equations. ∎

Thus, under our given conditions on F_0, we have proved the existence of a minimizer F_1 of $I(F)$. Now consider an arbitrary element $F_0 \in \mathcal{E}$. Then there exists a convex function ζ on $I\!R_+$ as in Section 4 such that $\zeta(t)/t$ increases to infinity and

$$\int_{I\!R^d} \zeta(|v|^2) F_0(v) \mathrm{d}v = C < \infty .$$

We approximate F_0 in $L^1(I\!R^d)$ by a sequence of densities $F_0^{(n)}$ such that

$$\int_{I\!R^d} \zeta(|v|^2) F_0^{(n)}(v) \mathrm{d}v < 2C$$

for all n, and such that for each n, $F_0^{(n)}$ is supported in B_{R_n} for some radius R_n. Let $F_1^{(n)}$ be the corresponding minimizer if $I(F)$. Then by Theorem 4.2, there are numbers $\alpha > 0$ and $K < \infty$ so that

$$\int_{I\!R^d} \zeta(\alpha |v|^2) F_1^{(n)}(v) \mathrm{d}v < K \tag{5.26}$$

for all n.

By passing to a subsequence, we may suppose that $F_1^{(n)}$ converges weakly to a probability density F_1. It is clear that the first moments converge, and by (5.26) it is clear that

the second moments converge as well, and hence $F_1 \in \mathcal{E}$. Moreover, since convergence in the 2–Wasserstein metric is equivalent to weak convergecne and convergence of the second moments, $\lim_{n \to \infty} W_2^2(F_1^{(n)}, F_1) = 0$, and $\lim_{n \to \infty} W_2^2(F_0^{(n)}, F_0) = 0$. Therefore,

$$\lim_{n \to \infty} W_2^2(F_1, F_0) = W_2^2(F_1^{(n)}, F_0^{(n)}) \ .$$

Finally, by weak lower semicontinuity, $S(F_1) \leq \liminf_{n \to \infty} S(F_1^{(n)})$. It follows that F_1 is the minimizer we seek.

Then by dominated convergence, $F_1 = \lim_{n \to \infty} F_1^{(n)} \in \mathcal{E}$ and F_1 is the desired minimizer. It is unique by strict convexity. Thus we have proven the following result:

Theorem 5.2 *For all $F_0 \in \mathcal{E}$, there exists a unique $F_1 \in \mathcal{E}$ such that*

$$I(F_1) \leq I(F)$$

for all $F \in \mathcal{E}$, where $I(F)$ is defined in (4.1).

We note that on the basis of this result, there is a unique solution to the discrete time evolution problem in which, given initial data $F_0 \in \mathcal{E}$ and a time step $h > 0$, F_n is defined iteratively in terms of F_{n-1} by setting F_n to be the minimizer of

$$\left[\frac{W_2^2(F_{n-1}, F)}{\theta} + hS(F) \right]$$

over $\mathcal{E}$. It is then easy to see, using the results of section 4, that if one defines $F^{(h)}(t, v)$ by an appropriate interpolation as in [12], then $\lim_{h \to 0} F^{(h)}(t, v) = F(t, v)$ where $F(t, v)$ solves the Fokker–Planck equation

$$\frac{\partial}{\partial t} F(t, v) = \nabla \cdot \left(e^{-|v-u|^2/2\theta} \nabla (e^{|v-u|^2/2\theta} F(t, v)) \right)$$

with initial data F_0. This equation is of course already well understood, but we shall show that this way of aproaching it extends to the nonlinear spatially inhomogenous kinetic Fokker–Planck equation, which is much less well understood, in a related paper.

Open Problems: We close this section by commenting on two open problems. First, consider the variational problem employed by Jordan, Kinderleher and Otto [12] to construct solutions of the heat equation:

$$\inf\{hS(F) + W_2^2(F, F_0)\} \tag{5.27}$$

in which *no constraint* is imposed on the variance of F. We conjecture that

$$\int_{\mathbb{R}^d} |v|^2 F_1(v) \mathrm{d}v > \int_{\mathbb{R}^d} |v|^2 F_0(v) \mathrm{d}v \tag{5.28}$$

where F_1 is the minimizer for (5.27). We can prove this under several additonal assumptions – when h is not too small, when F_0 is radial, etc., and we note that if F_t solves the heat equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{I\!\!R^d} |v|^2 F_t(v)\mathrm{d}v = 2d \qquad (5.29)$$

for any initial data F_0 with finite variance. In section 3, we have given the exact solution of this variational problem, and we see a similar behavior in that case. However, we have not been able to prove (5.28) in general. It would be most unfortunate if the discrete time problem did not possess a good analog of the basic montonicity property (5.29), and we do not believe that this is the case. If (5.28) were true, it would make it easy to prove Theorem 5.2 by adding on a Lagrange multiplier $\lambda \int_{I\!\!R^d} |v|^2 F(v)\mathrm{d}v$ to the functional in (5.27). The existence (and uniqueness) of minimizers would follow by the argument in [12] for all $\lambda > 0$. Let $F^{(\lambda)}$ denote the minimizer corresponding to a given value of $\lambda \geq 0$. If (5.28) were true, it would be easy to show the existence of a value $\lambda_0 > 0$ for which $\int_{I\!\!R^d} |v|^2 F^{(\lambda_0)}(v)\mathrm{d}v = \int_{I\!\!R^d} |v|^2 F_0(v)\mathrm{d}v$. It would then follow that $F^{(\lambda_0)}$ is the minimizer provided by Theorem 5.3.

Another open problem concerns the growth of higher moments. We note that if F_t solves the heat equation for any initial data F_0 with zero mean and finite fourth moments,

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_{I\!\!R^d} |v|^4 F_t(v)\mathrm{d}v = 12d\theta \ .$$

This leads one to hope that if F_1 is the mimimizer for (5.27), and F_0 has zero mean and, say, finite sixth moments, there is a constant C depending only on, say, the sixth moments so that

$$\int_{I\!\!R^d} |v|^4 F_1(v)\mathrm{d}v \leq (1 + Ch) \int_{I\!\!R^d} |v|^4 F_0(v)\mathrm{d}v \ . \qquad (5.30)$$

This would be helpful in studying the non–linear kinetic Fokker–Planck equation by these methods. We conjecture that this is true. We note that to prove (5.30), one needs an upper bound on the moments of the minimizer F_1, while to prove (5.28), one needs a lower bound.

References

[1] A.D. Alexandroff. *Existence and uniqueness of a convex surface with a given integral curvature*, Doklady Akad Nauk., **35**, 131–134, 1942.

[2] P. Appell. *Mémoire sur les déblais et les remblais des systèmes continues ou discontinus*, Mémoires présentes par divers Savants à l'Academie des Sciences de l'Institut de France, Paris, I. N. **29**, 1–208, 1887.

[3] Y. Brenier. *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., **44** 375–417, 1991.

[4] L.A. Caffarelli, *The regularity of mappings with a convex potential* J. Amer. Math. Soc., **5** 99–104, 1992.

[5] L.A. Caffarelli, *Boundary regularity of maps with convex potentials II*, Ann. of Math., **144** 453–496, 1996.

[6] L.A. Caffarelli, *Monotonicity properties of optimal transportation and the FKG and related inequalities* to apear in Comm. Math. Phys., 2001.

[7] C. Cercignani, *Boltzmann Equation*, 1996.

[8] L.C. Evans. *Partial differential equations and Monge-Kantorovich mass transfer*, Int. Press, Current Dev. Math, **26**, 26–78, 1997.

[9] L.C. Evans and W. Gangbo, *Differential equations methods for the Monge–Kantorovich mass transfer problem Mem. Amer. Math Soc.*, **137**, 1999.

[10] W. Gangbo, *An elementary proof of the polar factorization of vector-valued functions* Arch. Rational Mech. Anal., **128** 381–399, 1994.

[11] W. Gangbo and R.J. McCann, *The geometry of optimal transportation*, Acta Math., **177** 113–161, 1996.

[12] Jordan, Kinderleher and Otto, *The variatinal formulation of the Fokker–plack equation*, SIAM Jour. Math Anal., **29** 1–17, 1998.

[13] L. Kantorovich. *On the translocation of masses* C.R. (Doklady) Acad. Sci. USSR (N.S.), **37** 199–201, 1942.

[14] L.V. Kantorovich and G.S. Rubinstein. *On a functional space and certain extremum problems*, Dokl. Akad. Nauk USSR (N.S.), **115** 1058–1061, 1957.

[15] R.J. McCann, *A Convexity Theory for Interacting Gasses and Equilibrium Chrystals*, Thesis, Princeton University, 1994.

[16] R.J. McCann, *A convexity principle for interacting gases* Adv. Math. **128** 153–179, 1997

[17] G.Monge. *Mémoire sur la théorie des déblais et de remblais*, Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année, pages 666–704 (1781).

[18] F. Otto, *The geometry of dissipative evolution eqautions: the porous medium eqaution.* Comm. P.D.E., **26** 156–186 2001.

[19] S.T. Rachev and L. Rüschendorf, *Mass Transportation Problems*; Vol. I & II, Probability and its Applications, Springer Verlag, Heidleberg 1998.

[20] S.T. Rachev, *The Monge-Kantorovich mass transference problem and its stochastic applications* Thy Prob. Appl., **29**, 647–676, 1984.

[21] M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal., **6**, 587–600, 1996.

[22] C. Villlani *Topics in Mass Transportation* Lecture notes from a course of the author, private communication.

[23] L.N. Wasserstein, *Markov processes over denumerable products of spaces describing large systems of automata*, Problems of Information Transmission, **5**, 47–52, 1969.