OPTIMAL TRANSPORT APPLIED TO GEOMETRIC AND FUNCTIONAL INEQUALITIES

BENJAMIN SCOTT

ABSTRACT. In recent years optimal transport has found many varied applications across pure and applied mathematics. In this report we will detail how the tools of optimal transport can be developed in order to establish two inequalities: one from geometry and one from functional analysis. In this pursuit, we will show how optimal transport can lend differential and convex structure to the space of absolutely continuous measure on $\mathbb{R}^d$, made possible by the optimal transport map.

CONTENTS

1. INTRODUCTION

While originating in physical problems of earth-moving and resource allocation, the techniques of optimal transport have found surprising pure applications in geometry, probability, and differential equations. This report will detail how optimal transport can be employed to establish two inequalities, one from geometry and one from functional analysis.

First, an aside on notation. In what follows X and Y will be two Polish spaces. We denote the space of signed Radon measures on X as $\mathscr{M}(X)$ and the space of Borel probability measures on X as $\mathscr{P}(X)$; in either case a subscript ac will indicate the subspace of absolutely continuous measures when the underlying space is $\mathbb{R}^d$, and we freely associate between absolutely continuous measures and their densities. We will also denote by $C_c(X)$ the space of compact, continuously supported functions, which in the case of X locally compact satisfies $C_c(X)^* = \mathscr{M}(X)$ by the Riesz representation theorem. [1]

Date: DEADLINES: Draft AUGUST 14 and Final version AUGUST 28, 2021.

[1] See [10, Theorem 6.19].

Now we are able to introduce the two ways transport between measures is understood.

Definition 1.1. Given $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, and a Borel map $T : X \to Y$, we define the *pushforward measure* $T_{\#}\mu$ by

$$T_{\#}\mu(B) = \mu(T^{-1}(B)), \text{ for all } B \subset Y \text{ Borel.}$$

We say T is a *transport map from* μ *to* ν if $T_{\#}\mu = \nu$.

Theorem 1.2 (pushforward change of variables)**.** *Let* $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$. *Then* $T_{\#}\mu = \nu$ *for* $T : X \to Y$ *if and only if*

$$\int_Y \varphi \, d\nu = \int_X \varphi \circ T \, d\mu$$

for all $\varphi : Y \to \mathbb{R}$ *Borel and bounded.*

Intuitively a transport map specifies a way to carry infinitesimal pieces of μ from each point $x \in X$ to a point $T(x) \in Y$ in such a way that ν results. Another notion of transport between measures is given by what is called a *coupling*.

Definition 1.3. Given $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, we say $\gamma \in \mathscr{P}(X \times Y)$ is a *coupling* of μ and ν if

$$\gamma(A \times Y) = \mu(A), \quad \gamma(X \times B) = \nu(B) \quad \text{for all } A \subset X,\, B \subset Y \text{ Borel}$$

or equivalently

$$\pi^X_{\#}\gamma = \mu, \quad \pi^Y_{\#}\gamma = \nu,$$

where $\pi^X : X \times Y \to X$ and $\pi^Y : X \times Y \to Y$ are the two projection maps. We denote the set of couplings of μ and ν by $\Gamma(\mu, \nu)$.

A coupling no longer specifies explicit destinations for each point in the source space. Instead, it encodes the amount of mass transported from a Borel set $A \subset X$ to a Borel set $B \subset Y$ by the value $\gamma(A \times B)$. This relaxation will resolve some difficulties encountered when ones tries to find transport between arbitrary measures, since it enables mass in the source to be "split."

We are now ready to introduce the two problems which lie at the foundation of optimal transport: those of Monge and Kantorovich. We let $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$, and suppose $c : X \times Y \to \mathbb{R}$ is a Borel function which when evaluated at (x, y) gives the unit cost to transport mass from $x \in X$ to $y \in Y$.

Problem 1.4 (Monge)**.** *Find a transport map* $\overline{T}$ *from* μ *to* ν *such that*

$$\int_X c(x, \overline{T}(x)) \, d\mu(x) = \inf \left\{ \int_X c(x, T(x)) \, d\mu(x) \,\middle|\, T_{\#}\mu = \nu \right\}.$$

Problem 1.5 (Kantorovich)**.** *Find a coupling* $\overline{\gamma}$ *of* μ *and* ν *such that*

$$\int_{X \times Y} c(x, y) \, d\overline{\gamma}(x, y) = \inf \left\{ \int_X c(x, y) \, d\gamma(x, y) \,\middle|\, \gamma \in \Gamma(\mu, \nu) \right\}.$$

Although the Monge problem is more appealing to the intuition it need not admit a solution. For instance, if $X = Y = \mathbb{R}$, $\mu = \delta_0$, and $\nu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_2$, then the set of transport maps from μ to ν is empty. The reason for this is that any Borel map $T : X \to Y$ can specify only one destination for the atom at 0. Additional assumptions can recover the guarantee of a solution in the Monge problem, but in the Kantorovich problem the set of couplings is always nonempty, since we can define the product measure $\mu \otimes \nu(A \times B) = \mu(A) \times \nu(B)$. For this reason we will perform our analysis in the setting of the Kantorovich problem then derive results in the Monge problem as a special case.

Strictly speaking, we will neglect the goal of minimizing the cost in favor of deriving transport maps with the greatest possible regularity. The reasoning for this is twofold. First, optimality in terms of cost is actually irrelevant to our stated goal of resolving functional and geometric inequalities. Second, the transport map we seek can be had with weaker assumptions than a solution to the Monge problem. Nevertheless, we will note in the text where an additional assumption bridges the gap and guarantees optimality.

In the subsequent sections we will care chiefly about the case $X = Y = \mathbb{R}^d$ and $c(x, y) = |x - y|^2$, and we will take as given m, the Lebesgue measure on the Borel sets. First, some useful definitions and results from convex analysis will be established. These ideas will then be applied to the question of existence and uniqueness in the problem of finding a transport map between two measures. After that, an alternative notion of convexity for functionals on the space of absolutely continuous measures will be developed. Finally, we will resolve the Brunn-Minkowksi and Sobolev inequalities.

2. Review of Convex Analysis

The theory of convex functions and their derivatives will form the basis of our investigations into solutions of the Monge and Kantorovich problems. Many of the classical differentiability results for sufficiently smooth functions can be realized in an analogous form for convex functions in an appropriately weakened sense. In this section we will issue some necessary definitions to realize the aforementioned differentiability results.

First, we will make precise our notion of convex function.

Definition 2.1. A function $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is said to be *convex* if

$$\varphi((1-t)x + ty) \leq (1-t)\varphi(x) + t\varphi(y)$$

whenever $x, y \in \mathbb{R}^d$ and $t \in (0, 1)$. The *domain* of a convex function, dom φ is the convex set $\{\varphi < +\infty\}$. We say φ is *closed* whenever its epigraph $\{(x, y) \mid y \geq \varphi(x)\}$ is closed or equivalently[1] when φ is lower semicontinuous.

Hereafter we will suppose $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is a (not necessarily closed) convex function. While φ is always continuous on int dom φ, we have no hope of ensuring differentiability at every point. However, we can relax the notion of gradient to that of *subgradient*, which describes a hyperplane lying below the graph at a given point.

Definition 2.2. Given $x \in \mathbb{R}^d$, we say $y \in \mathbb{R}^d$ is a *subgradient* of φ at x if

$$f(z) \geq f(x) + \langle z - x, y \rangle$$

for all $z \in \mathbb{R}$. The collection of all subgradients of φ at a point x defines a multifunction called the *subdifferential* of φ at x and is denoted $\partial\varphi(x)$. We identify the subdifferential with its graph in the following way

$$\partial\varphi = \bigcup_{x \in \mathbb{R}^d} \{x\} \times \partial\varphi(x).$$

[1] See [8, Theorem 7.1].

If φ is closed, then the subdifferential is continuous in the sense that for a sequence $((x_n, y_n)) \subset \partial\varphi$ with $x_n \to x$ and $y_n \to y$, we have $y \in \partial\varphi(x)$, which is to say the set $\partial\varphi$ is closed. Since φ is locally Lipschitz at every point of int dom φ, one can conclude by Rademacher's theorem and Definition 2.2 the following

Proposition 2.3. *For almost all $x \in$ dom φ, $\nabla\varphi(x)$ exists. Moreover, $\partial\varphi(x) = \{\nabla\varphi(x)\}$ if and only if $\nabla\varphi(x)$ exists.*

Proof. See [8, Theorem 25.1 and 25.5]. ■

In fact, we can go one step further and supply a condition for second differentiability of φ at the points where $\nabla\varphi$ exists. This is done by indentifying a second derivative $\nabla^2\varphi(x)$ at a point x where $\nabla\varphi(x)$ exists if

$$\lim_{h\to 0} \sup_{y\in\partial\varphi(x+h)} \frac{|y - \nabla\varphi(x) - \nabla^2\varphi(x)(h)|}{|h|} = 0.$$

This derivative, called the Aleksandrov derivative, leads to the following theorem.

Theorem 2.4 (Aleksandrov's)**.** *The Aleksandrov derivative $\nabla^2\varphi(x)$ exists for $x \in$ dom φ almost everywhere.*

Proof. See [4, Theorem 6.9]. ■

This next definition will give a simple but powerful way to characterize the subdifferential of convex functions.

Definition 2.5. We say a set $S \subset \mathbb{R}^d \times \mathbb{R}^d$ is *cyclically monotone* if for all $(x_1, y_1), \cdots, (x_n, y_n) \subset S$, we have

$$\sum_{i=1}^{n} \langle x_{i+1} - x_i, y_i \rangle \leq 0, \text{ where } x_{n+1} = x_1. \tag{2.6}$$

The analogy here with the smooth case is evident. Paraphrasing Rockafellar, the condition in (2.6) can be thought of as a discretization of the classical requirement that the gradient of a conservative vector field induced by a convex potential should have positive semidefinite derivative and closed line integrals over the field should vanish. Indeed, we can identify the subdifferential of a convex function by the cyclical monotonicity property.

Theorem 2.7 (Rockafellar's)**.** *The subdifferential of a convex function is cyclically monotone, and a cyclically monotone set is contained in the subdifferential of a closed convex function.*

Proof. See [8, Theorem 24.8]. ■

Another useful definition arises out of the need to define the inverse derivative and is called the *Legendre transform*.

Definition 2.8. The *Legendre transform* of φ, $\varphi^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is defined by

$$\varphi^*(y) = \sup_{x\in\mathbb{R}^d} \langle x, y \rangle - \varphi(x).$$

Some useful properties of φ^*, along with the inverse property we have just mentioned, are now introduced.

Proposition 2.9. *The following hold for $x, y \in \mathbb{R}^d$:*

(i) $\varphi(x) + \varphi^*(y) \geq x \cdot y$.
(ii) $\varphi(x) + \varphi^*(y) = x \cdot y$ *if and only if* $y \in \partial\varphi(x)$.
(iii) If φ *is closed, then* $y \in \partial\varphi(x)$ *if and only if* $x \in \partial\varphi^*(y)$.

Proof. See [8, Theorem 23.5] ∎

The preceeding definitions and results are all hallmarks of convex analysis. We now discuss a few of more specialized results. Each of these was proved by McCann in [6] or [7].

Theorem 2.10 (convex inverse function theorem)**.** *Let* $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *be convex. Suppose there is* $x_0 \in \mathbb{R}^d$ *such that* $\nabla\varphi(x_0)$ *exists,* φ *is finite in a neighborhood of* x_0*, and* φ *has Aleksandrov derivative* Λ *at* x_0*. If* Λ *is invertible, then* φ^* *has Aleksandrov derivative* Λ^{-1} *at* $\nabla\varphi(x_0)$*. If* Λ *is not invertible, then* φ^* *has no Aleksandrov derivative at* $\nabla\varphi(x_0)$*.*

Theorem 2.11 (convex implicit function theorem)**.** *Let* $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *and* $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *be convex. Suppose there is* $x_0 \in \mathbb{R}^d$ *such that* $\varphi(x_0) = \psi(x_0)$*, but* $\nabla\varphi(x_0) \neq \nabla\psi(x_0)$*. Assuming* $\nabla\varphi(x_0) - \nabla\psi(x_0)$ *is directed along* x_1*, without loss of generality, then there is a neighborhood* U *of* p *and a 1-Lipschitz function* $f : \mathbb{R}^{d-1} \to \mathbb{R}^d$ *on which* $\varphi(x) = \psi(x)$ *if and only if* $x_1 = f(x_2, \cdots, x_d)$*. It follows that* φ *and* ψ *are locally equal on a set of measure zero.*

Theorem 2.12 (Jacobian theorem)**.** *Let* $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *be convex. If* φ *has Aleksandrov derivative* Λ *at* x_0*, then*

$$\frac{m(\partial\varphi(B_r(x_0)))}{m(B_r)} \to \det \Lambda, \text{ as } r \to 0.$$

Moreover, if Λ *is invertible, then* $\partial\varphi(B_r(x_0))$ *shrinks nicely*[1] *to* $\nabla\varphi(x_0)$*.*

3. Existence and Characterization of the Optimal Map

We now turn our view towards establishing some results regarding solutions to the problems of Monge and Kantorovich. We will follow the geometric and topological method of McCann as in [6], which establishes the necessary result in greatest generality. Historically, this has also been approached via the factorization of vector fields or through Kantorovich duality.

Here we will solve[2] the Kantorovich problem for finite combinations of Dirac measures. Then we will demonstrate that such combinations form a dense subset of the space of probability measures under an appropriate topology. Passing into a limit will demonstrate existence, and uniqueness will follow as a result of Theorem 2.11. We begin with the first order of business, solving the problem of Kantorovich in the discrete case.

Proposition 3.1 (existence of a cyclically monotone pairing)**.** *Given* n *points* $x_1, \cdots, x_n$ *and* n *points* $y_1, \cdots, y_n$*, there exists a permutation* σ *such that the set* $\{(x_{\sigma(1)}, y_1), \cdots, (x_{\sigma(n)}, y_n)\}$ *is cyclically monotone.*

[1] A sequence of Borel sets (E_n) *shrinks nicely to* x if there is a sequence $r(n) \to 0$ such that $E_n \subset B_{r(n)}(x)$ and $\inf_n m(E_n)/m(B_{r(n)}) > 0$.

[2] We will use the word "optimal" to mean in the sense of cylically monotone support, since this property is much more relevant to regularity results on the optimal transport map. The relationship between this property and optimality in the cost will be demonstrated afterwards.

Proof. Since the set of permutations is finite, we can simply select σ such that

$$\sum_{i=1}^{n} \langle x_{\sigma(i)}, y_i \rangle \tag{3.2}$$

is maximized. Without loss of generality we will assume $\sigma = I$. Now we wish to demonstrate cyclical monotonicity.

Take k of the points $(x_{i_1}, y_{i_1}), \cdots, (x_{i_k}, y_{i_k})$, and let $i_{k+1} = i_1$. Now we define a permutation σ' as

$$\sigma'(i) = \begin{cases} i_{j+1}, & \text{if } i = i_j \text{ for some } 1 \leq j \leq k \\ i, & \text{otherwise.} \end{cases}$$

Since we chose σ to maximize (3.2), we have

$$\sum_{i=1}^{n} \langle x_{\sigma'(i)}, y_i \rangle - \sum_{i=1}^{n} \langle x_i, y_i \rangle = \sum_{j=1}^{k} \langle x_{i_{j+1}} - x_{i_j}, y_{i_j} \rangle \leq 0,$$

as desired. ■

Corollary 3.3. *In the case of* $\mu = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}, \nu = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$, *there exists a coupling with cyclically monotone support.*

Proof. We assume the points have been indexed so that $(x_1, y_1), \cdots, (x_n, y_n)$ forms the cyclically monotone pairing as above. Then we define $\gamma = \frac{1}{n} \sum_{i=1}^{n} \delta_{(x_i, y_i)}$. One can check then that γ is a coupling of μ and ν, and that its support is the cyclically monotone set $\{(x_1, y_1), \cdots, (x_n, y_n)\}$. ■

Before we go further, we should briefly make a note on the topology of measures. Since $\mathbb{R}^d$ is a locally compact Hausdorf space, we have $\mathscr{M}(\mathbb{R}^d) = C_c(\mathbb{R}^d)^*$, and as a result we can use the idea of weak-$*$ convergence.

Definition 3.4. Given a sequence $(\mu_n) \subset \mathscr{M}(\mathbb{R}^d)$, we say (μ_n) converges *weak-$*$* to a limit $\mu \in \mathscr{M}(\mathbb{R}^d)$, sometimes denoted $\mu_n \xrightarrow{*} \mu$, if

$$\int \varphi \, d\mu_n \to \int \varphi \, d\mu, \text{ for all } \varphi \in C_c(\mathbb{R}^d).$$

This notion of convergence induces what is called the weak-$*$ topology. It turns out that this topology can be metrized by the transportation cost for linear distance, $c(x, y) = |x - y|$ by what is known as the *Wasserstein distance.* Those interested should see [5] for more.

The following theorem will enable us to take the previous results and translate them to the continuous case through a limit process.

Lemma 3.5. *The following set forms a weak-$*$ dense subset of* $\mathscr{P}(\mathbb{R}^d)$

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \,\middle|\, n \in \mathbb{N},\ x_1, x_2, \cdots, x_n \in \mathbb{R}^d \right\}. \tag{3.6}$$

Proof. The set $\mathscr{P}(\mathbb{R}^d)$ is convex, and is weak-$*$ compact by the Banach-Alaoglu theorem.[1] Moreover, the Dirac measures form the set of extreme points for $\mathscr{P}(\mathbb{R}^d)$, therefore by the Krein-Milman theorem, we know the closure of the set of convex

[1]See [9, §15.1].

combinations of Dirac masses, $\sum_{i=1}^n a_i\delta_{x_i}$ is dense. One can take the coefficients a_i to be rational, then we can recover a form more like (3.6) by taking n to be the least common denominator across every a_i and encoding the numerator with appropriate multiplicity of the points x_i. ■

Now we demonstrate that the relevant properties of a sequence of couplings is preserved under weak-$*$ limits.

Lemma 3.7. *Let $(\gamma_n) \subset \mathscr{P}(\mathbb{R}^d \times \mathbb{R}^d)$ and suppose $\gamma_n \overset{*}{\rightharpoonup} \gamma \in \mathscr{P}(\mathbb{R}^d)$. If $\operatorname{supp}\gamma_n$ is cyclically monotone for all n, then $\operatorname{supp}\gamma$ is cyclically monotone. If $\mu_n \overset{*}{\rightharpoonup}$ and $\nu_n \overset{*}{\rightharpoonup} \nu$, where μ_n and ν_n are the left and right marginals of γ_n, then γ has left and right marginals μ and ν.*

Proof. We prove the first claim by contradiction. If γ did not have cyclically monotone support, then there would exist k points $(x_i, y_i) \subset \operatorname{supp}\gamma$ that violate the inequality in (2.6). Moreover, we can select neighborhoods U_i of (x_i, y_i) such that the same inequality is violated whenever we select k points, one from each of the neighborhoods U_i. Since these points belong to $\operatorname{supp}\gamma$, we know $\gamma(U_i) > 0$ for all U_i, which implies that for large enough n, $\gamma_n(U_i) > 0$ for all U_i. However, this implies that $\operatorname{supp}\gamma_n \cap (\bigcup_{i=1}^k U_i) \neq \emptyset$, and $\operatorname{supp}\gamma_n$ therefore contains a subset that is not cyclically monotone, in contradiction with our supposition. This proves the first claim.

Let $f \in C_c(\mathbb{R}^d)$. We can treat f as a function on $\mathbb{R}^d \times \mathbb{R}^d$ independent of its second argument and then extend appropriately to take $f \in C_c(\overline{\mathbb{R}}^d \times \overline{\mathbb{R}}^d)$, where $\overline{bbR}^d$ denotes the one point compactification of $\mathbb{R}^d$. Since (γ_n) embeds in $C_c(\overline{\mathbb{R}}^d \times \overline{\mathbb{R}}^d)^*$, we can extract a weak-$*$ limit point of (γ_n), γ', in $C_c(\overline{\mathbb{R}}^d \times \overline{\mathbb{R}}^d)^*$. Since $C_c(\overline{\mathbb{R}}^d \times \overline{\mathbb{R}}^d)$ contains the constant functions, we know by the definition of weak-$*$ convergence, $\gamma'(\overline{\mathbb{R}}^d \times \overline{\mathbb{R}}^d) = 1$. Additionally, γ and γ' coincide on $\mathbb{R}^d \times \mathbb{R}^d$, which can be checked by an ascending limit of $C_c(\mathbb{R}^d \times \mathbb{R}^d)$ functions converging pointwise to the indicator of $\mathbb{R}^d \times \mathbb{R}^d$. Therefore, $\gamma = \gamma'$. If $f \in C_c(\overline{\mathbb{R}}^d \times \overline{\mathbb{R}}^d)^*$ is independent of its second argument and supported on a compact subset of $\mathbb{R}^d$ in its first, then

$$\int f\, d\gamma' = \lim_{n\to\infty} \int f\, d\gamma_n = \lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu.$$

Using Urysohn's Lemma and taking limits,[1] we can deduce that $\gamma(A \times \overline{\mathbb{R}}^d) = \mu(A)$ for all $A \subset \mathbb{R}^d$ compact,[2] which is to say μ is the left marginal of γ'. The same proof can be carried out to check that ν is the right marginal of γ'. It follows then that $\gamma = \gamma'$ has the correct marginals. ■

Theorem 3.8. *Let $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$, then there exists a coupling γ of μ and ν with cyclically monotone support.*

Proof. Using Lemma 3.6, one can obtain sequences

$$\mu_k = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \quad \nu_k = \frac{1}{n}\sum_{i=1}^n \delta_{y_i},$$

[1] That is, taking continuous functions $f : \mathbb{R}^d \to [0,1]$ such that $f(A) = \{1\}$ and $f(\{d(x,A) \geq 1/n\}) = 0$.

[2] In $\mathbb{R}^d$ the Borel σ-algebra is also generated by the compact sets.

where (x_i, y_i) is ordered such that the support of

$$\gamma_k = \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_i,y_i)}$$

is cyclically monotone and γ_k has marginals $\mu_k \overset{*}{\rightharpoonup} \mu$ and $\nu_k \overset{*}{\rightharpoonup} \nu$. Now γ_k lies in the unit ball of $C_c(\mathbb{R}^d \times \mathbb{R}^d)^*$, so extracting a limit point γ tells us that γ is a nonnegative measure with cyclically monotone support and μ and ν as its marginals. This last fact forces γ to have unit mass, as desired. ■

The next theorem will be central to the our stated goal of proving various inequalities, since it equips the optimal transport map with a great deal of regularity.

Theorem 3.9 (Brenier's)**.** *Let $\mu \in \mathscr{P}_{ac}(\mathbb{R}^d)$, $\nu \in \mathscr{P}(\mathbb{R}^d)$. Then there exists a unique (μ almost everywhere) optimal transport map from μ to ν given by the gradient of a convex function $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$.*

Proof. We will address existence. From Theorem 3.8, we have that there exists γ with cyclically monotone support and μ and ν as its marginals. Rockafellar's theorem tells us that $\operatorname{supp}\gamma \subset \partial\varphi$, for some closed convex function $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. Therefore $\partial\varphi$ is closed. The set of points where φ is differentiable, dom $\nabla\varphi$ is a G_δ set,[1] which is to say it is Borel. Moreover, since the boundary of dom φ is measure zero, we know because $\nabla\varphi$ exists in dom φ almost everywhere and μ is absolutely continuous that $\mu(\text{dom } \nabla\varphi) = 1$.

Now we should demonstrate that $(I \times \nabla\varphi)_{\#}\mu = \gamma$. Consider the set

$$S = \{(x, \nabla\varphi(x)) \mid x \in \text{dom } \nabla\varphi\} = (\text{dom}\nabla\varphi \times \mathbb{R}^d) \cap \partial\varphi.$$

Since $\partial\varphi$ is closed and dom $\nabla\varphi$ is Borel, we know S is Borel. Moreover, S is of full measure for γ, since $\mu(\text{dom } \nabla\varphi) = 1$ and $\operatorname{supp}\gamma \subset \partial\varphi$, so $\gamma(Z \cap S) = \gamma(Z)$ for $Z \subset \mathbb{R}^d \times \mathbb{R}^d$. This implies that for $M, N \subset \mathbb{R}^d$ Borel

$$\gamma(M \times N) = \gamma((M \cap (\nabla\varphi)^{-1}(N)) \times \mathbb{R}^d \cap S) = (I \times \nabla\varphi)_{\#}\mu(M \times N).$$

Since the Borel sets in the product space are generated by products of Borel sets, we have $(I \times \nabla\varphi)_{\#}\mu = \gamma$, and therefore $\nabla\varphi_{\#}\mu = \nu$, as desired.

Uniqueness is more laborious and a full proof can be found in [6]. ■

We will now relate the property of cylical monotonicity to that of optimality in the cost. The reason we have neglected to do so until now is that cyclical monotonicity, and the structure it places on the optimal transport map, is all one needs to deal with functional and geometric inequalities. However, to derive optimality results one requires a further assumption: thatf finite second order moments, i.e.

$$\int |x|^2 \, \mu(x) + \int |y| \, \nu(y) < +\infty.$$

First, a necessity result.

Proposition 3.10. *If γ is optimal for the Kantorovich cost, then* supp γ *is cyclically monotone.*

Through the following duality, we can make this condition into a sufficient one.

[1] See [8, Note after Theorem 25.5].

Theorem 3.11 (Kantorovich Duality)**.** *Let $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$. If μ and ν have finite second order moments, then*

$$\min_{\gamma \in \Gamma(\mu,\nu)} \int -x \cdot y \, d\gamma(x,y) = \max_{\varphi \text{ convex}} \int -\varphi \, d\mu + \int -\varphi^* \, \nu.$$

These results culminate in the following theorem.

Corollary 3.12 (fundamental theorem of optimal transport)**.** *The following are equivalent for the Kantorovich problem with μ, ν having finite second moments:*

(i) γ *is optimal*
(ii) $\operatorname{supp} \gamma$ *is cyclically monotone*
(iii) $\operatorname{supp} \gamma \subset \partial\varphi$ *for some convex* $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$.

Towards the end of establishing the regularity on the optimal transport map when μ and ν are absolutely continuous, we will now follow the program of McCann in [7]. This will culminate in an integral change of variables type theorem between μ and ν, along with a differential equation relating them, both based on the derivative of the optimal transport map.

Lemma 3.13. *Let $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be convex and let $\Omega = int\ dom\ \Omega$. If $\mu \in \mathscr{M}_{ac}(\Omega)$, then for a Borel set $M \subset \mathbb{R}^d$, $\nabla\varphi_{\#}\mu(M)$ is equal to $\mu(\partial\varphi^*(M))$.*

Proof. By definition, $\nabla\varphi_{\#}\mu(M) = \mu((\nabla\varphi)^{-1}(M))$. If $\nabla\varphi(x) \in M$, then $x \in \partial\varphi^*(\nabla\varphi(x)) \subset \partial\varphi^*(M)$, so $(\nabla\varphi)^{-1}(M) \subset \partial\varphi^*(M)$. Moreover, the difference is a subset of the points where $\partial\varphi$ is not single valued, which is a set of Lebesgue (and thereby μ) measure zero, therefore

$$\nabla\varphi_{\#}\mu(M) = \mu((\nabla\varphi)^{-1}(M)) = \mu(\partial\varphi^*(M)). \qquad \blacksquare$$

Proposition 3.14. *Let $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be convex, let $\Omega = int\ dom\ \varphi$, and let $\mu \in \mathscr{M}_{ac}(\Omega)$. Suppose x is a Lebesgue point of μ at which is the Aleksandrov derivative $\nabla^2\varphi$ exists and is invertible. Then at $\nabla\varphi(x)$, the symmetric derivative of $\nabla\varphi_{\#}\mu$ exists and is equal to $\mu \det(\nabla^2\varphi)^{-1}$.*

Proof. If $x \notin \Omega$, then the result is automatic since μ vanishes in a neighborhood of x, so we may assume $x \in \Omega$. Then φ is finite in a neighborhood of x, $\nabla\varphi(x)$ exists, and $\nabla^2\varphi(x)$ exists and is invertible, therefore by Theorem 2.10 we have that $\nabla^2\varphi^*(\nabla\varphi(x)) = (\nabla^2\varphi(x))^{-1}$, which combined with Theorem 2.12 gives

$$\frac{m(\partial\varphi^*(B_r(\nabla\varphi(x))))}{m(B_r(\nabla\varphi(x)))} \to \det(\nabla^2\varphi(x))^{-1}, \text{ as } r \to 0. \tag{3.15}$$

Moreover, since $\partial\varphi^*(B_r(\nabla\varphi(x)))$ shrinks nicely to x and x is a Lebesgue point of μ, we have

$$\frac{\mu(\partial\varphi^*(B_r(\nabla\varphi(x))))}{m(\partial\varphi^*(B_r(\nabla\varphi(x))))} \to \mu(x), \text{ as } r \to 0. \tag{3.16}$$

Multiplying (3.15) and (3.16) then applying the identity in Lemma 3.13 yields the result. $\blacksquare$

Corollary 3.17. *Let $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be convex and let $\Omega = int\ dom\ \Omega$. Then $\det(\nabla^2\varphi) \in L^1_{loc}(\Omega)$. Moreover, the pushforward of $\det(\nabla^2\varphi)$ through $\nabla\varphi$ is equal to Lebesgue measure restricted to $\partial\varphi(M)$, where M consists of points x such that the Aleksandrov derivative exists, is invertible, and x is a Lebesgue point of $\det(\nabla^2\varphi)$.*

Proof. Let $K \subset \Omega$ be compact. Since $\partial\varphi$ is continuous, by Lemma 3.13 we know that $\nabla\varphi^*_\# m(K) = m(\partial\varphi(K)) < \infty$, therefore $\nabla\varphi^*_\# m$ is Radon and its absolutely continuous part is locally integrable on Ω. On $M \subset \Omega$, where $\nabla^2\varphi$ is invertible, we can apply Proposition 3.14 to the measure $\nabla\varphi^*_\# m$ along with Theorem 2.10 to find that its absolutely continuous part is equal to $\det(\nabla^2\varphi)$. Since $\det(\nabla^2\varphi)$ is zero outside M, we find that it is equal to a locally integrable function almost everywhere, and is thus locally integrable.

Now if we apply Proposition 3.14 to $\nabla\varphi_\# \det(\nabla^2\varphi)$, we find that the symmetric derivative is equal to 1 on $\partial\varphi(M)$, and because M is full measure for $\det(\nabla^2\varphi)$, $\partial\varphi(M)$ must be full measure for $\nabla\varphi_\# \det(\nabla^2\varphi)$. Thus, $\nabla\varphi_\# \det(\nabla^2\varphi)$ is equal to Lebesgue measure restricted to $\partial\varphi(M)$. ∎

Theorem 3.18 (monotone change of variables)**.** *Let $\mu, \nu \in \mathscr{P}_{ac}(\mathbb{R}^d)$, and let $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function such that $\nabla\varphi_\#\mu = \nu$. Let $\Omega = int\ dom\ \varphi$, and let $X \subset \Omega$ be the points where the Alexandrov derivative $\nabla^2\varphi$ exists and is invertible. Then X is of full measure for μ. Moreover, if $F : [0,\infty) \to \mathbb{R}$ is measurable and $F(0) = 0$, then*

$$\int_{\mathbb{R}^d} F(\mu(y))\, dy = \int_X F\left(\frac{\mu(x)}{\det(\nabla^2\varphi(x))}\right) \det(\nabla^2\varphi(x))\, dx.$$

Proof. Since $\nabla\varphi_\#\mu = \nu$, we know φ is finite on a set of full measure for μ, therefore $\nabla^2\varphi$ exists μ almost everywhere, and is invertible but for the set $\partial\varphi^*(Z)$, where $Z = \{y \mid \nabla^2\varphi^*(y) \text{ does not exist}\}$. Since ν is absolutely continuous, we have by Lemma 3.13 that $\mu(\partial\varphi^*(Z)) = \nu(Z) = 0$, so $\mu(X) = 1$. This proves the first result.

Let M be the set of Lebesgue points for $\det(\nabla^2\varphi)$. Since $\det(\nabla^2\varphi)$ is $L^1_{loc}(\Omega)$, we know M and X differ on a set of Lebesgue (and thereby μ) measure zero. Thus, $\partial\varphi(M)$ has full measure for ν. Since $\nabla\varphi$ pushes $\det(\nabla^2\varphi)$ foward to Lebesgue measure on $\partial\varphi(M)$, the pushforward change of variables

$$\int_{\partial\varphi(M)} F(\nu(y))\, dy = \int_M F(\nu(\nabla\varphi(x))) \det(\nabla^2\varphi(x))\, dx. \tag{3.19}$$

Taking ν to coincide with its symmetric derivative, we find by Proposition 3.14 that $\nu(\nabla\varphi(x)) = \mu(x)/\det(\nabla^2\varphi(x))$ on X. Thus, using the fact that $F(0) = 0$ to change domains of integration, we find

$$\int_{\mathbb{R}^d} F(\nu(y))\, dy = \int_X F\left(\frac{\mu(x)}{\det(\nabla^2\varphi(x))}\right) \det(\nabla^2\varphi(x))\, dx. \qquad \blacksquare$$

The first part of the proof in Theorem 3.18 and an application of Proposition 3.14 gives the following.

Theorem 3.20 (Monge-Ampere equation holds almost everywhere)**.** *Let $\mu, \nu \in \mathscr{P}_{ac}(\mathbb{R}^d)$ and let φ be the convex map such that $\nabla\varphi_\#\mu = \nu$. Then the Monge-Ampere equation*

$$\nu(\nabla\varphi(x)) \det(\nabla^2\varphi(x)) = \mu(x)$$

holds for $x \in \mathbb{R}^d$ μ-almost everywhere, where $\nabla^2\varphi$ is taken in the Aleksandrov sense.

4. Displacement Interpolation and Displacement Convexity

In this section we will describe an alternate notion of interpolation between measures in the space $\mathscr{P}_{ac}(\mathbb{R}^d)$. This definition was originally introduced by McCann in [7]. There he sought to investigate the behavior of a cloud of gas in $\mathbb{R}^d$ modeled by a probability density ρ subject to an energy functional of the form

$$E(\rho) = \underbrace{\int_{\mathbb{R}^d} U(\rho(x))\, dx}_{\mathcal{U}(\rho)} + \underbrace{\frac{1}{2}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} V(x-y)\, d\rho(x)d\rho(y)}_{\mathcal{V}(\rho)},$$

where $\mathcal{U}$ describes the internal energy of the gas cloud and $\mathcal{V}$ its interaction energy. Under certain assumptions, he was able to derive a convex structure of E, called *displacement convexity*, which was instrumental in showing that the gas cloud possesses a unique ground state. For our purposes this same definition can furnish a simple proof of the Brunn-Minkowski inequality. It is this definition we will develop now.

A simple example illustrating why an alternate definition may be appealing is as follows. Let $x, y \in \mathbb{R}^d$. If we select an attractive interaction potential such as $V(r) = |r|^2$, then $\mathcal{V}$ will not be convex on $\mathscr{P}(\mathbb{R}^d)$. That is, $\mathcal{V}(\delta_x) = \mathcal{V}(\delta_y) = 0$, but $\mathcal{V}((1-t)\delta_x + t\delta_y) > 0$ for all $t \in (0,1)$. The interpolation we get from $(1-t)\delta_x + t\delta_y$ is of course nonphysical: it sees mass traveling in an instant from x to y. If we interpret δ_x as a particle residing at x, then a natural way to carry it to y is linearly, i.e. via $\delta_{(1-t)x+ty}$ for $t \in [0,1]$. This interpolant will minimize $\mathcal{V}$ just as δ_x and δ_y did since $\mathcal{V}$ is translation invariant. This takes advantage of the linearity of the ambient space over the linearity of the space of measures, and turns out to be the appropriate generalization.

We begin with a definition.

Definition 4.1 (displacement interpolant and convexity)**.** Given $\mu, \nu \in \mathscr{P}_{ac}(\mathbb{R}^d)$, Theorem 3.9 implies that there is a unique map $\nabla\varphi$ given by the gradient of a convex function such that $\nabla_{\#}\mu = \nu$. Then the *displacement interpolant between μ and ν* is the measure

$$[\mu, \nu]_t = ((1-t)I + t\nabla\varphi(x))_{\#}\mu.$$

We say a functional $F : \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}$ is *displacement convex* if for any $\mu, \nu \in \mathscr{P}_{ac}(\mathbb{R}^d)$

$$t \to F([\mu, \nu]_t)$$

is convex for $t \in [0,1]$.

As was hinted at before, this notion of interpolation allows us to establish the displacement convexity of $\mathcal{V}$.

Theorem 4.2 (displacement convexity of interaction energy)**.** *Let $\mu, \nu \in \mathscr{P}_{ac}(\mathbb{R}^d)$, and let γ_t denote the displacement interpolant along the convex map φ such that $\nabla\varphi_{\#}\mu = \nu$. If V is convex, then $\mathcal{V}$ is displacement convex.*

However, displacement convexity of $\mathcal{U}$ will require more effort. First we should establish some basic facts of the displacement interpolant.

Proposition 4.3. *Given $\mu, \nu \in \mathscr{P}_{ac}(\mathbb{R}^d)$. For $t \in [0,1]$ and $\gamma_t = [\mu, \nu]_t$ we have*

(i) $\gamma_0 = \mu$ *and* $\gamma_1 = \nu$.

(ii) $\gamma_t \in \mathscr{P}_{ac}(\mathbb{R}^d)$.

Proof. See [7, Proposition 1.3]. ∎

Lemma 4.4. *Let Λ be a nonnegative symmetric $d \times d$ matrix, and let $v(t) = \det((1-t)I + t\Lambda)$. Then $v^{1/d}$ is concave.*

Proof. Since the invertible matrices are dense in the space of square matrices, we can assume Λ is invertible. Since $v^{1/d}$ is continuous, we need only show the mean value property and the rest follows by the density of dyadic rationals.

Since Λ is nonnegative, symmetric, and invertible, we know it admits a diagonalization

$$\Lambda = Q \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_d \end{pmatrix} Q^{-1},$$

where $\lambda_1, \lambda_2, \cdots, \lambda_d \geq 0$ are the eigenvalues of Λ. Therefore, if $s, t \in [0,1]$ and $s < t$, then

$$\begin{aligned} \det((1 - \tfrac{s+t}{2})I + \tfrac{s+t}{2}\Lambda)^{1/d} &= \prod_{i=1}^{d} (1 - \tfrac{s+t}{2} + \tfrac{s+t}{2}\lambda_i)^{1/d} \\ &\geq \prod_{i=1}^{d} \left(\tfrac{1}{2}(1 - s + s\lambda_i)^{1/d} + \tfrac{1}{2}(1 - t + t\lambda_i)^{1/d} \right) \\ &\geq \tfrac{1}{2} \prod_{i=1}^{d} (1 - s + s\lambda_i)^{1/d} + \tfrac{1}{2} \prod_{i=1}^{d} (1 - t + t\lambda_i)^{1/d} \\ &= \tfrac{1}{2} \det((1-s)I + s\Lambda)^{1/d} + \tfrac{1}{2} \det((1-t)I + t\Lambda)^{1/d}. \end{aligned}$$ ∎

Theorem 4.5 (displacement convexity of internal energy)**.** *Let $\mu, \nu \in \mathscr{P}_{ac}(\mathbb{R}^d)$, and let $\gamma_t = [\mu, \nu]_t$ denote the displacement interpolant along the optimal transport map $\nabla\varphi$. If $r \mapsto r^d U(r^{-d})$ is convex and nonincreasing on $[0, +\infty)$, then $\mathcal{U}$ will be displacement convex.*

Proof. Let X denote the set where the Aleksandrov derivative $\nabla^2\varphi(x)$ is defined and invertible. Theorem 3.18 implies that X has full measure in μ and because $U(0) = 0$,

$$\mathcal{U}(\gamma_t) = \int_X U \left(\frac{\mu(x)}{\det((1-t)I + t\nabla^2\varphi(x))} \right) \det((1-t)I + t\nabla^2\varphi(x))\, dx. \tag{4.6}$$

A small subtlety to note is that the integral above should be taking place over the set where $(1-t)I + t\nabla^2\varphi(x)$ is invertible as per Theorem 3.18, which is potentially larger than X. However, since μ and thereby $U(\mu/\det((1-t)I + t\nabla^2\varphi))$ die outside of X but for a set of measure zero, the two are equal can neglect this consideration.

Let $x \in X$, $\Lambda = \nabla^2\varphi(x)$, and v as in the statement of Lemma 4.4. Then $v^{1/d}$ is concave and $r \to r^d U(r^{-d}\mu(x))$ is convex and nonincreasing, therefore their composition, which is exactly the integrand in (4.6), is convex for $t \in [0,1]$. Integrating with respect to x yields that $t \to \mathcal{U}(\gamma_t)$ is convex as desired. ∎

The sufficient conditions just given for the two energy functionals to be displacement convex both turn out to be necessary as well. Those interested should see [11].

5. Brunn-Minkowski Inequality

With the major results of optimal transport and a criteria for displacement convexity in hand, we are now able to turn our view towards establishing a result in the class of geometric inequalities.

Theorem 5.1 (Brunn-Minkowski inequality)**.** *Let $K, K' \subset \mathbb{R}^d$ be compact. Then*

$$m(K+K')^{1/d} \geq m(K)^{1/d} + m(K')^{1/d},$$

where $K + K'$ denotes the Minkowski sum $\{k + k' \mid k \in K,\ k' \in K'\}$.

Proof. We start by fixing the measures $\mu = m(K)^{-1}\chi_K$ and $\nu = m(K')^{-1}\chi_{K'}$. It follows that $\mu \in \mathscr{P}_{ac}(\mathbb{R}^d)$ and $\nu \in \mathscr{P}(\mathbb{R}^d)$. Therefore, we can define $\gamma_t = [\mu, \nu]_t$ to be the displacement interpolant between μ and ν.

Now we wish to demonstrate $\operatorname{supp} \gamma_t \subset (1-t)K + tK'$. Since $\nabla\varphi_{\#}\mu = \nu$, we know $\nabla\varphi(x) \in K'$ for almost all $x \in K$. Therefore, $((1-t)I + t\nabla\varphi)(x) = (1-t)x + t\nabla\varphi(x) \in (1-t)K + tK'$ for almost all $x \in K$, so $\gamma_t((1-t)K + tK') = 1$. It follows by the definition of γ_t that $\operatorname{supp} \gamma_t \subset (1-t)K + tK'$, since the latter is compact.

Let $\mathcal{U} : \mathscr{P}_{ac}(\mathbb{R}^d) \to \mathbb{R}$ be defined by

$$\mathcal{U}(\rho) = \int_{\mathbb{R}^d} U(\rho(x))\, dx, \quad U(y) = -y^{(d-1)/d}.$$

The choice of U is evident after some reflection: $\mathcal{U}(\mu) = -m(K)^{1/d}$ and the negative sign ensures Theorem 4.5 applies. Therefore, by the dispacement convexity of $\mathcal{U}$,

$$\mathcal{U}(\gamma_t) \leq -(1-t)m(K)^{1/d} - tm(K')^{1/d}, \tag{5.2}$$

and by Jensen's inequality,

$$\mathcal{U}(\gamma_t) \geq -m(\operatorname{supp}\gamma_t)\left(\frac{1}{m(\operatorname{supp}\gamma_t)}\int \gamma_t\right)^{(d-1)/d} = -m(\operatorname{supp}\gamma_t)^{1/d}. \tag{5.3}$$

It follows combining (5.2) and (5.3) that

$$m(\operatorname{supp}\gamma_t)^{1/d} \geq (1-t)m(K)^{1/d} + tm(K')^{1/d},$$

and since $\operatorname{supp} \gamma_t \subset (1-t)K + tK'$, we can conclude

$$m((1-t)K + tK')^{1/d} \geq (1-t)m(K)^{1/d} + tm(K')^{1/d}. \qquad \blacksquare$$

6. Review of Weak Derivatives and Sobolev Spaces

To understand the discussion in the next section, one needs to understand the functional analytic notions of weak derivatives and the spaces they induce. These ideas and more will be introduced in this section.

The inspiration for the definition of weak derivative comes from the integration by parts formula. That is, if $u, v : \mathbb{R} \to \mathbb{R}$ are two continuously differentiable functions, then for any $a, b \in \mathbb{R}$

$$\int_a^b uv' = uv|_a^b - \int_a^b u'v.$$

If we assume u and v have compact support, then

$$\int uv' = -\int u'v.$$

This formula, when treated as definition of differentiability, allows us to expand the class of functions which are differentiable, while retaining many familiar properties of the normal derivative.

Definition 6.1. Given $u \in L^1_{loc}(\mathbb{R}^d)$. We say a function $v \in L^1_{loc}(\mathbb{R}^d)$ is a *weak partial derivative* of u in the x_i direction if for all smooth compactly supported functions $\varphi : \mathbb{R}^d \to \mathbb{R}$,

$$\int u \partial_{x_i} \varphi \, dx = - \int v \varphi \, dx.$$

The weak partial derivative in the x_i direction is unique almost everywhere, so we denote it by $\partial_{x_i} u$.

It can be shown, as was done in [3], that the weak partial derivative satisfies a number of desirable properties. In particular, it is uniquely defined almost everywhere, it is linear, and the space of integrable functions possessing integrable weak derivatives

$$W^{1,p} = \{f \in L^p(\mathbb{R}^d) \mid \partial_{x_i} f \text{ exists and is in } L^p(\mathbb{R}^d) \text{ for all } 1 \leq i \leq d\}$$

equipped with the norm

$$\|u\|_{W^{1,p}} = \|u\|_{L^p} + \sum_{i=1}^{d} \|\partial_{x_i} u\|_{L^p}$$

forms a Banach space.

This last property is an improvement over a classical space like C^∞, which may fail to be complete under uniform convergence. Importantly, the completeness of the Sobolev space allows us to derive results by a limiting process, all the while preserving regularity. We will now discuss a few useful results which will be instrumental in proving the Sobolev inequality, which allows one to embed $W^{1,p}$ in some higher order Lebesgue space.

Theorem 6.2 (Meyers-Serrin)**.** *Let $u \in W^{1,p}$. Then there exists a sequence $(u_n) \subset C^\infty \cap W^{1,p}$ such that $u_n \to u$ in $W^{1,p}$.*

There exists an even more general notion of derivative, which discards the requirement that the derivative of a function be a function itself. This leads to the idea of *distributional derivative*, which is a linear functional on the space of smooth compactly supported functions that satisfies an analogous "integration by parts" formula, but we will not pursue this idea to any major lengths. Instead, we will mention a special case, which is of the distributional second partial derivative of a convex function, which will take the form of a signed Radon measure.

Theorem 6.3 (Hessian measure of convex function)**.** *Let $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be convex. Then there exist signed Radon measures $\mu^{ij} = \mu^{ji}$ such that for all i, j,*

$$\int f \partial_{x_i x_j} \varphi \, dx = \int \varphi \, d\mu^{ij}.$$

Moreover, the Aleksandrov derivative of f coincides with the absolutely continuous part of the matrix valued measure given by the entries μ^{ij}.

Proof. See [4, Theorem 6.8] ■

This result is crucial for the Sobolev inequality, as it will allow us to take advantage of the regularity of the optimal transport map established by Brenier's theorem.

7. Sobolev Inequality

We finally turn our attention to the last result of this report, the Sobolev inequality. We follow the method of [2], making use of the results developed in Section 3, specifically the differential structure of the monotone rearrangement map.

In what follows we will take $1 \leq p < d$, and we define the Holder conjugate of p, $p' = p/(p-1)$, and the Sobolev conjugate of p, $p^* = dp/(n-p)$. Moreover, we define the function h_p by

$$h_p(x) = \frac{1}{(\sigma + |x|^{p'})^{(d-p)/p}},$$

where the constant σ has been selected so that $\|h_p\|_{L^{p^*}} = 1$. The functions h_p will form the extremal functions in the Sobolev inequality to follow and will yield a sharp constant therein.

Theorem 7.1 (Sobolev inequality)**.** *Let $p \in (1, d)$. If $f, g \in L^{p^*}(\mathbb{R}^d)$ are two functions satisfying $\|f\|_{L^{p^*}} = \|g\|_{L^{p^*}}$ and $\nabla f \in L^p(\mathbb{R}^d)$, then*

$$\frac{\int |g(y)|^{p^*(1-1/d)}\, dy}{\left(\int |y|^{p'} |g(y)|^{p^*}\, dy\right)^{1/p'}} \leq \frac{p(d-1)}{d(d-p)} \|\nabla f\|_{L^{p^*}},$$

with equality achieved when $f = g = h_p$.

Proof. We begin by reducing the problem in a few ways. Since $|\nabla f| = |\nabla |f||$, we can assume $f, g \geq 0$. From Theorem 6.2, we can assume $f, g \in C_c^\infty(\mathbb{R}^d)$. Finally, we can normalize $\|f\|_{L^{p^*}} = \|g\|_{L^{p^*}} = 1$.

Define the following probability densities

$$F(x) = f^{p^*}(x), \quad G(y) = g^{p^*}(y), \tag{7.2}$$

and take φ to be the convex map such that $\nabla\varphi_\# F = G$. By the definition of pushforward measure we have

$$\int G(y) G(y)^{-1/d}\, dy = \int F(x) G(\nabla\varphi(x))^{-1/d}\, dx,$$

and the Monge-Ampere equation implies

$$\int G(y) G(y)^{-1/d}\, dy = \int F(x) F(x)^{-1/d} \det(\nabla^2 \varphi(x))\, dx.$$

Now we use the AM-GM inequality to get

$$\int G^{1-1/d} \leq \frac{1}{d} \int F^{1-1/d} \Delta_A \varphi.$$

The Alexandrov Laplacian, $\Delta_A \varphi$, which is the sum of the diagonal of the Alexandrov matrix, forms the absolutely continuous part of the distributional Laplacian, $\Delta_{D'}\varphi$, itself being the trace of the matrix valued Hessian measure. Therefore, we only increase the value of the integral by replacing $\nabla_A \varphi$ with $\nabla_{D'}\varphi$, so

$$\int G^{1-1/d} \leq \frac{1}{d} \int F^{1-1/d} \Delta_{D'} \varphi.$$

Now we can apply the integration by parts formula in each coordinate,

$$\int G^{1-1/d} \leq -\frac{1}{d} \int \langle \nabla (F^{1-1/d}), \nabla \varphi \rangle,$$

substitute based on our definitions in (7.2),

$$\int g^{p(d-1)/(d-p)} \leq -\frac{p(d-1)}{d(d-p)} \int f^{p^*/p'} \langle \nabla f, \nabla\varphi \rangle,$$

and apply Holder's inequality in each coordinate to find.

$$-\int f^{p^*/p'} \langle \nabla f, \nabla\varphi \rangle \leq \|\nabla f\|_{L^p} \left(\int f^{p^*} |\nabla\varphi|^{p'} \right)^{1/p'}. \tag{7.3}$$

Again by the definition of pushforward measure,

$$\int f^{p^*} |\nabla\varphi|^{p'} = \int |y|^{p'} g^{p^*}(y)\, dy. \tag{7.4}$$

Combining (7.3) and (7.4) proves the inequality.

To check the claim that h_p yields equality, one should simply substitute $f = g = h_p$ and verify equality at every step. We can no longer treat h_p as a test function in the steps involving distributions, but they follow regardless due to the fact that $\nabla\varphi = I$. ■

Acknowledgments

I would like to thank my mentor, Antonis Zitridis, who introduced me to the rich field of optimal transport and suggested the idea of pursuing mass-transport methods for inequalities as a research topic. His advice was also helpful for getting familiar with the necessary background knowledge to support my readings in optimal transport. I also would like to thank Professor J. Peter May for organizing the REU. His efforts made my first experience conducting research in mathematics an enjoyable and enriching one.

References

[1] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer New York, 2010.
[2] D. Cordero-Erausquin, B. Nazaret, and C. Villani. A mass-transportation approach to sharp sobolev and gagliardo–nirenberg inequalities. *Advances in Mathematics*, 182(2):307–332, 2004.
[3] L.C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 2010.
[4] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions, Revised Edition*. Textbooks in Mathematics. CRC Press, 2015.
[5] A. Figalli and F. Glaudo. *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows*. EMS Press, Aug 2021.
[6] Robert J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309 – 323, 1995.
[7] Robert J. McCann. A convexity principle for interacting gases. *Advances in Mathematics*, 128(1):153–179, 1997.
[8] R.T. Rockafellar. *Convex Analysis: (PMS-28)*. Princeton Landmarks in Mathematics and Physics. Princeton University Press, 2015.
[9] H.L. Royden and P. Fitzpatrick. *Real Analysis*. Prentice Hall, 2010.
[10] Walter Rudin. *Real and Complex Analysis*. Higher Mathematics Series. McGraw-Hill Education, 1987.
[11] Cédric Villani. *Topics in Optimal Transportation*. American Mathematical Society, Mar 2003.