Lipschitz Optimal Transport Maps

Davido Zhang

Under the direction of

Shrey Aryan
Department of Mathematics
Massachusetts Institute of Technology

Research Science Institute
July 30, 2024

Abstract

In optimal transport, Caffarelli famously constructed a Lipschitz map between two log-concave densities $f = e^{-V}$ and $g = e^{-W}$ using the quadratic cost $c(x, y) = -xy$. We considered variations of this problem. Firstly, we built foundations for extending this result to the transport between unequal dimensions $T : \mathbb{R}^n \to \mathbb{R}$. We found an analog to the Monge-Ampère Equation and conditions that guarantee continuity to its unique solution. Secondly, we recovered Caffarelli's result for the perturbed cost function $c(x, y) = -xy - \varepsilon|x|^2|y|^2$, which suggests the possibility of constructing contraction maps for a greater variety of costs.

Summary

Optimal transport theory studies how to minimize the total cost of transporting masses from a source to a target. Particularly, researchers became interested in settings where this optimal transportation contracts, meaning that after any two points are mapped to the target region, the distance between them shrinks. In the past, studies have analyzed scenarios where the cost equals the distance moved from source to target and were able to construct contracting maps. We modified this cost function by adding small error terms and recovered the same contracting map. As optimal transport is closely related to differential geometry, our contracting maps provide insights on how to smoothly transform between surfaces. Whereas in the other sciences, optimal transport is widely applied to particle systems, economic models, and deep learning.

1 Introduction

Optimal Transportation Theory has its applications in numerous fields: predicting particle systems, analyzing stochastic models, and establishing gradient flows in deep learning models [1]. Most famously, the Wasserstein Distance is widely applied to AI models such as the generative model WGAN [2].

The original problem of optimal transport considers how to move masses from different locations to destinations such that the quota is satisfied for each destination and the total cost is minimized. Our work concerns optimal transport between continuous rather than discrete mass distributions such that the transport map is Lipschitz, meaning that the distance between any two points shrinks by at least a certain factor after the mapping.

In the past, Lipschitz properties have been proven for the optimal transport between Euclidean spaces $\mathbb{R}^d$ of equal dimensions with the quadratic cost $c(x, y) = \frac{1}{2}|x - y|^2$, including Carlier, Figalli, and Santambrogio's work [3] as well as Caffarelli's contraction theorem (Theorem 2 in [4]). Caffarelli constructed a contracting optimal transport with the quadratic cost, and the probability measures $\mu = e^{-V}dx$ and $\nu = e^{-W}dx$. We establish parts of this result in more generalized settings of unequal dimensions and non-quadratic costs, following the convention of assuming the domains X and Y to be open and connected.

Firstly, we seek a contraction map between unequal dimensions. We consider optimal transport from $\mathbb{R}^n$ to $\mathbb{R}$ by analyzing the more complicated PDE derived in [5] by McCann and Pass for unequal dimension transports. We show that the quadratic cost retains continuity on T in Section 3.1. Then, we present a localized PDE similar to the Monge-Ampère Equation in Section 3.2, stated in Theorem 3.2.

Secondly, we analyze the case of equal dimensions but varying costs. Transport on spheres and other surfaces operates under different metrics than Euclidean spaces, motivating us to understand transport under various costs. So far, a contraction is known for only the quadratic cost, described in Caffarelli's theorem [4]. We generalize this result to a perturbed version of the quadratic cost $c(x, y) = -xy - \varepsilon|x|^2|y|^2$ in Theorem 4.3, which operates under a framework similar to the original proof of Caffarelli's Theorem.

Finally, in Section 5, we suggest some future directions as well as more distant speculations in the fields of differential geometry, quantum mechanics, and economics.

2 Preliminaries

In Section 2.1, we define measures in preparation for introducing optimal transport. In Section 2.2, we first formulate the two versions of the optimal transport problem: the Monge and Kantorovich problems. Then, we introduce Brenier's theorem [6] as foundations for Section 4.1, where we outline Caffarelli's theorem of contracting transport maps [4]. Finally, we identify the dual of the Kantorovich problem, which we heavily rely on in Section 3 and 4's proofs. In Section 2.3, we introduce Lipschitz functions as preparation for Section 4 (not required for reading Section 3).

2.1 Measures

Without getting into excessive details, a measure μ can be seen as a function that takes a set as an argument equivalent to a real-valued and globally non-negative function f. For a domain X, we have

$$\mu(A) = \int_A f(x)dx$$

for any $A \subseteq X$. Any measure μ possesses *countable additivity*, meaning that

$$\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$$

for any sequence of disjoint sets $\{A_k\}$ with countably infinite length.

Specifically, the d-dimensional *Lebesgue measure* λ_d is used in Section 3. It can be defined on the Euclidean space $\mathbb{R}^d$ as

$$\lambda_d(E) = \inf \left\{ \sum_{k=1}^{\infty} \mathrm{vol}(C_k) \mid E \subset \bigcup_{k=1}^{\infty} C_k \right\},$$

where C_k is a d-dimensional cuboid and vol denotes its volume. The Lebesgue measure essentially measures the amount of space occupied by any set $E \in \mathbb{R}^d$.

Now with an understanding of measures, we may introduce optimal transport.

2.2 Optimal Transport

Gaspard Monge first proposed a version of the optimal transport problem in 1781 [7]. In his formulation, we have $T : X \to Y$ transporting masses in X described by the density measure $\mu(A) = \int_A f(x)dx$, onto Y described by the density measure $\nu(A) = \int_A g(y)dy$. We require T to be a *push-forward map* from μ to ν, meaning that

$$\mu(T^{-1}(A)) = \nu(A)$$

for any $A \subseteq Y$. Figure 1 illustrates how a push-forward T preserves the area of the region when going reversely ν to μ.

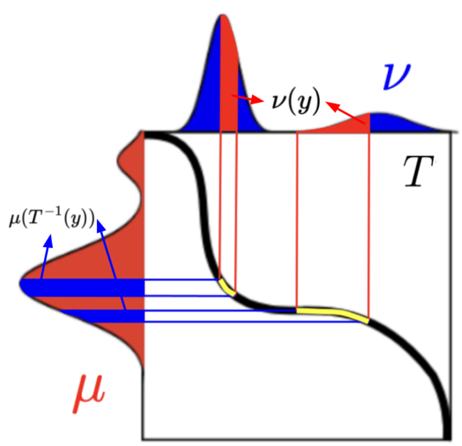

Figure 1: $\mu(T^{-1}(y))$ and $\nu(y)$ represents equal areas for any interval y (Adapted from [8]).

Suppose the transportation cost can be described by $c(x, T(x)) \in \mathbb{R}$. Then a solution to the problem is a map $T(x)$ that minimizes the total cost

$$\int_X c(x, T(x))d\mu(x).$$

When $X, Y \subset \mathbb{R}^d$, the Monge-Ampère Equation, first studied by Monge and later André-

Marie Ampère [9], gives the optimal transport map by the following PDE.

$$\det(\nabla T(x)) = \frac{f(x)}{g(T(x))}.$$

However, in certain compact domains X and Y, this PDE does not yield a continuous optimal mapping. Leonid Kantorovich later proposed a more relaxed version of the Monge Problem [10] to resolve the issue of discontinuity. We call a measure γ a *coupling* if $\pi_X(x, y) = x$ pushes forward γ to μ and $\pi_Y(x, y) = y$ pushes forward γ to ν. Let the set of couplings be $\Gamma(\mu, \nu)$. Such couplings allow masses at any location in X to split to different destinations in Y, so that the total cost can be rewritten as

$$\int_{X\times Y} c(x, y) d\gamma(x), \quad \gamma \in \Gamma(\mu, \nu).$$

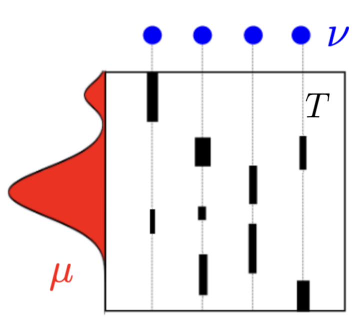

Figure 2: An optimal transport mapping to four discrete point destinations where the mass splits (Adapted from [8])

In Figure 2, several intervals in X have masses transported to more than one of the four destinations in Y. Only the Kantorovich problem encompasses this form of optimal transport. Whereas in Figure 1, the curve outlined by T is a one-to-one mapping with no masses splitting.

The Kantorovich formulation has a dual problem in terms of real-valued potential functions u, v. The minimum of the Kantorovich problem equals the maximum of its dual, as in

$$\inf \left\{ \int_{X\times Y} c(x,y) d\gamma(x) \mid \gamma \in \Gamma(\mu,\nu) \right\} = \sup \left(\int_X u(x) d\mu(x) + \int_Y v(y) d\nu(y) \mid u(x) + v(y) \leq c(x,y) \right).$$

It has been shown that the solution to the dual problem always satisfies c-convexity in the sense that

$$Du(x) + D_x c(x,y) = 0, \quad \text{and} \quad D^2 u(x) + D_{xx} c(x,y) \geq 0.$$

More recently, Yann Brenier [6] showed that the optimal transport map for the quadratic cost function $c(x,y) = -xy$ can always be expressed as $T = \nabla\varphi$ where φ is convex. This optimal T uniquely exists for the Kantorovich formulation while the masses never split. For the quadratic cost, the original PDE becomes

$$\det\left(D^2\varphi\right) g(\nabla\varphi) = f. \tag{1}$$

Brenier's result allows more variations in choosing the density functions f and g, as T is guaranteed to be differentiable. More broadly, Brenier unified the Monge and Kantorovich problems, as he proved that masses will never split in the optimal transport for the quadratic cost. Consequently, the quadratic cost setting is relatively more well-researched: the aforementioned theorems by Caffarelli and Figalli were able to impose regularity conditions stricter than differentiability for T, namely Lipschitz continuity.

2.3 Lipschitz Functions

A function $T : \mathbb{R}^m \to \mathbb{R}^n$ is called *Lipschitz* if it satisfies

$$|T(x_1) - T(x_2)| \leq L|x_1 - x_2|, \quad \text{for any } x_1, x_2 \in \text{dom}(T)$$

given a certain positive number L called the *Lipschitz constant.* The Lipschitz property can also be understood as a strict global upper bound $|\nabla T| \leq L$. It is a regularity condition, such as continuity or smoothness, but stronger than both.

As an example, consider the following picture distinguishing between Lipschitz and non-Lipschitz 1D functions. If the cross $y = \pm 1.2x$ at every point on the function covers the entire curve on the left and right quadrants, it is Lipschitz with Lipschitz constant 1.2. Here, $f(x) = \frac{x}{1+e^{-x}}$ is Lipschitz but $g(x) = \frac{x^2}{1+e^{-x}}$ is not, which can be verified by computing ∇f and ∇g.

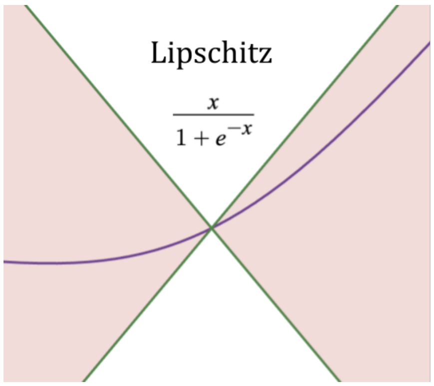

Figure 3: Example of a Lipschitz function with Lipschitz constant 1.2

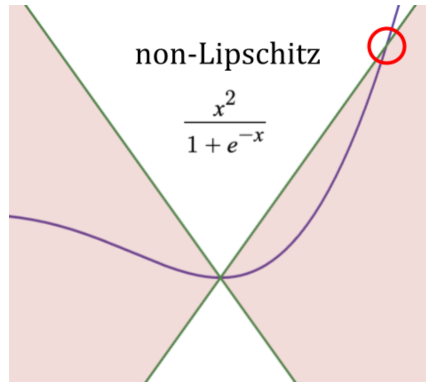

Figure 4: Example of a non-Lipschitz function

3 $\mathbb{R}^n \to \mathbb{R}$ Transport

We outline a derivation of the PDE for optimal transport similar to the generalized case of $\mathbb{R}^m \to \mathbb{R}^n$ by McCann and Pass [5], which was also reproduced by Cassini and Hamfeldt [11]. Given our special case of transporting from n-dimensions to one dimension, we establish necessary conditions for the transport to be continuous. Then, we reduce the PDE to a more simplified form.

3.1 Continuity of the transport from $\mathbb{R}^n$ to $\mathbb{R}$

We impose continuity on T using the condition of *nestedness* introduced by Chiappori, McCann, and Pass in [12], as nestedness has demonstrated capabilities of gaining a variety of regularity conditions when combined with other restrictions to (c, μ, ν). First, denote level sets of c as

$$X_=(y,k) := \{x \in X \mid c_y(x,y) = k\},$$

and similarly denote sub-level sets as

$$X_\leq(y,k) := \{x \in X \mid c_y(x,y) \leq k\}.$$

The same notation applies to $\geq, <, >$. Define $k(y) : Y \to \mathbb{R}$ as the unique function that satisfies

$$\mu(X_\leq(y,k(y))) = \nu((-\infty, y]), \quad \text{for any } y \in Y.$$

In other words, $k(y)$ forms sub-level sets that divide μ at the same ratio as y divides ν.

We may now establish the notion of nestedness, which has appeared in numerous studies in the field, including [12, 13, 14].

Definition 3.1. The triple (c, μ, ν) is nested if for any $y_0 < y_1$,

$$X_\leq(y_0, k(y_0)) \subseteq X_<(y_1, k(y_1)).$$

Using nestedness, we prove the following proposition.

Lemma 3.1. *If μ, ν are absolutely continuous with respect to Lebesgue measure, the optimal transport for the cost function*

$$c(x,y) = \sum_{i=1}^{n} \frac{1}{2}(x_i - y)^2$$

is always continuous on the interior of X.

Proof. We show that the quadratic cost is in *pseudo-index* form. That is, $c(x,y) = \alpha(x) + \sigma(I(x), y)$, where $\alpha : X \to \mathbb{R}$, $\sigma : \mathbb{R} \times Y \to \mathbb{R}$, and $I : \mathbb{R}^n \to \mathbb{R}$ are C^1 functions. Proposition 4.3 in [12] showed that for any absolutely continuous measures μ and ν, the triplet (c, μ, ν) is nested as long as c is in pseudo-index form and $D^2_{xy}c$ never vanishes. Taking

$$\alpha = \frac{1}{2}\sum_{i=1}^{n} x_i^2, \ I(x) = \sum_{i=1}^{n} x_i, \ \sigma(r,y) = ry + \frac{n}{2}y^2,$$

we observe that quadratic cost is in pseudo-index form.

Thus, by Theorem 4.2 in [12], nestedness implies continuity of T on the interior of X. □

3.2 Analog of the Monge-Ampère Equation

Theorem 3.2. *For sets $X \subset \mathbb{R}^n, Y \subset \mathbb{R}$, denote $\lambda_d(x)$ as the d-dimensional Lebesgue measure on X and*

$$X_1(y, v'(y)) = \{x \in X \mid v'(y) + ny = \sum_{i=1}^{n} x_i\}, \quad \textit{for all } y \in Y.$$

Suppose f, g are absolutely continuous. If a c-convex potential function v satisfies

$$g(y) = \frac{2^{n-1}(\frac{n-1}{2})!}{\pi^{\frac{n-1}{2}}} \int_{X_1(y,v'(y))} \frac{(v''(y)+n)}{\sqrt{n}} f(x) d\lambda_{n-1}(x),$$

then the optimal transport dual's solution $T : X \to Y$ is uniquely optimal.

Proof. By the push-forward property of T,

$$\begin{aligned} \int_Y \Phi(y)g(y)dy &= \int_Y \Phi(y)f(T^{-1}(y))dy \\ &= \int_X \Phi(T(x))f(x)dx \ \text{ for all } \Phi \in L^1(Y). \end{aligned} \tag{2}$$

Since $Y \subset \mathbb{R}$, note that the Jacobian $JT = \sqrt{\det(\nabla T(x) \nabla T(x)^T)} = |\nabla T|$. So, we can use the coarea formula to transform the right-hand side to

$$\begin{aligned} \int_X \Phi(T(x))f(x)dx &= \int_X \left(\frac{\Phi(T(x))f(x)}{|\nabla T|} \cdot |\nabla T| \right) dx \\ &= \int_{\mathbb{R}} \int_{T^{-1}(Y)} \frac{\Phi(T(x))f(x)}{|\nabla T|} dH_{n-1}(x)dy \\ &= \int_{\mathbb{R}} \Phi(y) \int_{T^{-1}(Y)} \frac{f(x)}{|\nabla T|} dH_{n-1}(x)dy, \end{aligned}$$

where H_{n-1} is the $(n-1)$-dimensional Hausdorff measure. Since the equation holds for any test function Φ, we may eliminate it. We have simplified Equation (2) to

$$g(y) = \int_{T^{-1}(Y)} \frac{f}{|\nabla T|} dH_{n-1}(x) \qquad \text{(a.e.)}. \tag{3}$$

Now, consider the dual problem of optimal transport, which is to find a pair of c-convex functions $(u, v) \in C^0(X) \times C^0(Y)$ such that $u(x) + v(y) + c(x, y) \geq 0$ with equality at $T(x) = y$. In our setting, such a pair would satisfy

$$D_y(u(x) + v(y) + c(x, y)) = 0,$$

where $y = T(x)$. Differentiating this equation, we get

$$D_{yy}v(T(x))\nabla T(x) + D_{yy}c(x,T(x))\nabla T(x) + D_{xy}c(x,T(x)) = 0.$$

Note that in the equation, terms of ∇T have scalar coefficients, so we conclude that

$$\nabla T(x) = -\frac{T(x) + \nabla(D_y c(x,T(x)))}{D_{yy}v(T(x)) + D_{yy}c(x,T(x))}.$$

On the other hand, the Hausdorff measure is a rescaling of the Lebesgue measure λ_d. So, switching $T^{-1}(Y)$ for X and H_d for λ_d, we can write Equation (3) as

$$\begin{aligned} g(y) &= \int_{T^{-1}(Y)} \frac{D_{yy}v(T(x)) + D_{yy}c(x,T(x))}{|\nabla(D_y c(x,T(x))|} f(x)dH_{n-1}(x) \\ &= \frac{2^{n-1}(\frac{n-1}{2})!}{\pi^{\frac{n-1}{2}}} \int_{T^{-1}(Y)} (v''(y) + D_{yy}c(x,y)) \frac{f(x)}{|\nabla D_y c(x,y)|} d\lambda_{n-1}(x). \end{aligned} \tag{4}$$

Note that we still require c-convexity, namely $D^2_{yy}(v(y) + c(x,y)) \geq 0$, $y \in T^{-1}(Y)$. While incorporating this condition into the differential equation, we attempt to localize the domain of integration by restricting certain regularities.

Define the following sets:

$$\begin{cases} \partial_c v(y) := \left\{ x \in X \mid v(y) + c(x,y) = \inf_{z \in Y}\{v(z) + c(x,z)\} \right\} \\ X_1(y, v'(y)) := \{x \in X \mid v'(y) + c_y(x,y) = 0\} \\ X_2(y, v'(y), v''(y)) := \{x \in X_1 \mid v''(y) + c_{yy}(x,y) \geq 0\} \end{cases}.$$

The first set is referred as the c-subgradient of v.

McCann and Pass have shown in Theorem 4.2(b) of [12] that for a nested triplet (c, μ, ν), these sets are all equal to $T^{-1}(Y)$. Furthermore, when replacing $T^{-1}(Y)$ with $X_1(y, \nabla v(y))$ in the integration domain in Equation (4), the PDE always produces the solution to the optimal transport problem by solving Equation (4) for a unique v and then solving $v'(T(x)) + c_y(x, T(x)) = 0$ for T. Using Lemma 3.1, the quadratic cost gains nestedness for any absolutely continuous μ and ν. Hence, for the quadratic cost, Equation (4) becomes

$$g(y) = \frac{2^{n-1}(\frac{n-1}{2})!}{\pi^{\frac{n-1}{2}}} \int_{X_1(y,v'(y))} \frac{(v''(y) + n)}{\sqrt{n}} f(x)d\lambda_{n-1}(x),$$

while

$$X_1(y, \nabla v(y)) = \{x \in X \mid v'(y) + ny = \sum_{i=1}^{n} x_i\}.$$

□

Hereby, we produce a version of the Monge-Ampère Equation that analyzes transport between unequal dimensions. In summary, starting from the push-forward equation, we get a PDE considering equal dimensions through the coarea formula. Then, we utilize properties of the Kantorovich dual problem to ensure that T is optimal. Finally, we apply the previous lemma to adjust the PDE to be local.

3.3 Cases where higher regularity fails

Now, with our version of the Monge-Ampère Equation, we come back to the idea of seeking Lipschitz transport maps. Intuitively, a contraction ought to be easily attainable when transporting from a higher dimension to a lower one, taking the projection map as an example. We briefly present a possible counter example to this idea.

Consider the optimal transport from the sphere $\mathbb{S}^n$ to its diameter $D = [-1, 1]$. By considering the patterns of the first Neumann eigenvalue (see Appendix A for more details), this transport from $\mathbb{S}^n$ to $\mathbb{R}$ seems unlikely to contract. This is because

$$\lambda_1(\mathbb{S}^n) = n \geq \lambda_1(D) = \frac{\pi^2}{4},$$

which implies that any smooth transformation should make the surface expand. Therefore, more analysis is needed for the case of transport between unequal dimensions.

4 Perturbation of the quadratic cost

In this section, we first sketch a proof of Cafferelli's theorem of contraction. Following a similar structure, we compute the necessary components for the perturbed quadratic cost before proving Theorem 4.3 by applying the Ma-Trudinger-Wang tensor from their work in [15].

4.1 Cafferelli's theorem

Theorem 4.1. *Let $T = \nabla\varphi$ be the optimal transportation mapping pushing forward a probability measure $\mu(x) = e^{-V}dx$ onto a probability measure $\nu(y) = e^{-W}dy$ for $V, W \in C^2$ with respect to cost $c(x, y) = -xy$. If*

$$\sup_{\substack{x\in\mathbb{R}^n \\ e\in\mathbb{S}^{n-1}}} V_{ee} \leq M \quad \textit{and} \quad |D^2 W| \geq K,$$

then T is Lipschitz. In particular, if μ is the standard Gaussian measure and $K \geq 1$, then T is a contraction.

In the spirit of reducing the complexity of Theorem 4.3, we sketch a proof of Cafferellli's contraction theorem in Appendix B with notations similar to the ones used in our proof of Theorem 4.3.

4.2 Setting up the perturbed quadratic cost

To start, consider the original Monge-Ampère Equation

$$\det(\nabla T(x)) = \frac{f(x)}{g(T(x))}.$$

The potential function u of the dual problem satisfies $-D_x c(x, T(x)) = \nabla u(x)$. Differentiating this once gives us $-D_{xx}c(x, T(x)) - D_{xy}c(x, T(x))DT(x) = D^2u(x)$. Therefore, substituting this into the Monge-Ampère Equation,

$$\det\left(D^2 u + D_{xx}c\right) = |\det\left(D_{xy}c\right)| \frac{f(x)}{g(T(x))}. \tag{5}$$

Note that $D^2u + D_{xx}c$ is positive definite as a consequence of c-convexity. We work from this equation for the rest of this section. Before further computation, we introduce the notations in [15] of writing $c_{ij} = (D_{xx}c)_{ij}$ and $c_{i,j} = (D_{xy}c)_{ij}$, as well as

$$c_{ijk} = \frac{\partial}{\partial x_k} c_{ij}, \quad c_{ij,k} = \frac{\partial}{\partial y_k} c_{ij},$$
$$c_{ijkl} = \frac{\partial^2}{\partial x_k x_l} c_{ij}, \quad c_{ij,kl} = \frac{\partial}{\partial y_k y_l} c_{ij}, \quad c_{ijk,l} = \frac{\partial}{\partial x_k y_l} c_{ij}.$$

The same applies to matrices w and D^2u. Also, write

$$T_{i,j} = \frac{\partial}{\partial x_j} T_i, \qquad (c_{i,j})^{-1} = c^{i,j}, \qquad (w_{ij})^{-1} = w^{ij}.$$

With the simplified notation, we compute the derivatives of the perturbed quadratic cost, which yield

$$c_{ij} = -\delta_{ij} 2\varepsilon |T|^2, \quad c_{i,j} = -\delta_{ij} - 4\varepsilon x_i y_j, \quad c_{ii,k} = -4\varepsilon y_k, \quad c_{i,kk} = -4\varepsilon x_i, \quad c_{ii,jj} = -4\varepsilon. \tag{6}$$

From there, we may also compute

$$\det(c_{i,j}) = (-1)^n (1 + 4\varepsilon xy), \qquad c^{i,j} = -\delta_{ij} + \frac{4\varepsilon x_i y_j}{1 + 4\varepsilon xy}. \tag{7}$$

Any other form of derivatives yield 0.

Now, we may introduce the Ma-Trudinger-Wang Tensor:

$$\mathfrak{S}_{(x,y)}(\xi, \eta) := \sum_{i,j,k,l,p,q,r,s} \left(c_{ij,p} c^{p,q} c_{q,rs} - c_{ij,rs} \right) c^{r,k} c^{s,l} \xi^i \xi^j \eta^k \eta^l, \qquad \xi, \eta \in \mathbb{R}^n. \tag{8}$$

The following lemma is true for our perturbed quadratic cost.

Lemma 4.2. *The cost $c = -xy - \varepsilon |x|^2 |y|^2$ satisfies*

$$\mathfrak{S}_{(x,y)}(\xi, \eta) \geq \frac{16\varepsilon}{(1 + 4\varepsilon R^2)^3} |\xi|^2 |\eta|^2$$

for any $\xi, \eta \in \mathbb{R}^n$.

Proof. The tensor can now be expressed as

$$\begin{aligned}
\mathfrak{S}_{(x,y)}(\xi, \eta) = & \sum_{i,j,k,l,p,q,s,t} \left(\frac{16\varepsilon^2 x_q y_p}{1 + 4\varepsilon xy} \cdot \left(-\delta_{pq} + \frac{4\varepsilon x_p y_q}{1 + 4\varepsilon xy} \right) + 4\varepsilon \right) \\
& \qquad \cdot \left(-\delta_{sk} + \frac{4\varepsilon x_s y_k}{1 + 4\varepsilon xy} \right) \cdot \left(-\delta_{tl} + \frac{4\varepsilon x_t y_l}{1 + 4\varepsilon xy} \right) \cdot \xi^i \xi^j \eta^k \eta^l \\
& = \sum_{i,j,k,l,s,t} \frac{4\varepsilon}{1 + 4\varepsilon xy} \cdot \left(-\delta_{sk} + \frac{4\varepsilon x_s y_k}{1 + 4\varepsilon xy} \right) \cdot \left(-\delta_{tl} + \frac{4\varepsilon x_t y_l}{1 + 4\varepsilon xy} \right) \xi^i \xi^j \eta^k \eta^l \\
& \geq \frac{4\varepsilon}{1 + 4\varepsilon R^2} \sum_{i,j,k,l,s,t} \left(-\delta_{sk} + \frac{4\varepsilon x_s y_k}{1 + 4\varepsilon xy} \right) \cdot \left(-\delta_{tl} + \frac{4\varepsilon x_t y_l}{1 + 4\varepsilon xy} \right) \xi^i \xi^j \eta^k \eta^l \\
& \geq \frac{16\varepsilon}{(1 + 4\varepsilon R^2)^3} |\xi|^2 |\eta|^2.
\end{aligned}$$

Note that the last inequality holds because if we write ξ or η as $(r\cos t, r\sin t)$, we obtain

$$\sum_{i,j} \xi^i \xi^j = r^2(\cos t + \sin t) \leq 2r^2 = 2|\xi|^2.$$

□

With this lemma, we may present our proof to Theorem 4.3.

4.3 Extension of Caffarelli's theorem

Theorem 4.3. *Let $T : X \to Y$ be the optimal transportation mapping pushing forward a probability measure $\mu(x) = e^{-V}dx$ onto a probability measure $\nu(y) = e^{-W}dy$ for $V, W \in C^2$ with respect to cost $c(x,y) = -x \cdot y - \varepsilon|x|^2|y|^2$. Suppose $X, Y \subseteq B_R \subset \mathbb{R}^2$ where B_R is the open ball with radius R and $\varepsilon R^2 \to 0$ but R is sufficiently large. If*

$$\sup_{\substack{x\in\mathbb{R}^2 \\ e\in\mathbb{S}^1}} D^2_{ee}V \leq M, \quad |\nabla W| < R, \quad \text{and} \quad |D^2 W| \geq K > 0,$$

then T is Lipschitz with Lipschitz constant bounded by

$$\mathrm{Lip}(T) < \frac{M}{K}.$$

Proof. Now, we consider only the optimal transport in between $\mathbb{R}^2$ and let w be the 2×2 matrix satisfying $w_{ij} = u_{ij} + c_{ij}$. This w is analogous to $D^2\varphi$ in our proof of Theorem 4.1, but without Brenier's results on the quadratic cost, computations become more microscopic and we need to consider individual entries of w.

Similar to thep roof of Theorem 4.1, we take the log of Equation (5) when $f(x) = e^{-V}$, and $g(y) = e^{-W}$. We obtain

$$\begin{aligned} \varphi := \log\left(w_{11}w_{22} - w_{12}w_{21}\right) &= \log|1 + 4\varepsilon x_1 y_1 + 4\varepsilon x_2 y_2| + W(T(x)) - V(x) \\ &= \log|1 + 4\varepsilon x \cdot T(x)| + W(T(x)) - V(x). \end{aligned}$$

Differentiating once along the direction of x_1, we obtain

$$\varphi_1 = \frac{w_{11,1}w_{22} + w_{11}w_{22,1} - w_{12,1}w_{21} - w_{12}w_{21,1}}{\det(w)} = \frac{4\varepsilon x_1 T_{1,1} + 4\varepsilon T_1}{1 + 4\varepsilon xT} + \nabla W(T) \cdot \frac{\partial T}{\partial x_1} - V_1. \tag{9}$$

Differentiating again, we obtain

$$\varphi_{11} = \frac{w_{11,11}w_{22} + w_{11}w_{22,11} - w_{12,11}w_{21} - w_{12}w_{21,11}}{\det(w)} - \varphi_1^2 + \frac{2w_{11,1}w_{22,1} - 2w_{12,1}w_{21,1}}{\det(w)}$$
$$= -\frac{(4\varepsilon x_1 T_{1,1} + 4\varepsilon T_1)^2}{(1 + 4\varepsilon xT)^2} + \frac{8\varepsilon T_{1,1} + 4\varepsilon x_1 T_{1,11}}{1 + 4\varepsilon xT} + D^2W(T)\left(\frac{\partial T}{\partial x_1}\right)^2 + \nabla W(T) \cdot \frac{\partial^2 T}{\partial x_1^2} - V_{11}. \tag{10}$$

Assume that the quantity

$$G(w, t) = w \cdot [\cos t, \sin t] \cdot [\cos t, \sin t]^T$$

attains its maximum at $x_0 \in X$ and $t_0 \in [0, 2\pi)$. By a shift along the t-axis, we may further assume that $t_0 = 0$. Thus,

$$\frac{\partial}{\partial t} G(w, t) = -2w_{11} \cos t \sin t + (w_{12} + w_{21})(\cos^2 t - \sin^2 t) + 2w_{22} \cos t \sin t = 0$$

implies that $w_{12} + w_{21} = 0$, as $G_t(w, 0) = 0$. Assuming continuity on w, we can exchange the indices of w. So, $w_{12} = w_{21} = 0$ and w is diagonal. This exchangeability also establishes $w_{12,1}w_{21,1} = w_{12,1}^2 \geq 0$. Notice that if $t_0 = 0$, we also have $w_{11,i} = 0$ and $w_{11,ij} \leq 0$ for $i, j = 1, 2$ at the maximum point x_0.

With these assumptions at the maximal point x_0, Equation (10) gives the inequality

$$\varphi_{11} \leq \frac{w_{11,11}}{w_{11}} + \frac{w_{22,11}}{w_{22}} \leq \frac{w_{22,11}}{w_{22}}. \tag{11}$$

Now, our goal is to convert this inequality to a bound on w_{11} in terms of the derivatives of c and φ. First, we compute the fourth derivatives of w.

The relation $-D_{xx}c(x, T(x)) - D_{xy}c(x, T(x))\nabla T(x) = D^2 u(x)$ implies that for any i, j,

$$w_{ij} = c_{i,1}T_{1,j} + c_{i,2}T_{2,j}. \tag{12}$$

Differentiating again, we have

$$w_{ij,k} = \sum_{p,q=1}^{2} c_{ik,p}T_{p,j} + c_{i,pq}T_{p,j}T_{q,k} + c_{i,p}T_{p,jk}. \tag{13}$$

Taking one more derivative, we have

$$\begin{aligned} w_{22,11} = u_{2211} + c_{2211} + 2c_{221,1}T_{1,1} + 2c_{221,2}T_{2,1} + c_{22,1}T_{1,11} \\ + c_{22,2}T_{2,11} + c_{22,11}T_{1,1}^2 + 2c_{22,12}T_{1,1}T_{2,1} + c_{22,22}T_{2,1}^2. \end{aligned} \tag{14}$$

We try to express all derivatives of T in terms of derivatives of w and c before returning to Equation (11). Repeating the identity (12) for different indices and inverting the equations, we have

$$T_{p,i} = c^{p,1}w_{1i} + c^{p,2}w_{2i} \tag{15}$$

for any indices p, i. We run through a similar process for Equation (13) and see that

$$T_{p,ij} = \sum_{k=1}^{2} c^{p,k}w_{ij,k} + \sum_{k,s,t=1}^{2} c^{p,k}(2c_{ki,s}T_{s,j} + c_{ij,s}T_{s,k} + c_{k,st}T_{s,i}T_{t,j}). \tag{16}$$

Combining this fact with $w_{12} = w_{21} = 0$, $\det(D_{xy}c) = 1 - \Omega(\varepsilon)$ and our computations (6) and (7), we may infer that

$$c_{i,j}, c^{i,j} \sim \varepsilon R^2 + \delta_{ij}, \quad c_{ii,k}, c_{i,kk} \sim \varepsilon R, \quad c_{ii_jj} \sim \varepsilon. \tag{17}$$

From Equation (15), we also have

$$T_{p,i} \sim w_{11}(\varepsilon R^2 + \delta_{ij}). \tag{18}$$

Now, starting from inequality (11), we recover the Ma-Trudinger-Wang tensor at the point x_0. By substituting in Equation (14) and noting that $c_{ijk,l} = 0$, we obtain

$$0 \geq -(c_{22,1}T_{1,11} + c_{22,2}T_{2,11} + c_{22,11}T_{1,1}^2 + 2c_{22,12}T_{1,1}T_{2,1} + c_{22,22}T_{2,1}^2)w^{22} + \varphi_{11}.$$

Then, Equation (15) plus the estimates (17) and (18) reduces Equation (16) to

$$T_{k,11} = \sum_{l,s,t} c^{k,l}c_{l,st}T_{s,1}T_{t,1} + \Omega(\varepsilon^2 R^3).$$

With these computations, the Ma-Trudinger-Wang tensor emerges by taking $\xi = (0, \sqrt{w_{22}})$

and $\eta = (w_{11}, 0)$. After applying Equation (8), our original inequality becomes

$$0 \geq \sum_{k,l,s,t} w^{22}(c^{k,l}c_{22,k}c_{l,st} - c_{22,st})c^{s,1}c^{t,1}w_{11}^2 + \varphi_{11} + \Omega(\varepsilon^3 R^4)$$

$$= \mathfrak{S}_{(x,y)}(\xi, \eta) + \varphi_{11} + \Omega(\varepsilon^3 R^4).$$

Finally, we apply Lemma 4.2 and conclude that

$$\frac{16\varepsilon}{(1+4\varepsilon R^2)^3} w_{11}^2 w^{22} + \varphi_{11} + \Omega(\varepsilon^3 R^4) \leq 0.$$

Since $w^{22} \geq w^{11}$, we may simplify our error term into $\Omega(\varepsilon w_{11} + \varepsilon^3 R^4)$. Again, we apply Equation (10) and estimate derivatives of T by Equations (16) and (18) before absorbing the ε terms of higher degrees, resulting in

$$C(\varepsilon K R^2 w_{11} + \varepsilon^2 R^4 w_{22}) + w_{11}^2 \leq \frac{V_{11}}{K}$$

for an arbitrary constant C. Recall that $w_{ii} = u_{ii} - 2\varepsilon|T|^2$ and $u_{11} = -c_{11} + c_{1,1}T_{1,1} + c_{1,2}T_{2,1}$, so we replace w_{11} with $T_{1,1}$, absorb ϵ terms, and yield

$$(\varepsilon C R^2 + 1)|T_{1,1}| \leq \sqrt{\frac{V_{11}}{K}}$$

Since $|D_{ee}^2 u| \leq |u_{11}|$ for any direction e, we may bound T's Lipschitz constant by

$$\mathrm{Lip}(T) < \sup_{\substack{x \in \mathbb{R}^2 \\ e \in \mathbb{S}^1}} \frac{D_{ee}^2 V(x)}{K}$$

for sufficiently small ε. □

Thus, we conclude with a Lipschitz bound that resembles Cafferelli's theorem. In summary, similar to Caffarelli's proof, we start by taking the log of the Monge-Ampère Equation and differentiate it twice. Then, we apply the maximum principle and operate at a specific maximal point x_0, providing conditions on the derivatives of the potential function u. However, without the result of Brenier ($T = \nabla\varphi$), we work with the relation $Du(x) + D_x c(x, T(x)) = 0$, which complicates certain computations. Eventually, we apply the Ma-Trudinger-Wang tensor, a tensor of c that is upper-bounded for $c(x,y) = -xy - \varepsilon|x|^2|y|^2$, bounding the first derivative of T.

5 Future Directions

Nenna and Pass pointed out that the optimal transport between unequal dimensions can potentially map between strategy spaces in Cournot-Nash equilibria [14].

On the other hand, our study also has its profound impact on differential geometry. We potentially contribute to current works looking at the Shing-Tung Yau conjecture [16] through the lenses of optimal transport.

Conjecture 5.1. *The first eigenvalue of any compact embedded minimal surface* Σ *in* $\mathbb{S}^{n+1}$ *is* n.

Ultimately, the goal is to construct a transport map $T : \mathbb{S}^{n+1} \to \Sigma$ that is Lipschitz. Furthermore, if T satisfies the Lipschitz property for $L = 1$, the map is a contraction, and Theorem 1.7 in Milman's work [17] gives

$$L\lambda_1(\Sigma) \geq \lambda_1(\mathbb{S}^{n+1}) = n.$$

This proves the conjecture. Both of our directions help gain insights to this ultimate goal.

First, since $\mathbb{S}^{n+1}$ is a dimension higher than its embedded minimal surface, a transport map between unequal dimensions is potentially needed. Despite establishing a local PDE, the conventional techniques used in Caffarelli's theorem, Figalli's theorem [3], and Theorem 4.3 do not naturally generalize to the optimal transport of unequal dimensions. In particular, our PDE is non-linear, making the tactic of taking the log and differentiating twice insufficient. Future work on linearizing the equation in Theorem 3.2 may be crucial.

From a different perspective, the Monge-Ampère Equation on non-Euclidean manifolds includes a term to describe the metric, which we may simulate by using specific cost functions. In the future, we hope to first generalize our perturbed quadratic cost result to maps of $\mathbb{R}^n \to \mathbb{R}^n$. Then, we may attempt to generalize our perturbation to cost functions in the form of

$$c(x, y) = -xy - \varepsilon f(x)g(y).$$

From there, we can potentially motivate contraction maps for a greater variety of costs, especially costs that simplify the problem of optimal transport in non-Euclidean spaces. For example, the reflector antenna cost $c(x, y) = -2 \log \|x - y\|$ simplifies optimal transport on a sphere [18].

Acknowledgements

I would like to express my deepest gratitude to my mentor, Shrey Aryan, for proposing this project and guiding me through the past month with great insights. I would also like to express my appreciation to Dr. Tanya Khovanova, Prof. David Jerison and Dr. Jenny Sendova for their weekly feedback on my progress and their inspirational views on conducting mathematical research. I also thank last week's teaching assistants Mihika Dusad and Viktor Kolev as well as Hanming Ye and Assylbek Olzhabayev for their invaluable advice on my drafts. My gratitudes also extend to the first teaching assistants Srikara Vishnubhatla, Lauren Shen, and Ella Lan for their instructions on using Latex and Beamer. Finally, I thank the Center for Excellence in Education (CEE), the Massachusetts Institute of Technology (MIT), the Research Science Institute (RSI), and my sponsor, Princetonow Education, for making this once-in-a-lifetime research opportunity.

References

[1] J. Carrillo. *Optimal Transport and Partial Differential Equations*. University of Oxford.

[2] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN, 2017. URL `https://arxiv.org/abs/1701.07875`. `1701.07875`.

[3] G. Carlier, A. Figalli, and F. Santambrogio. On optimal transport maps between 1/d-concave densities, 2024. URL `https://arxiv.org/abs/2404.05456`. `2404.05456`.

[4] L. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214(3):pp. 547–563, 2000.

[5] R. J. McCann and B. Pass. Optimal transportation between unequal dimensions. *Archive for Rational Mechanics and Analysis*, (238):p. 1475–1520, 2020. URL `https://doi.org/10.1007/s00205-020-01569-5`.

[6] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, (44):pp. 375–417, 1991. URL `https://doi.org/10.1002/cpa.3160440402`.

[7] G. Monge. *Mémoire sur le calcul intégral des équations aux différences partielles*. Mémoires de l'Académie des Sciences. Paris, France: Imprimerie Royale., 1784.

[8] P. Dvurechensky. *An Introduction to Optimal Transport*. Weierstrass Institute for Applied Analysis and Stochastics.

[9] A.-M. Ampère. *Mémoire contenant l'application de la théorie exposée dans le XVII. e Cahier du Journal de l'École polytechnique, à l'intégration des équations aux différentielles partielles du premier et du second ordre*. Paris: De l'Imprimerie royale., 1819.

[10] L. Kantorovich. On the translocation of masses. *Management Science*, 5(1), 1958.

[11] M. A. Cassini and B. F. Hamfeldt. Numerical optimal transport from 1d to 2d using a non-local monge-ampère equation, 2023. URL `https://arxiv.org/abs/2307.06820`. `2307.06820`.

[12] P. Chiappori, R. J. McCann, and B. Pass. Multi-to one-dimensional optimal transport. *Communications on Pure and Applied Mathematics*, 70(12):pp. 2405–2444, 2017. doi: 10.1002/cpa.21707. URL `http://dx.doi.org/10.1002/cpa.21707`.

[13] P. Chiappori, R. J. McCann, and B. Pass. Transition to nestedness in multi- to one-dimensional optimal transport. *European Journal of Applied Mathematics*, 30(6):pp. 1220—1228, 2019. doi:10.1017/S0956792518000578.

[14] L. Nenna and B. Pass. Variational problems involving unequal dimensional optimal transport. *Journal de Mathématiques Pures et Appliquées*, 139:pp. 83–108, 2020. doi: 10.1016/j.matpur.2020.05.004.

[15] X. N. Ma, N. Trudinger, and X. J. Wang. Regularity of potential functions of the optimal transportation problem. *Archive for Rational Mechanics and Analysis*, 177:pp. 151–183, 2005. doi:10.1007/s00205-005-0362-9.

[16] S. T. Yau. *Seminar on Differential Geometry*. Princeton Univ. Press, 1982.

[17] E. Milman. Spectral estimates, contractions and hypercontractivity. *Journal of Spectral Theory*, (8):pp. 669–714, 2018. URL `https://arxiv.org/abs/1508.00606`.

[18] B. F. Hamfeldt and A. G. R. Turnquist. Convergent numerical method for the reflector antenna problem via optimal transport on the sphere. *J. Opt. Soc. Am. A*, 38(11):pp. 1704–1713, 2021. doi:10.1364/JOSAA.439679.

Appendix

A The First Neumann Eigenvalue

First define the Laplacian operator Δ on a function $u : \mathbb{R}^d \to \mathbb{R}$ as

$$\Delta u = \sum_{i=1}^{d} \frac{\partial^2 u}{\partial x_i^2}.$$

For a bounded region $\Omega \in \mathbb{R}^d$ with boundary $\partial\Omega$, denote $N(x)$ as the unit normal vector at point $x \in \Omega$. Then, *Neumann eigenvalues* are non-negative numbers λ that produce C^2 solutions to the set of differential equations

$$\begin{aligned} -\Delta u(x) = \lambda u(x) \qquad & x \in \Omega, \\ \nabla u(x) \cdot N(x) = 0 \qquad & x \in \partial\Omega. \end{aligned}$$

Hence, for any surface S embedded in a Euclidean space, we have a sequence of eigenvalues $\{\lambda_k\}$ such that

$$0 = \lambda_0(S) \leq \lambda_1(S) \leq \dots$$

We call $\lambda_1(S)$ the *first Neumann eigenvalue* of S.

B Complete Proof of Theorem 4.1

Proof. Since we are free to rotate the coordinates, assume that φ_{ee}, the second directional derivative with respect to unit vector e, takes its maximum when e is in the direction of x_1. Thus, $\nabla\varphi_{11} = 0$ and $D^2\varphi_{11} \leq 0$.

Now, we start with the Monge-Ampère Equation after applying Brenier's Theorem, which is Equation (1).

$$e^{-V} = e^{-W(\nabla\varphi)} \det\big(D^2\varphi\big)$$

From now on, denote the matrix $D^2\varphi$ as w and $D^2\left(\frac{\partial\varphi}{\partial x_i}\right)$ as w_i. Taking the log of both sides, we get

$$V = W(T) - \log(\det(w)).$$

Differentiating along the direction of x_1, we have

$$V_1 = \nabla W(T) \cdot T_1 - \frac{\partial_1 \det(w)}{\det(w)}$$
$$= \nabla W(T) \cdot T_1 - \operatorname{tr}\left(w^{-1} w_1\right).$$

Differentiating along x_1 again and noting that $T_{11} = 0$, we get

$$V_{11} = D^2 W(T) \cdot T_1^2 - \operatorname{tr}\left(w^{-1} w_{11}\right) - \operatorname{tr}\left(w^{-1} w_1 w w_1^{-1}\right)$$

Note that $(w \cdot w^{-1})_1 = w_1 w^{-1} + w w_1^{-1} = 0$, so $\operatorname{tr}(w^{-1} w_1 w w_1^{-1}) = -\operatorname{tr}((w^{-1} w_1)^2) \leq 0$. Since φ is convex, w is positive definite and $\operatorname{tr}(w^{-1} w_{11}) \leq 0$. Thus,

$$V_{11} \geq D^2 W(T) \cdot T_1^2 \geq K T_1^2.$$

The same argument applies to bounding T_i for any principal direction x_i. Therefore, $|\nabla T|$ is upper bounded and T is Lipschitz. Furthermore, when V is the standard Gaussian $e^{-\frac{|x|^2}{2}}$, $V_{11} = 1$. Thus, $K \geq 1$ guarantees that $\operatorname{Lip}(T) \leq 1$. □