WEIZMANN INSTITUTE OF SCIENCE
מכון ויצמן למדע

Thesis for the degree
Master of Science

עבודת גמר (תזה) לתואר
מוסמך למדעים

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By
Daniel Tsodyks

מאת
דניאל צודיקס

יציבות של טרנספורט אופטימלי במרחב אוקלידי
Stability of Optimal Transport in Euclidean Space

Advisor: Bo’az Klartag

מנחה: בועז קלרטג

June 30th, 2025

ד' בתמוז, התשפ"ה

1 Introduction

The problem of optimal transport has been studied since Monge in 1781. In its modern formulation, we are given two probability measures, say on $\mathbb{R}^n$, and are tasked with finding a map T, which transfers μ to ν, and is optimal in the sense that it minimizes an average cost $\mathbb{E}_{x\sim\mu}[c(x,T(x))]$, for some predetermined cost function $c:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$. The most studied choice of such a cost function is $c(x,y)=|x-y|^p/p$, for some $p>0$, though we will consider only the case $p\geq 1$.

The most natural question one can ask of this problem is the existence and uniqueness of solutions to it. Under the mild assumption that the source measure μ is absolutely continuous, it turns out that there is always such an optimal transport map, and if $p>1$ then it is unique. If $p=1$, such a map need not be unique even for simple cases. For example, if μ,ν are the uniform probability measures on $[0,1],[1,2]$ respectively, then both $T(x)=x+1$ and $T(x)=2-x$ are optimal. However, there is some uniqueness property: For any measures μ,ν, we can associate a Lipschitz function with Lipschitz constant 1, called the Kantorovich potential of the pair, and it is unique up to an addition of a constant.

In general, the potentials are a pair of real-valued functions maximizing a dual formulation of the optimal transport problem. Specifically, they are a pair (φ,ψ), such that $\varphi(x)+\psi(y)\leq c(x,y)$ for all $x,y\in\mathbb{R}^n$, and which maximize $\mathbb{E}_{x\sim\mu}[\varphi(x)]+\mathbb{E}_{x\sim\nu}[\psi(x)]$ among such functions. For example, for $p=1$, the potentials turn out to be $\pm u$, where u is the aforementioned Lipschitz function; and for $p=2$, they are convex functions which are Legendre transforms of each other. Another natural question is robustness of these potentials, or of the optimal transport maps, depending on variations in the measures μ,ν. It is indeed true that the potentials and the transport maps depend continuously on the measures μ,ν. A more complex question is about the modulus of continuity of such maps - e.g., are they α-Hölder continuous, for some $\alpha>0$. This turns out to be much more difficult to prove. Only recently have there been proofs of some results under reasonably general assumptions. These results apply to the cases $p>1$, and results relating to stability of potentials have not been obtained for the case $p=1$. The research shown in this thesis works in this direction.

The thesis is organized as follows. Chapter 2 contains a brief background on optimal transport. Chapter 3 is a survey of recent results on stability results that have been obtained recently in the cases of p greater than 1. Then, the final two chapters have novel results. Chapter 4 details results I have obtained on stability in the case $p=1$, and various assumptions on the source and target measure. Finally, chapter 5 shows some applications of these stability results, mainly focusing on a stable version of the transport convolution inequality.

2 Background on Optimal Transport and Other Results

In this section, we lay the basic theory of optimal transport. The main sources we will use throughout this chapter are the works of Ambrosio [1, 2], Santambrogio [12], and Villani [13].

2.1 Basic Definitions and Formulation

Let μ,ν be probability measures over some measure spaces X,Y respectively. We denote by $\Pi(\mu,\nu)$ the set of all joint probability measures π on $X\times Y$ that have μ,ν as their marginals. In concrete terms, $\pi\in\Pi(\mu,\nu)$ if and only if $\pi(X\times A)=\mu(A),\pi(B\times X)=\nu(B)$ for all measurable subsets $A\subseteq X,B\subseteq Y$. Note that $\pi(\mu,\nu)$ is always nonempty: For instance, the product measure $\mu\otimes\nu$ is in $\Pi(\mu,\nu)$. Elements of $\Pi(\mu,\nu)$ will also be called transport plans.

A cost function will be a measurable function $c : X \times Y \to [0, \infty]$. Given a transport plan $\pi \in \Pi(\mu, \nu)$ and a cost function c, we define the total cost of the plan as

$$I_c(\pi) = \int_{X \times Y} c(x, y) d\pi(x, y),$$

And the Kantorovich optimal transport cost between μ and ν is defined as

$$T_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} I_c(\pi). \tag{1}$$

It is possible to have $I_c(\pi) = \infty$ for all transport plans π, in which case $T_c(\mu, \nu) = \infty$. If $\pi \in \Pi(\mu, \nu)$ minimizes (1), it is called an optimal transport plan.

This is the Kantorovich formulation of optimal transport. The original problem Monge described was slightly different; informally, in the Kantorovich formulation, each $x \in X$ can map to multiple $y \in Y$, but Monge described each $x \in X$ mapping to a single $y \in Y$. More formally, we look at the set of all maps $T : X \to Y$ such that $T\#\mu = \nu$, which means that T is measurable and $\nu(A) = \mu(T^{-1}(A))$ for all measurable subsets $A \subseteq Y$. Denote this set as $M(\mu, \nu)$, and note that unlike the Kantorovich formulation, this set can be empty; for example, if μ is a Dirac mass, then $T\#\mu$ will also be a Dirac mass for any map T. Therefore, if μ is a Dirac mass but ν is not, $M(\mu, \nu)$ is empty. Given a transport map T and a cost function, we can analogously define its total cost to be

$$I_c(T) = \int_X c(x, T(x)) d\mu,$$

And the Monge optimal transport cost between μ and ν is defined as

$$T_c^m(\mu, \nu) = \inf_{T \in M(\mu, \nu)} I_c(T). \tag{2}$$

Similarly to the Kantorovich formulation, a minimizer to (2) is called an optimal transport map. If T is a transport map, we can define $\pi = (\mathrm{Id}, T)\#\mu$; then it is easy to check that $\pi \in \Pi(\mu, \nu)$, and $I_c(\pi) = I_c(T)$. Therefore, the Kantorovich formulation is broader than the Monge formulation; in particular, we always have $T_c(\mu, \nu) \leq T_c^m(\mu, \nu)$.

We will mostly be deal with the Monge formulation, i.e. with optimal transport maps. However, the Kantorovich formulation will be important to us. It is very hard to prove directly that the Monge problem admits a solution. As we will see, however, the Kantorovich problem has a solution in a broad number of cases, and it is then possible to prove that a solution to the Kantorovich problem is also a solution to the Monge problem.

2.2 Existence and Duality of Optimal Transport Plans

In this section, we prove the existence of optimal transport plans under rather general assumptions. We narrow down the spaces X, Y to be Polish spaces, which means that they are separable complete metric spaces. We use Polish spaces because probability measures on such spaces obey some convenient properties. In particular, Prokhorov's theorem, described below, allows us to use a compactness argument to prove the existence of a minimizer to the Kantorovich problem.

Theorem 1 (Prokhorov's Theorem)**.** *Let X be a Polish space. A set Σ of probability measures on X is compact in the weak topology if and only if for every $\varepsilon > 0$ there exists a compact subset $K \subseteq X$ such that $\mu(X \backslash K) \leq \varepsilon$ for every $\mu \in \Sigma$.*

The proof of Prokhorov's theorem is rather technical, see [11] for example. The special case where Σ consists of a single element is easier.

Lemma 1. *Let X be a Polish space, μ a probability measure on X. Then for every $\varepsilon > 0$ there exists a compact subset $K \subseteq X$ such that $\mu(X \backslash K) \leq \varepsilon$.*

Proof. Let $\{x_n\}_{n=1}^{\infty}$ be a countable dense subset of X. Then $\bigcup_{n=1}^{\infty} B_{1/m}(x_n) = X$ for every $m \in \mathbb{N}$, so there exists some $k_m > 0$ such that

$$\mu\left(X \backslash \bigcup_{n=1}^{k_m} B_{1/m}(x_n)\right) \leq \varepsilon/2^m$$

Denote $K_m = \bigcup_{n=1}^{k_m} B_{1/m}(x_n)$, and let $K = \bigcap_{m=1}^{\infty} K_m$. By definition, then, K is closed and totally bounded, hence compact; and

$$\mu(X \backslash K) \leq \sum_{m=1}^{\infty} \mu(X \backslash K_m) \leq \sum_{m=1}^{\infty} \frac{\varepsilon}{2^m} = \varepsilon.$$

$\square$

To state the theorem we want to prove, we also define a notion of semicontinuity.

Definition 1. *Let X be a Polish space. A function $f : X \to [-\infty, \infty]$ is called lower semicontinuous if*
$f(x) \leq \liminf_{n\to\infty} f(x_n)$ *whenever* $\lim_{n\to\infty} x_n \to x$.

There is another, equivalent definition to lower semicontinuity, at least in the case of nonnegative functions.

Lemma 2. *Let X be a Polish space. A function $f : X \to [0, \infty]$ is lower semicontinuous if and only if there exist bounded nonnegative continuous functions f_n such that $f = \sup_n f_n$.*

Proof. If $f = \sup_n f_n$, with f_n continuous, and $x_m \to x$, then

$$f(x) = \sup_n f_n(x) = \sup_n \lim_{m\to\infty} f_n(x_m) \leq \liminf_{m\to\infty} \sup_n f_n(x_m) = \liminf_{m\to\infty} f(x_m).$$

The other direction is a bit more involved. First, note that any nonnegative continuous function f is certainly the supremum of bounded continuous functions (simply take $f_n = \min\{f, n\}$). Therefore, it's enough to show that any nonnegative lower semicontinuous function is the supremum of nonnegative continuous functions. If $f \geq 0$ is lower semicontinuous and d is the distance on X, we define f_n by $f_n(x) = \inf_{y\in X} f(y) + nd(x, y)$. Then we need to verify some things: That f_n is continuous, that $f_n \leq f$, and that $\sup_n f_n = f$.

For continuity of f_n, note that fixing y, the function $f(y) + nd(x, y)$ is Lipschitz continuous and nonnegative, with Lipschitz constant n. This means that f_n is the infimum of Lipschitz functions with the same Lipschitz constant and which are bounded from below, so it is continuous, and even Lipschitz.

By definition, we have $f_n(x) \leq f(x) + nd(x, x) = f(x)$, so $\sup_n f_n(x) \leq f(x)$. The hard part is proving that $\sup_n f_n(x) \geq f(x)$, and this is where lower semicontinuity of f will come into play. Let $y_n \in X$ such that $f_n(x) \geq f(y_n) + nd(x, y_n) - 2^{-n}$. Then $f_n(x)$ is bounded by $f(x)$, so

$$nd(x, y_n) \leq f(x) + 2^{-n} - f(y_n) \leq f(x) + 2^{-n} \leq f(x) + 1.$$

In other words, $nd(x, y_n)$ is bounded as $n \to \infty$. Therefore, $d(x, y_n) \to 0$ as $n \to \infty$, or equivalently $y_n \to x$. By lower semicontinuity of f, we get

$$\sup_n f_n(x) \geq \liminf_{n\to\infty} f(y_n) + nd(x, y_n) - 2^{-n} \geq \liminf_{n\to\infty} f(y_n) \geq f(x).$$

□

Now, we can state the theorem proving existence of optimal transport plans.

Theorem 2. *Let X, Y be Polish spaces, μ, ν probability measures on X, Y respectively, $c : X \times Y \to [0, \infty]$ lower semicontinuous. Then there exists an optimal transport plan between μ, ν for the cost c.*

Proof. We first prove that $\Pi(\mu, \nu)$ is weakly compact. Indeed, by lemma 1, for every $\varepsilon > 0$ there exists compact subsets $K_1 \subseteq X, K_2 \subseteq Y$ such that $\mu(X\backslash K_1), \nu(Y\backslash K_2) \leq \varepsilon/2$. Then for every $\pi \in \Pi(\mu, \nu)$ we have

$$\pi((X \times Y)\backslash(K_1 \times K_2)) \leq \pi((X\backslash K_1) \times Y) + \pi(X \times (Y\backslash K_2)) = \mu(X\backslash K_1) + \nu(Y\backslash K_2) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

therefore, by Prokhorov's theorem $\Pi(\mu, \nu)$ is compact.

Now, let $\pi_n \in \Pi(\mu, \nu)$ be a sequence of transport plans such that $I_c(\pi_n) \to T_c(\mu, \nu)$. By compactness of $\Pi(\mu, \nu)$, there exists a subsequence π_{n_k} converging to some $\pi \in \Pi(\mu, \nu)$ in the weak topology. Also, by lemma 2, we can write $c = \sup_m c_m$, where c_m are bounded and continuous. Then by the monotone convergence theorem and the definition of weak convergence, we have

$$\begin{aligned} I_c(\pi) = \mathbb{E}_\pi[c(x, y)] = \mathbb{E}_\pi\left[\sup_m c_m(x, y)\right] &= \sup_m \mathbb{E}_\pi[c_m(x, y)] = \sup_m \lim_{k\to\infty} \mathbb{E}_{\pi_{n_k}}[c_m(x, y)] \\ \leq \liminf_{k\to\infty} \sup_m \mathbb{E}_{\pi_{n_k}}[c_m(x, y)] &= \liminf_{k\to\infty} \mathbb{E}_{\pi_{n_k}}[c(x, y)] = \liminf_{k\to\infty} I_c(\pi_{n_k}) = T_c(\mu, \nu). \end{aligned}$$

So $I_c(\pi) \leq T_c(\mu, \nu)$, and since $I_c(\pi) \geq T_c(\mu, \nu)$ by definition, we get that $I_c(\pi) = T_c(\mu, \nu)$, so that π is optimal. □

Next, we discuss a dual formulation to the Kantorovich problem.

Theorem 3. *Let X, Y be Polish spaces, μ, ν probability measures on X, Y, $c : X \times Y \to [0, \infty]$ lower semicontinuous. Then*

$$\min_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} c(x, y)d\pi = \sup_{(\varphi,\psi)\in\Phi_c(\mu,\nu)} \int_X \varphi d\mu + \int_Y \psi d\nu, \tag{3}$$

where

$$\Phi_c(\mu, \nu) = \{(\varphi, \psi) : \varphi \in L^1(\mu), \psi \in L^1(\nu), \varphi(x) + \psi(y) \leq c(x, y) \forall x \in X, y \in Y\}.$$

We call a maximizer to the right hand side in (3) a Kantorovich potential. They will prove useful to us later, when analyzing the Monge problem, since we will be able to construct optimal transport maps using the potentials.

Note that one direction of the equality in (3) is easy: If $\varphi(x) + \psi(y) \leq c(x, y)$, then for every $\pi \in \Pi(\mu, \nu)$:

$$\int_{X\times Y} c(x,y) d\pi \geq \int_{X\times Y} \varphi(x) + \psi(y) d\pi = \int_X \varphi(x) d\mu + \int_Y \psi d\nu.$$

The hard part, will therefore be to prove the other direction of the inequality.

To gain an idea of the proof, we can imagine fixing some function $\varphi : X \to [-\infty, \infty]$, and considering what is the largest corresponding ψ so that $\varphi(x) + \psi(y) \leq c(x, y)$ for all $x \in X, y \in Y$. Swapping sides, we get $\psi(y) \leq c(x, y) - \varphi(x)$; therefore, the largest value $\psi(y)$ can achieve is the infimum of $c(x, y) - \varphi(x)$, over all possible $x \in X$. This leads us to the following definition.

Definition 2. *Let $\varphi : X \to [-\infty, \infty]$ such that $\varphi(x) > -\infty$ for some $x \in X$. The c-transform of φ is defined as*

$$\varphi^c(y) = \inf_{x\in X} c(x, y) - \varphi(x).$$

We say that a function $\psi : Y \to [-\infty, \infty)$ is c-concave if $\psi = \varphi^c$ for some function φ. Similarly, for $\psi : Y \to [-\infty, \infty]$ such that $\psi(y) > -\infty$ for some $y \in Y$, we define

$$\psi^c(x) = \inf_{y\in Y} c(x, y) - \psi(y),$$

and $\varphi : X \to [-\infty, \infty)$ is c-concave if $\varphi = \psi^c$ for some function ψ.

Note that our definition of c-concavity includes the constant function $-\infty$; however we will mostly ignore this case.

There is a technical subtlety in the definition - it is not a priori obvious that φ^c is even measurable, since it is the supremum of a possibly uncountable family. However, if c is continuous and $y_n \to y$, then $-\varphi^c$ is lower semicontinuous, since it is the supremum of continuous functions. It is not hard to see that lower semicontinuity is equivalent to $(\varphi^c)^{-1}(-\infty, -a) = (-\varphi^c)^{-1}(a, \infty)$ being open for every $a \in \mathbb{R}$, which implies that φ^c is measurable in this case.

For convenience, we show some basic properties of the c-transform.

Proposition 1. *Let $\varphi : X \to [-\infty, \infty]$. Then*

a. For all $x \in X, y \in Y$ we have
$$\varphi(x) + \varphi^c(y) \leq c(x, y).$$

b. $\varphi \leq \varphi^{cc}$.

c. $\varphi^{ccc} = \varphi^c$.

d. $\psi : Y \to [-\infty, \infty)$ is c-concave if and only if $\psi^{cc} = \psi$.

Analogous properties also hold for $\psi : Y \to [-\infty, \infty]$.

Proof. (a) follows directly from the definition of the c-transform. Swapping sides, we get $\varphi(x) \leq c(x, y) - \varphi^c(y)$, and taking the infimum we see that

$$\varphi(x) \leq \inf_{y\in X} c(x, y) - \varphi^c(y) = \varphi^{cc}(x),$$

which proves (b). For (c), we have

$$\begin{aligned}
\varphi^{ccc}(y) &= \inf_{x\in X} c(x,y) - \varphi^{cc}(x) = \inf_{x\in X} c(x,y) - \inf_{y'\in X} c(x,y') - \varphi^{c}(y') \\
&= \inf_{x\in X} c(x,y) - \inf_{y'\in X} c(x,y') - \inf_{x'\in X} c(x',y') - \varphi(x') \\
&= \inf_{x\in X} \sup_{y'\in X} \inf_{x'\in X} c(x,y) - c(x,y') + c(x',y') - \varphi(x') \\
&\le \inf_{x,x'\in X} \sup_{y'\in X} c(x,y) - c(x,y') + c(x',y') - \varphi(x') \\
&\le \inf_{x\in X} \sup_{y'\in X} c(x,y) - c(x,y') + c(x',y') - \varphi(x) \\
&= \inf_{x\in X} c(x,y) - \varphi(x) = \varphi^{c}(x).
\end{aligned}$$

This proves that $\varphi^{ccc} \le \varphi^{c}$. By (b), we know that $\varphi^{ccc} \ge \varphi^{c}$, so we conclude that $\varphi^{ccc} = \varphi^{c}$. Finally, (d) is an easy consequence of (c). $\square$

To connect between optimal transport plans and c-concave functions, we introduce the notion of cyclical monotonicity.

Definition 3. *A set $\Gamma \subseteq X \times Y$ is c-cyclically monotone if, for any finite set of points $(x_1, y_1), \ldots, (x_k, y_k) \in \Gamma$, it holds that*

$$\sum_{i=1}^{k} c(x_i, y_i) \le \sum_{i=1}^{k} c(x_i, y_{i+1}),$$

where we define $y_{k+1} := y_1$.

Since every permutation is a product of cycles, this definition is equivalent to

$$\sum_{i=1}^{k} c(x_i, y_i) \le \sum_{i=1}^{k} c(x_i, y_{\pi(i)}),$$

for every $\pi \in S_k$.

The link between optimal transport and cyclical monotonicity is demonstrated in the following proposition.

Proposition 2. *a. Suppose c is continuous, and let μ, ν be probability measures over X, Y such that $T_c(\mu, \nu) < \infty$. If π is an optimal transport plan between μ and ν, then the support of π is c-cyclically monotone.*

b. Moreover, the union of the supports of all optimal transport plans between μ and ν is c-cyclically monotone.

Proof. We first prove (a). Assume otherwise, so there exist $(x_1, y_1), \ldots, (x_k, y_k) \in \operatorname{supp}(\pi)$ such that

$$\sum_{i=1}^{k} c(x_i, y_i) > \sum_{i=1}^{k} c(x_i, y_{i+1}).$$

Then by continuity of c and the definition of support, there exist open neighborhoods $x_i \in U_i, y_i \in V_i$ such that $\pi(U_i \times V_i) > 0$, the sets $U_i \times V_i$ are pairwise disjoint, and

$$\sum_{i=1}^{k} c(x_i', y_i') > \sum_{i=1}^{k} c(x_i', y_{i+1}'),$$

for any $x_i' \in U_i, y_i' \in V_i$. Let $A = \bigcup_{i=1}^{k} U_i \times V_i, B = \bigcup_{i=1}^{k} U_i \times V_{i+1}$. Let $\pi_i = \pi|_{U_i \times V_i}/\pi(U_i \times V_i)$. Define the measure σ to be the product measure of all the π_is on $\prod_{i=1}^{k} X \times Y$. Also denote by u_i, v_i the projections onto the i-th copy of X, Y respectively; then $(u_i, v_i)\#\sigma = \pi_i$. Now, for $\alpha > 0$ let π' be the signed measure defined by

$$\pi' := \pi + \alpha \sum_{i=1}^{k} ((u_i, v_{i+1})\#\sigma - (u_i, v_i)\#\sigma) = \pi + \alpha \sum_{i=1}^{k} ((u_i, v_{i+1})\#\sigma - \pi_i).$$

If α is small enough so that $\alpha\pi(U_i \times V_i) \leq 1$ for all i, then $\alpha\pi_i \leq \pi$ on $U_i \times V_i$, so π' is a positive measure. In fact, π' is a probability measure, since

$$\pi'(X \times Y) = \pi(X \times Y) + \alpha \sum_{i=1}^{k} ((u_i, v_{i+1})\#\sigma(X \times Y) - \pi_i(X \times Y)) = 1 + \alpha \sum_{i=1}^{k} (1 - 1) = 1.$$

Also, π' has the same marginals as π, since for example, denoting π_x to be the projection to the first coordinate coordinate,

$$\begin{aligned}
\pi_x\#\pi'(X \times Y) &= \pi_x\#\pi + \alpha \sum_{i=1}^{k} (\pi_x \circ (u_i, v_{i+1}))\#\sigma - \pi_x \circ (u_i, v_i))\#\sigma) \\
&= \pi_x\#\pi + \alpha \sum_{i=1}^{k} (u_i\#\sigma - u_i\#\sigma) = \pi_x\#\pi.
\end{aligned}$$

Now, a simple computation gives us:

$$\begin{aligned}
I(\pi') - I(\pi) &= \alpha \sum_{i=1}^{k} \left(\int_{X \times Y} c(x, y) d((u_i, v_{i+1})\#\sigma) - \int_{X \times Y} c(x, y) d((u_i, v_i)\#\sigma) \right) \\
&= \alpha \int_{\prod_{i=1}^{k} X \times Y} c(u_i, v_{i+1}) - c(u_i, v_i) d\sigma < 0,
\end{aligned}$$

which is a contradiction, since we assumed π was optimal.

For (b), let $(x_1, y_1), \ldots, (x_k, y_k)$ belong to optimal transport plans $\pi_1, \ldots, \pi_k$ respectively. Then

$$\pi := \frac{1}{n} \sum_{i=1}^{n} \pi_i,$$

is also an optimal transport plan, whose support contains $(x_1, y_1), \ldots, (x_k, y_k)$. Therefore (b) follows from (a). $\square$

The next definition draws inspiration from convex analysis.

Definition 4. *Let $\varphi : X \to [-\infty, \infty)$ be a c-concave function. The c-superdifferential of φ is defined for each $x \in X$ as*

$$\partial_c \varphi(x) = \{y \in Y : \varphi(z) \leq \varphi(x) + c(z, y) - c(x, y) \forall z \in X\},$$

and $\partial_c \varphi = \bigcup_{x \in X} \partial_c \varphi(x)$.

Note that if $\psi = \varphi^c$, then $\varphi(x) + \psi(y) = c(x, y)$ if and only if $y \in \partial_c \varphi(x)$. Indeed, we always have $\varphi(x) + \psi(y) \leq c(x, y)$, and

$$\begin{aligned} \varphi(x) + \psi(y) \geq c(x, y) \Leftrightarrow \\ \psi(y) \geq c(x, y) - \varphi(x) \Leftrightarrow \\ c(z, y) - \varphi(z) \geq c(x, y) - \varphi(x) \forall z \Leftrightarrow \\ \varphi(z) \leq \varphi(x) + c(z, y) - c(x, y) \forall z \Leftrightarrow \\ y \in \partial_c \varphi(x) \Leftrightarrow . \end{aligned}$$

Also, suppose φ is not identically $-\infty$, and $\varphi(x) = -\infty$. Then there exists some $z \in X$ such that $\varphi(z) > -\infty$, and then $\varphi(z) > \varphi(x) + c(x, z) - c(x, y)$ for all y, since the right hand side is always $-\infty$. Therefore, in this case $\partial_c \varphi(x)$ is empty.

The c-superdifferential allows us to prove the following theorem.

Theorem 4. *A set $\Gamma \subseteq X \times Y$ is c-cyclically monotone if and only if it is contained in the c-superdifferential of a c-concave function $\varphi : X \to [-\infty, \infty)$ that is not identically $-\infty$.*

Proof. First, assume $\Gamma \subseteq \partial_c \varphi$, and let $(x_1, y_1), \ldots, (x_k, y_k) \in \Gamma$. By the remark preceding the theorem, $\varphi(x_i) > -\infty$ for all i. Then we have

$$\varphi(x_{i-1}) - \varphi(x_i) \leq c(x_{i-1}, y_i) - c(x_i, y_i),$$

and summing over all possible i-s, the terms on the left hand side cancel out, and we are left with

$$0 \leq \sum_{i=1}^{k} c(x_{i-1}, y_i) - c(x_i, y_i) = \sum_{i=1}^{k} c(x_i, y_{i+1}) - c(x_i, y_i).$$

For the other direction, let Γ be c-cyclically monotone. Fix some $(x_0, y_0) \in \Gamma$. Define $\varphi : X \to [-\infty, \infty)$ as

$$\varphi(x) := \inf_{k \in \mathbb{N}, (x_1, y_1), \ldots, (x_k, y_k) \in \Gamma} c(x, y_1) + c(x_1, y_2) + \ldots + c(x_k, y_0) - c(x_0, y_0) - c(x_1, y_1) - \ldots - c(x_k, y_k).$$

Then, by c-cyclical monotonicity of Γ, we have $\varphi(x_0) = 0$, so φ is not identically $-\infty$.

Now, let $(x, y) \in \Gamma$, and $z \in X$. Take $(x_1, y_1) := (x, y)$ and $(x_2, y_2), \ldots, (x_k, y_k) \in \Gamma$, then

$$\begin{aligned} \varphi(z) &\leq c(z, y_1) + c(x_1, y_2) + \ldots + c(x_k, y_0) - c(x_0, y_0) - c(x_1, y_1) - \ldots - c(x_k, y_k) \\ &= c(z, y) - c(x, y) + c(x, y_2) + \ldots + c(x_k, y_0) - c(x_0, y_0) - c(x_2, y_2) - \ldots - c(x_k, y_k), \end{aligned}$$

and taking the infimum over all $(x_2, y_2), \ldots (x_k, y_k)$, we get

$$\varphi(z) \leq c(z, y) - c(x, y) + \varphi(x).$$

Since z was arbitrary, we have $y \in \partial_c \varphi(x)$, which is what we wanted to prove. □

Now, we can begin proving theorem 3. We first discuss a special case, where the supremum in (3) is actually a maximum.

Proposition 3. *Assume c is continuous, and that $T_c(\mu,\nu) < \infty$. Then there exists a c-concave pair of functions φ, ψ such that if π is an optimal transport plan between μ and ν, and (x,y) are in the support of π, then $\varphi(x) + \psi(y) = c(x,y)$. If, in addition, we assume that the set of all $y \in Y$ for which*

$$\int_X c(x,y)d\mu(x) < \infty, \tag{4}$$

has positive ν-measure, and similarly the set of all $x \in X$ for which

$$\int_Y c(x,y)d\nu(y) < \infty, \tag{5}$$

has positive μ-measure, then $\varphi \in L^1(\mu), \psi \in L^1(\nu)$, and we have

$$\int_{X\times Y} c(x,y)d\pi = \int_X \varphi d\mu + \int_Y \psi d\nu.$$

Proof. By proposition 2 and theorem 4, the union of the supports of all optimal transport plans is contained in the c-superdifferential of a c-concave function φ. If $\psi = \varphi^c$, then we have $y \in \partial_c\varphi(x)$ whenever $(x,y) \in \operatorname{supp}(\pi)$ for an optimal transport plan π, which as we saw is equivalent to $\varphi(x) + \psi(y) = c(x,y)$. Therefore we can try and write

$$\int_{X\times Y} c(x,y)d\pi = \int_X \varphi d\mu + \int_Y \psi d\nu. \tag{6}$$

To make this argument correct, we need to prove that the integrals on the right hand side actually exist, and that the expression is not of the form $\infty - \infty$. This would then prove the integrals are finite, since the left hand side is. To this end, we will show the integrals exist and are not ∞ by proving that the nonnegative parts φ^+, ψ^+ are integrable.

We first prove that $\varphi, \psi > -\infty$ almost everywhere. Indeed, if $\varphi(x) = -\infty$, then $\partial_c\varphi(x) = \emptyset$, but $\operatorname{supp}(\pi)$ is concentrated on the superdifferential of φ. If we define $A = \{x : \psi(x) = -\infty\}$, then this means that $A \times Y \cap \operatorname{supp}(\pi) = \emptyset$, so $\pi(A \times Y) = 0$, and therefore $\mu(A) = 0$. A similar argument holds for ψ.

Since $\varphi(x) + \psi(y) \leq c(x,y)$, we have $\varphi^+(x) \leq c(x,y) - \psi(y)$. Fixing y to be such that condition (4) holds and $\psi(y) \neq -\infty$, we see that $\varphi^+(x) \in L^1(\mu)$. Similarly, $\psi^+(x) \in L^1(\nu)$. Therefore, the integrals in (6) are not ∞, so they in fact converge. $\square$

Proof of theorem (3). For c that is bounded and continuous, the theorem follows from the previous result; in fact, in this case the dual problem admits a maximizer. For general c, we express $c = \sup_n c_n$, where c_n are bounded continuous functions. Then there exist $(\varphi_n, \psi_n) \in \Phi_{c_n}(\mu,\nu)$ that are maximizers to the dual problem for cost c_n. Passing to the limit, we get

$$\begin{aligned}\sup_{(\varphi,\psi)\in\Phi_c(\mu,\nu)} \int_X \varphi d\mu + \int_Y \psi d\nu &\geq \limsup_{n\to\infty} \int_X \varphi_n d\mu + \int_Y \psi_n d\nu = \limsup_{n\to\infty} \int_{X\times Y} c_n(x,y)d\pi \\ &\geq \int_{X\times Y} \lim_{n\to\infty} c_n(x,y)d\pi = \int_{X\times Y} c(x,y)d\pi,\end{aligned}$$

and the other direction of the duality was already proven. $\square$

This concludes the proof of the duality of optimal transport. The final result we will need relates back to cyclical monotonicity. As it turns out, this property is rather useful, and we would like to show optimal transport plans are concentrated on c-cyclically monotone sets even when c isn't continuous.

Proposition 4. *Suppose c is lower semicontinuous, and let π be an optimal transport plan. Then there exists a c-cyclically monotone set $\Gamma \subseteq X \times Y$ such that $\pi(\Gamma) = 1$.*

Proof. For the proof, we will use duality. Let $(\varphi_n, \psi_n) \in \Phi_c(\mu, \nu)$ be such that

$$\lim_{n\to\infty} \int_X \varphi_n d\mu + \int_Y \psi_n d\nu = \int_{X\times Y} c(x,y) d\pi,$$

then the functions $f_n(x,y) = c(x,y) - \varphi_n(x) - \psi_n(y)$ are nonnegative, and $f_n \to 0$ in $L^1(\pi)$. Therefore, taking a subsequence if necessary, we have $f_n \to 0$ π-almost everywhere. Let Γ be a set of full measure on which $f_n \to 0$. Then if $(x_1, y_1), \dots (x_k, y_k) \in \Gamma$, we have

$$\sum_{i=1}^k c(x_i, y_i) = \lim_{n\to\infty} \sum_{i=1}^k \varphi_n(x_i) + \psi_n(y_i) = \lim_{n\to\infty} \sum_{i=1}^k \varphi_n(x_i) + \psi_n(y_{i+1}) \leq \sum_{i=1}^k c(x_i, y_{i+1}).$$

□

For a single polish space $X = Y$, the most important examples of cost functions are $c_p(x,y) = d(x,y)^p$, where $p \geq 1$. We define $W_p = T_{c_p}(\mu,\nu)^{1/p}$. This turns out to define a distance, called the Wasserstein metric. The only difficult part is the triangle inequality. We will use the following theorem, allowing us to "glue" probability measures.

Theorem 5 (Lebesgue gluing theorem)**.** *Let μ_1, μ_2, μ_3 be probability measures over X_1, X_2, X_3 respectively. Let π_1 be a coupling of μ_1 and μ_2, and π_2 be a coupling of μ_2 and μ_3. Then there exists a unique joint probability distribution π on $X_1 \times X_2 \times X_3$, whose marginal on $X_1 \times X_2$ is π_1, and whose marginal on $X_2 \times X_3$ is π_2.*

If μ_1, μ_2, μ_3 are probability measures, π_1 is an optimal transport plan between μ_1, and π_2 is an optimal transport plan between μ_2 and μ_3, then let π be the joint probability measure given above, and let π_3 be the marginal of π on $X_1 \times X_3$. Then π_3 is a transport plan between μ_1 and μ_3, so by the Minkowski inequality

$$\begin{aligned} W_p(\mu_1,\mu_3) &\leq \|d(x,y)\|_{L^p(\pi_3)} = \|d(x_1,x_3)\|_{L^p(\pi)} \leq \|d(x_1,x_2)\|_{L^p(\pi)} + \|d(x_2,x_3)\|_{L^p(\pi)} \\ &= \|d(x,y)\|_{L^p(\pi_1)} + \|d(x,y)\|_{L^p(\pi_2)} = W_p(\mu_1,\mu_2) + W_p(\mu_2,\mu_3), \end{aligned}$$

which proves the triangle inequality.

2.3 Optimal Transport in Euclidean Space, $p > 1$

So far, our setting has been rather general, with measures over an arbitrary Polish space, cost functions which are arbitrary lower semicontinuous functions on that space. However, we will now restrict to the more concrete setting of $X = Y = \mathbb{R}^n$, and cost functions of the form $c(x,y) = |x-y|^p/p$, for some $p \geq 1$.

Our aim is to study the existence and uniqueness of optimal transport maps in the Euclidean setting. To state the theorems, we will need to define some notion of approximate differentiability. First, we define the density of a point in a set.

Definition 5. *Let $A \subseteq \mathbb{R}^n, x \in \mathbb{R}^n$. The density of A at x is defined to be*

$$\lim_{\varepsilon \to 0} \frac{|A \cap B_\varepsilon(x)|}{|A_\varepsilon(x)|}.$$

We recall also the well known Lebesgue density theorem.

Theorem 6 (Lebesgue density theorem)**.** *Let $A \subseteq \mathbb{R}^n$. Then almost every point $x \in A$ has density 1 in A.*

Now we define the notion of approximate differentiability.

Definition 6. *Let $f : \mathbb{R}^n \to [-\infty, \infty]$. We say that f is approximately differentiable at a point $x \in \mathbb{R}^n$ if there exists a function g differentiable at x, such that the set $\{y \in \mathbb{R}^n : f(y) = g(y)\}$ has density 1 at x. In this case, we denote $\tilde{\nabla} f(x) = \nabla g(x)$.*

Note that from this definition, it is obvious that if the approximate differential exists, it is unique since if g_1, g_2 are two functions that agree with f on a set with density 1 at x, then $g_1 = g_2$ on a set of density 1 at x, so $\nabla g_1(x) = \nabla g_2(x)$.

Aside from this, we will make use of some standard differentiability theorems, due to Rodemacher and Aleksandrov.

Theorem 7 (Rodemacher's theorem)**.** *Let U be an open subset of $\mathbb{R}^n$, and let $u : U \to \mathbb{R}$ be Lipschitz. Then u is differentiable almost everywhere.*

Theorem 8 (Aleksandrov's theorem)**.** *Let A be a convex subset of $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then f is twice differentiable almost everywhere, that is there exist $\nabla_A f(x), D^2_A f(x)$ such that*

$$f(z) = f(x) + \nabla_A f(x) \cdot (z - x) + D^2_A f(x)(z - x) \cdot (z - x) + o(|z - x|^2)$$

For almost every $x \in A$.

Now, we can state our main theorem. In the following, for $v \in \mathbb{R}^n, \alpha \in \mathbb{R}$, we write $v^\alpha = |v|^{\alpha-1} v$.

Theorem 9 (Optimal transport, the case $p > 1$)**.** *Let $p > 1$. Let μ, ν be probability measures on $\mathbb{R}^n$, with μ absolutely continuous. Then:*

- *There exists an almost everywhere unique optimal transport plan between μ and ν for the cost $c(x, y) = |x - y|^p/p$, and that plan is induced by a transport map T.*
- *There exists a c-concave function φ which is approximately differentiable μ-almost everywhere, and $T(x) = x - (\tilde{\nabla}\varphi)^{(1/(p-1))}$.*
- *The transport map T is approximately differentiable μ-almost everywhere, and the derivative $\tilde{\nabla} T$ is diagonalizable with nonnegative eigenvalues.*
- *If ν is also absolutely continuous, then T is injective almost everywhere and T^{-1} is the optimal transport map from ν to μ.*

Proof. By proposition 3, there exists a pair (φ, ψ) of c-concave functions, such that any optimal transport plan π between μ and ν is supported on the c-superdifferential of φ. Because φ is c-concave, we can write

$$\varphi(x) = \inf_{y \in \mathbb{R}^n} \frac{|x-y|^p}{p} - \psi(y). \tag{7}$$

Define

$$\varphi_R(x) = \inf_{y \in B_r(0)} \frac{|x-y|^p}{p} - \psi(y),$$

so φ_R is Lipschitz on $B_R(0)$, since it is the infimum of Lipschitz functions with the same Lipschitz constant. For μ-almost all $x \in \mathbb{R}^n$, the infimum on (7) is achieved for some $y \in \mathbb{R}^n$, so we have $\varphi(x) = \varphi_R(x)$ when $R > |y|$. Denoting $A_R := \{x \in B_r(0) : \varphi(x) = \varphi_R(x)\}$, by Lebesgue's density theorem and Rodemacher's theorem we see that φ is approximately differentiable at almost every point $x \in A_R$. Since μ is absolutely continuous, we get that φ is approximately differentiable μ-almost everywhere.

If $(x, y) \in \text{supp}(\pi)$, then we have

$$\varphi(x) + \psi(y) = \frac{|x-y|^p}{p},$$

so x maximizes

$$\frac{|x'-y|^p}{p} - \varphi(x'),$$

which means, by taking gradients, that

$$(x-y)^{(p-1)} = \tilde{\nabla}\varphi(x).$$

From here, it is obvious that y is unique almost everywhere and given by $y = x - (\tilde{\nabla}\varphi)^{(1/(p-1))}$.

For the proof of the third point, we will use Aleksandrov's theorem on the potential φ. Of course, this is not immediately applicable since φ is not convex. We will instead rely on a weak form of semi-concavity. For each triplet (B, B', B'') of open balls containing each other with rational centers and radii, let

$$N_{B,B',B''} = \{x \in B : T(x) \in B'' \backslash B'\},$$

and

$$\varphi_{B,B',B''} = \inf_{y \in B'' \backslash B'} \frac{|x-y|^p}{p} - \varphi^c(y).$$

Because the minimum defining φ is attained at $T(x)$, we have $\varphi_{B,B',B''} = \varphi$ on $N_{B,B',B''}$. Also, for $y \in B'' \backslash B'$, $|x-y|^p/p$ is differentiable on B and the derivative is C-Lipschitz, for some $C := C(B, B', B'') < \infty$ independent of y. Therefore, the functions

$$\frac{|x-y|^p}{p} - \varphi^c(y) - \frac{C}{2}|x|^2,$$

are all concave in x, and so is $\varphi_{B,B',B''} - \frac{C}{2}|x|^2$, since it is the infimum of concave functions. By Aleksandrov's theorem, $\varphi_{B,B',B''}$ is twice differentiable almost everywhere, and therefore φ is approximately twice differentiable almost everywhere on $A := \bigcup_{(B,B',B'')} N_{B,B',B''} = \{x : T(x) \neq x\}$. Since T is a smooth function of the approximate gradient of φ, it is approximately differentiable almost everywhere on A. Outside of A, T is the identity. Therefore it is approximately differentiable on any point with density 1, which is almost every point by the Lebesgue density theorem.

For simplicity, we denote $f(x) = \frac{|x-y|^p}{p}$ in what follows. We need to prove that $\tilde{\nabla} T$ is diagonalizable with nonnegative eigenvalues. This is clear if $T(x) = x$, so we assume otherwise. If φ is approximately twice differentiable at x, as shown before x minimizes $f(x' - y) - \varphi(x')$. Therefore, the hessian of this function is totally nonnegative, so

$$D^2 f(x - T(x)) \geq \tilde{D}^2 \varphi(x).$$

Letting f^* be the Legendre transform of f, we have that $\nabla f, \nabla f^*$ are inverses. Also note that $T(x) = x - \nabla f^*(\tilde{\nabla}\varphi(x))$, so

$$\left(D^2 f^*(\tilde{\nabla}\varphi(x))\right)^{-1} = D^2 f(\nabla f^*(\tilde{\nabla}\varphi)) = D^2 f(x - T(x)) \geq \tilde{D}^2 \varphi(x).$$

Define $A := D^2 f^*(x - T(x)), B := -\tilde{D}^2 \varphi(x)$. Note that by the chain rule, $\tilde{\nabla} T(x) = I + AB$, and we know that $A^{-1} \geq -B$. Let C be a symmetric positive definite matrix such that $C^2 = A$. Then we can write

$$C(\mathrm{id} + CBC)C^{-1} = CC^{-1} + C^2 BCC^{-1} = I + AB,$$

and since $I + CBC$ is symmetric, this proves that $\tilde{\nabla} T$ is diagonalizable. If v is any vector, we can write

$$\langle (I + CBC)v, v\rangle = \langle v, v\rangle + \langle Cv, BCv\rangle \geq |v|^2 \geq \langle Cv, A^{-1}Cv\rangle = |v|^2 - \langle v, v\rangle = 0,$$

which proves that all eigenvalues of $I + CBC$, and hence all eigenvalues of $I + AB$, are nonnegative.

For the fourth point, let π be the optimal transport plan induced by T; we can regard it as a transport plan π' from ν to μ, which is optimal as a transport plan. Since ν is absolutely continuous, by the first point, π' is induced by a transport map T' from ν to μ. Then the sets $\{(x, T(x) : x \in \mathbb{R}^n\}$ and $\{(y, T'(y)) : y \in \mathbb{R}^n\}$ have full π-measure, and so does their intersection; therefore $T(T'(x))$ for μ-almost all x, and $T'(T(y))$ for ν-almost all y. □

Let us discuss in more detail the implications of this theorem on the special case $p = 2$. Consider the dual problem. If φ, ψ are the potentials associated to μ, ν, then they maximize

$$\int \varphi d\mu + \int \psi d\nu,$$

subject to

$$\varphi(x) + \psi(y) \leq |x - y|^2/2 = \frac{1}{2}(|x|^2 + |y|^2) - x \cdot y.$$

Substituting $\varphi' = |x|^2/2 - \varphi$ and similarly for ψ, we end up with the equivalent minimization problem

$$\int \varphi' d\mu + \int \psi' d\nu,$$

subject to

$$\varphi'(x) + \psi'(y) \geq x \cdot y.$$

Recall that φ, ψ are c-transforms of each other. If we translate this to φ', ψ', we get

$$\psi'(y) = \frac{|y|^2}{2} - \psi(y) = \sup_x \frac{|y|^2}{2} - \frac{|x-y|^2}{2} + \varphi(x) = \sup_x x \cdot y - \frac{x^2}{2} + \varphi(x) = \sup_x x \cdot y - \varphi'(x) := (\varphi')^*(y),$$

and a similar expression for φ'. This star operation is known as the Legendre transform — note that it is the supremum of affine functions, so it defines a convex function, i.e. φ', ψ' are convex. By the theorem we proved before, we know that the optimal transport map is $x - \nabla\varphi = \nabla\varphi'$, so it is the gradient of a convex function. This map is called the Brenier map pushing μ to ν. The condition $\varphi(x) + \psi(T(x)) = c(x, y)$ turns into the Fenchell-Young inequality and equality:

$$\begin{aligned} &\varphi(x) + \psi(y) \geq x \cdot y, \\ &\varphi(x) + \psi(\nabla\varphi(x)) = x \cdot \nabla\varphi(x). \end{aligned}$$

We also note that as a consequence, the gradients $\nabla\varphi, \nabla\psi$ are inverses of each other.

Before moving on to the case $p = 1$, we will discuss interpolation. For two measures μ_0, μ_1, with μ_0 absolutely continuous, let T be the optimal transport map between μ_0 and μ_1 for the cost $|x - y|^p$. We interpolate between μ_0 and μ_1 by defining $T_t = tT + (1-t)\mathrm{id}$, and then $\mu_t = T_t \# \mu_0$. The relevant fact for us is that μ_t is absolutely continuous for every $t < 1$. This is a consequence of the following lemma:

Lemma 3. *Let μ_0, μ_1 be probability measures over $\mathbb{R}^n$, with μ_0 absolutely continuous, and let T be an optimal transport map from μ_0 to μ_1 for the cost $|x - y|^p$. Let $T_t(x) = tT(x) + (1-t)x$. Then for $t < 1$, T is injective μ_0-almost everywhere.*

Proof. The case $t = 0$ is obvious, so we may assume that $0 < t < 1$. Suppose that $T_t(x) = T_t(y)$. Denote $a = T(x) - x, b = T(y) - y$. Note that we have $x = T_t(x) - ta, y = T_t(y) - tb$. By c-cyclical monotonicity, for μ_0-almost all pairs x, y we have

$$\begin{aligned} |a|^p + |b|^p &= |T(x) - x|^p + |T(y) - y|^p \leq |T(y) - x|^p + |T(x) - y|^p \\ &= |ta + (1-t)b|^p + |tb + (1-t)a|^p \leq t|a|^p + (1-t)|b|^p + t|b|^p + (1-t)|a|^p = |a|^p + |b|^p, \end{aligned}$$

and by strict convexity of $|\cdot|^p$, this implies that $a = b$, and therefore $x = y$, μ_0-almost everywhere. ∎

2.4 Optimal Transport in Euclidean Space, $p = 1$

Next, we discuss the case $p = 1$. This subsection will also be relevant in sections 4 and 5, which are the crux of the thesis. In this case, there is no unique optimal transport plan. However, we have the following theorem.

Theorem 10 (Optimal transport, the case $p = 1$)**.** *Let μ, ν be probability measures on $\mathbb{R}^n$, with μ absolutely continuous. Then:*

a. *There exists a (not necessarily unique) optimal transport map T from μ to ν for the cost $c(x, y) = |x - y|$.*

b. *There exists a 1-Lipschitz function u such that, for any optimal transport map T, we have $|x - T(x)| = u(x) - u(T(x))$.*

While (a) will take some effort, (b) is quite easy to show, and follows from the fact that 1-Lipschitz functions are exactly the c-concave functions in the $p = 1$ case. Indeed if u is 1-Lipschitz, then $u(x) - u(y) \leq |x - y|$, so

$$u^c(y) = \inf_{x \in \mathbb{R}^n} |x - y| - u(x) \geq -u(y),$$

and

$$u^c(y) = \inf_{x \in \mathbb{R}^n} |x - y| - u(x) \leq |y - y| - u(y) = -u(y),$$

which proves that u is c-concave, and $u^c = -u$. On the other hand, an infimum of 1-Lipschitz functions is still 1-Lipschitz, as long as it is not $-\infty$.

According to proposition 3, there exists a 1-Lipschitz function u such that

$$u(x) - u(y) = u(x) + u^c(y) = |x - y|,$$

for every (x, y) in the support of an optimal transport plan π. This leads us to the following definition.

Definition 7. *Let u be 1-Lipschitz. A transport ray of u is a maximal segment $[x, y]$ in $\mathbb{R}^n$ such that $u(x) - u(y) = |x - y|$, maximal here meaning it is not contained in any larger such segment. We call a transport ray nondegenerate if $x \neq y$.*

Note that if $u(x) - u(y) = |x - y|$ and z lies between x and y, then

$$u(x) - u(z) = u(x) - u(y) - (u(z) - u(y)) \geq |x - y| - |z - y| = |x - z|.$$

On the other hand, $u(x) - u(z) \leq |x - z|$ because u is 1-Lipschitz; this shows that $u(z) = |x - z| - u(x)$, so u is linear on the segment $[x, y]$. This allows us to define a total ordering on each transport ray:

Definition 8. *Let u be 1-Lipschitz. We define a total order on each transport ray of u by $x \geq y$ if and only if $u(x) \leq u(y)$.*

If $(x, y) \in \operatorname{supp}(\pi)$, then as before $u(x) - u(y) = |x - y|$, so x, y belong to the same transport ray, and $x \leq y$ on this transport ray.

Given a transport ray $[x, y]$, we call the open segment (x, y) the interior of the transport ray.

Proposition 5. *Let u be a 1-Lipschitz function, and let $[x, y]$ be a transport ray. Then u is differentiable on its interior, and its derivative equals*

$$\nabla u = \frac{x - y}{|x - y|}.$$

Proof. Let $e := \frac{x-y}{|x-y|}$. The directional derivative of u along e is 1, since u increases linearly on $[x, y]$. We first prove the derivative in directions orthogonal to $x - y$ vanishes. Let z lie on (x, y), and let e' be a unit vector orthogonal to e. Define $\delta := u(z + te') - u(z)$, and let t_0 be small enough so that $z \pm t_0 e \in (x, y)$. Then

$$\delta + t_0 = u(z + te') - u(z) + u(z) - u(z - t_0 e) = u(z + te') - u(z - t_0 e) \leq |te' + t_0 e| = \sqrt{t^2 + t_0^2},$$

where the last step used the orthogonality of e and e'. Similarly,

$$\delta - t_0 = u(z+te') - u(z) + u(z) - u(z+t_0e) = u(z+te') - u(z+t_0e) \leq \sqrt{t^2+t_0^2}.$$

Squaring these inequalities, we get

$$\pm 2\delta t_0 \leq \delta^2 \pm 2\delta t_0 \leq t^2,$$

so that

$$|\delta| \leq \frac{t^2}{2t_0}.$$

Now, if $z' \in \mathbb{R}^n$ is close enough to z, we can write it as $z' = z'' + te'$, for some $t \in \mathbb{R}$, $z'' \in (x, y)$, and e' orthogonal to e. Also, if z' is close enough, we can pick a small t_0, depending only on z, so that $z'' \pm t_0 e$. Then by what we proved above, we have

$$u(z') - u(z) = u(z') - u(z'') + u(z'') - u(z) = O\left(t^2\right) + (z''-z)\cdot e = (z'-z)\cdot e + o(|z'-z|),$$

since $O(t^2) = o(t) = o(|z'-z|)$. □

Now, fix an optimal transport plan π, which among all optimal transport plans for the cost $|x-y|$, is optimal for the square cost $|x-y|^2$ (such a plan exists by compactness, though it is not immediate that it is unique). We want to prove that this plan is induced by a transport map. First, we prove the following monotonicity property.

Proposition 6. *Let π be a transport plan defined as above. Then π is concentrated on a set $\Gamma \subseteq \mathbb{R}^n \times \mathbb{R}^n$, with the following property: Let $(x_1, y_1), (x_2, y_2) \in \Gamma$, such that x_1, x_2, y_1, y_2 all belong to one transport ray. If $x_1 < x_2$, then $y_1 \leq y_2$.*

Proof. We can define a new cost, $\bar{c}$, by

$$\bar{c}(x,y) = \begin{cases} |x-y|^2 \text{ if } x \leq y, \\ \infty \text{ otherwise.} \end{cases}$$

Then $\bar{c}$ is lower semicontinuous, and the condition on π is equivalent to it being optimal for the cost $\bar{c}$. Therefore π is concentrated on a $\bar{c}$-cyclically monotone set Γ. We already know that $x_1 \leq y_1$ and $x_2 \leq y_2$, so we need to account for the possibility $x_1 < x_2 \leq y_2 < y_1$. In that case, $\bar{c}$-cyclical monotonicity gives us $(x_1 - x_2)\cdot(y_1 - y_2) \geq 0$, which implies $y_1 \leq y_2$, contradicting the assumption. □

Therefore, π is monotone on each transport ray. This also implies that the number of x for which there exist multiple y such that $(x, y) \in \Gamma$ is at most countable on each nondegenerate transport ray. We want to conclude from this that this set has measure zero. For this we will first need the following definition.

Definition 9. *A function f is called countably Lipschitz if there exists a sequence $\{A_n\}_{n=1}^{\infty}$ such that $\bigcup_{n=1}^{\infty} A_n$ has full measure and $f|_{A_n}$ is Lipschitz for each n.*

Proposition 7. *Suppose l_i is a collection of nondegenerate oriented intervals such that the direction $v(x)$ of the line passing through x is a countably Lipschitz function, and $B_i \subseteq l_i$ is countable for each i. Then $B := \bigcup B_i$ has measure zero.*

Proof. If we look at the countable set of hyperplanes orthogonal to vectors with rational coordinates, every nondegenerate line intersects at least one of these. By σ-additivity, we can assume that v is Lipschitz and that there exists a hyperplane Y intersecting each line in the collection. Let $f : A \to \mathbb{R}^n$ be defined as $f(x,t) = x + tv(x)$, defined on some set $A \subseteq Y \times \mathbb{R}$. Then, the preimage $f^{-1}(B)$ has only countably many points on each line of the form $\{x\} \times \mathbb{R}$. By Fubini's theorem, this implies that $f^{-1}(B)$ has measure zero, and therefore B has measure zero since the image of a measure zero set under a Lipschitz map is still measure zero. □

To complement this result, we want to prove that ∇u is countably Lipschitz. This will again make use of Aleksandrov's theorem.

Proposition 8. *Let u be 1-Lipschitz. Then ∇u is countably Lipschitz.*

Proof. Let E_h be the set of all points belonging to transport rays $[x,y]$ that are at distance at least h from the point x.

Let u_h denote the restriction of u to E_h. We define a new cost function c_h by $c_h(x,y) := f_h(x-y)$, where f_h is any C^2 function which is 1-Lipschitz and such that $f_h(x) = |x|$ if $|x| \geq h$. Choose λ_h to be greater than $\|D^2 f\|_\infty$. Then we can write

$$u(z) = |x - z| + u(x) = \inf_{x \in \mathbb{R}^n} f_h(x - z) + u(x),$$

and since the functions

$$f_h(x - z) + u(x) - \lambda_h \frac{|z|^2}{2},$$

are all concave, so is

$$u(z) - \lambda_h \frac{|z|^2}{2}.$$

Therefore, u is concave up to a smooth function. We finish by citing the following result. □

Proposition 9. *Let f be a convex function. Then ∇f is countably Lipschitz.*

Proof of theorem 10. Let π be the optimal transport plan for the distance cost $|x-y|$, which is also optimal for the cost $|x-y|^2$. Then, we proved there exists a set Γ of full π-measure, such that for every x, there are several options:

- x belongs to several transport rays, but then u is not differentiable at x; by Rodemacher's theorem, the set of all such x has Lebesgue measure 0.
- x belongs to a single nondegenerate transport ray, and there exist multiple y such that $(x,y) \in \Gamma$; by propositions 6, 7, and 8, the set of all such x also has Lebesgue measure 0.
- x belongs to a single nondegenerate transport ray, and there exists a single y on that ray such that $(x,y) \in \Gamma$.
- Finally, x can not belong to any nondegenerate transport ray, and in that case $u(x) - u(y) = |x - y|$ implies that $y = x$, i.e. y is determined uniquely by x.

Because μ is absolutely continuous, we get that for μ-almost all x, there exists a unique y such that $(x,y) \in \Gamma$. If we now define $T(x)$ to be that unique y, we get that π is induced by T, so T is an optimal transport map from μ to ν. □

Let us mention that according to proposition 6, for the transport map T constructed in the proof and $x_1 < x_2$, we have $T(x_1) \leq T(x_2)$. In other words, T is monotone on each transport ray.

Next, we define the transport density.

Definition 10. *Let μ, ν be probability measures, π an optimal transport plan. Then the transport density (for some optimal transport plan π) is defined by*

$$\sigma = \int_0^1 \pi_t \#(|y - x|\pi)dt,$$

where $\pi_t(x, y) = ty + (1 - t)x$.

Since the definition of σ depends on the choice of an optimal transport plan π, it is not necessarily unique, however see proposition 11 below.

One notable fact about the transport density σ is that it obeys a sort of partial differential equation.

Proposition 10. *Let μ, ν be probability measures, and u, σ the potential and transport density associated with μ, ν. Then*

$$-\nabla \cdot (\nabla u \cdot \sigma) = \mu - \nu,$$

where this equation is taken to hold in the sense of distributions.

Proof. Define E_t to be equal to $\pi_t \#(y - x)\pi$, so that $\sigma = \int_0^1 E_t dt$. Also define $\mu_t = \pi_t \# \pi$. For some test function $\varphi \in C_c^\infty$, we calculate:

$$\frac{d}{dt} \int_{\mathbb{R}^n} \varphi d\mu_t = \frac{d}{dt} \int_{\mathbb{R}^n \times \mathbb{R}^n} \varphi(ty + (1-t)x) d\pi = \int_{\mathbb{R}^n \times \mathbb{R}^n} \nabla\varphi(ty + (1-t)x) \cdot (y - x) = \int_{\mathbb{R}^n \times \mathbb{R}^n} \nabla\varphi dE_t.$$

Interpreting this equation in the sense of distributions, we get that

$$\frac{d}{dt}\mu_t = -\nabla \cdot E_t,$$

Note that $-E_t = \nabla u |E_t|$; this follows from the fact that if $(x, y) \in \text{supp}(\pi)$, then $u(x) - u(y) = |x - y|$, so $y - x$ is a positive multiple of $-\nabla u(x) = -\nabla u(y)$. Therefore, we get

$$\frac{d}{dt}\mu_t = \nabla \cdot (\nabla u |E_t|).$$

Finally, integrating t from 0 to 1 we get

$$\nu - \mu = \mu_1 - \mu_0 = \int_0^1 \frac{d}{dt}\mu_t dt = \int_0^1 \nabla \cdot (\nabla u |E_t|) dt = \nabla \cdot \left(\nabla u \int_0^1 |E_t| dt \right) = \nabla \cdot (\nabla u \cdot \sigma).$$

□

We will not use this, but this equation allows us to prove that the transport density is unique if μ or ν or absolutely continuous:

Proposition 11. *Let μ or ν be absolutely continuous. Then σ is independent of π, and therefore is unique.*

Proof. The proof uses the theory of disintegration of measures. The idea is that given a map $f : X \to Y$, a measure μ on X, and $\nu = f\#\mu$, then we can disintegrate μ alongside the fibers of f. We write $\mu = \mu_y \otimes \nu$, where μ_y is concentrated on $f^{-1}(y)$, and we take this to mean, for each $A \subseteq X$ measurable,

$$\mu(A) = \int_Y \mu_y(A) d\nu(y).$$

The Lebesgue disintegration theorem is the central result we will use. It states that this decomposition always exists, and is unique almost everywhere.

Now, we explain how to use the disintegration theorem. Without loss of generality, assume μ is absolutely continuous. Let S be the set of all points belonging to a single transport ray. Then S has full measure, by Rodemacher's theorem. Define a multivalued function $R : \mathbb{R}^n \times \mathbb{R}^n \to T_u$ by sending (x, y) to the transport rays containing x; then this function is uniquely defined μ-almost everywhere inside $S \times \mathbb{R}^n$, so by absolute continuity of μ, it is well-defined μ-almost everywhere. Let $\lambda = R\#\pi$. Then disintegrate π as $\pi = \pi_r \otimes \lambda$, and similarly $\sigma = \sigma_r \otimes \lambda$. Let the marginals of π_r be μ_r, ν_r. Then, note that σ_r is the transport density associated with μ_r. However, on one dimension, this means that σ_r depends only on μ_r, ν_r, by proposition 10. Therefore, to finish the proof, we need to prove that λ does not depend on π, and that μ_r, ν_r do not depend on π for λ-almost all r.

For λ, this is easy: Note that since R only depends on the first coordinate, $\lambda(A) = \pi(R^{-1}(A)) = \pi(A \times \Omega) = \mu(A)$. Also, we can write $\mu = \mu_r \otimes \lambda$, so by uniqueness of disintegration μ_r does not depend on π. For ν_r the argument is harder; note that we can not write $\nu = \nu_r \otimes \lambda$ since ν is not necessarily absolutely continuous, and hence R is not defined ν-almost everywhere. Still, we can write

$$\nu_r = \nu_r^S + \nu_r^{S^c},$$

where ν_r^S is the second marginal of $\pi_r|_{\mathbb{R}^n \times S}$, and $\nu_r^{S^c}$ is the second marginal of $\pi_r|_{\Omega \times S^c}$. Now, similarly to before, we have $\nu|_S = \nu_r^S \otimes \lambda$, and so ν_r^S does not depend on π. For $\nu_r^{S^c}$, note that this measure must be concentrated on the endpoints of the transport ray, since every other point on the ray is differentiable. In fact, for λ-almost all r, $\nu_r^{S^c}$ cannot give mass to the lower endpoint. This is because, denoting A to be the subset of S^c consisting of all $x \in \mathbb{R}^n$ which are lower endpoints of such rays, we see that

$$0 = \mu(S^c) \geq \mu(A) = \int_{R(\mathbb{R}^n \times \mathbb{R}^n)} \mu_r(A) d\lambda(r) = \int_{R(S^c \times \mathbb{R}^n)} \mu_r(\{x(r)\}) d\lambda(r),$$

where $x(r)$ is the lower endpoint of r. This proves the integrand is zero λ-almost everywhere.

To summarize, $\nu_r^{S^c}$ does not give mass to the lower endpoint of r, so it is a Dirac mass concentrated on the endpoint of the ray. However ν_r is a probability measure, and therefore the mass of ν_r^S is equal to 1 minus the mass of $\nu_r^{S^c}$, which is not dependent on π, as we proved. Therefore we are done. □

We will use the following uniqueness result.

Proposition 12. *Let μ or ν be absolutely continuous and with bounded supports. Then σ is absolutely continuous.*

Proof. Without loss of generality, assume μ is absolutely continuous, and let T be the optimal transport plan from μ to ν which is monotone on each transport ray. Let π be the transport plan induced by T, and define $\mu_t := ((1-t)\text{id} + tT)\#\mu = \pi_t\#\pi$. Then it follows that

$$\sigma \le C \int_0^1 \mu_t dt,$$

where C is chosen to be larger than $|y-x|$ for all $x \in \text{supp}(\mu), y \in \text{supp}(\nu)$; this is possible since we assumed μ, ν are supported on bounded sets. It is enough, then, to prove that μ_t is absolutely continuous for almost all t, and we will indeed prove it is absolutely continuous for all $t < 1$. We first argue when $\nu = \{y_1, \ldots, y_k\}$ is discrete. Define $\Omega_i := T^{-1}(y_i)$, and $\Omega_i(t) = T_t(\Omega_i)$. If x belongs to two different $\Omega_i(t)$s, then it must lie on the line between two points y_i, y_j in the support of ν. Denoting $x = T_t(x_i) = T_t(x_j)$, where $T(x_i) = y_i, T(x_j) = y_j$, we see that x_i, x_j, x, y_i, y_j all lie on the same line. If $y_i < y_j$, then $x_i < x_j$ by monotonicity, but then $x = T_t(x_i) < T_t(x_j) = x$, which is a contradiction.

Therefore, $\Omega_i(t)$ are all disjoint for each $t < 1$. Then for each subset $A \subseteq \mathbb{R}^n$, we have

$$\mu_t(A) = \sum_{i=1}^k \mu_t(A \cap \Omega_i(t)) = \sum_{i=1}^k \mu_t\left(\frac{A \cap \Omega_i(t) - tx_i}{1-t}\right) = \mu\left(\bigcup_{i=1}^k \frac{A \cap \Omega_i(t) - tx_i}{1-t}\right).$$

By translation invariance of the Lebesgue measure, we have

$$\left|\bigcup_{i=1}^k \frac{A \cap \Omega_i(t) - tx_i}{1-t}\right| \le \frac{|A|}{(1-t)^n}.$$

Because μ is absolutely continuous, for every ε there exists $\delta := \delta_\varepsilon > 0$ such that $|A| \le \delta$ implies $\mu(A) \le \varepsilon$. Then, choosing $\delta_\varepsilon(t) = \delta_\varepsilon/(1-t)^n$, we see that $|A| \le \delta_\varepsilon(t)$ implies that $\mu_t(A) \le \varepsilon$, so μ_t is absolutely continuous. If ν is a general measure, we can express it as a weak limit of discrete measures ν^n. If A is open and $|A| \le \delta_\varepsilon(t)$, then

$$\mu_t(A) \le \liminf_{n\to\infty} \mu_t^n(A) \le \varepsilon,$$

which shows that μ_t is absolutely continuous. □

2.5 Entropic Optimal Transport

Next, we discuss a variation on the optimal transport problem, called entropic optimal transport. The content of this subsection is taken from [9].

We first define the concept of relative entropy. For two probability measures π, ρ on a space X, we define the relative entropy to be

$$H(\pi|\rho) = \int_X \ln\left(\frac{d\pi}{d\rho}\right) d\pi = \mathbb{E}_\rho\left[\frac{d\pi}{d\rho} \ln\left(\frac{d\pi}{d\rho}\right)\right],$$

and if π is not absolutely continuous with respect to ρ, we define $H(\pi|\rho) = \infty$. By convexity of $h(z) = z \ln z$, we see that $H(\cdot|\rho)$ is convex and nonnegative (by Jensen's inequality).

Lemma 4 (Pinsker's inequality). *$d_{TV}(\mu, \nu) \le \sqrt{2H(\mu|\nu)}$. Here the total variation d_{TV} is the L^1 distance between the densities, according to some (arbitrary) reference measure.*

Proof. If μ is not absolutely continuous with respect to ν, the inequality is trivial. Therefore, assume μ is absolutely continuous with respect to ν, and let $f = \frac{d\mu}{d\nu}$. We have the inequality $3(x-1)^2 \le (4+2x)\cdot(x\ln x - x + 1)$, so by Hölder's inequality

$$\begin{aligned}
3\|\mu-\nu\|_{TV}^2 &= \mathbb{E}_\nu\left[\sqrt{3}|f-1|\right]^2 \le \mathbb{E}_\nu\left[\sqrt{4+2f}\cdot\sqrt{f\ln f - f + f}\right]^2 \\
&\le \mathbb{E}_\nu[4+2f]\cdot\mathbb{E}_\nu[f\ln f - f + 1] = 6H(\mu|\nu).
\end{aligned}$$

□

Proposition 13. *Suppose that π_n is a sequence of probability distributions on a space X, such that $H(\pi_n|\rho)$ is bounded. Then π_n has a weakly convergent subsequence.*

Proof. Denote by f_n the Radon-Nikodym derivatives of π_n. Then we know that $\mathbb{E}_\rho[h(f_n)]$ is bounded by $M < \infty$. Let $\varepsilon > 0$, and choose $C = e^{-\varepsilon}$. Then we have that $\varepsilon h(x) > x$ for $x > C$, and therefore

$$\int_{f_n(x)>C} f_n d\rho \le \varepsilon \int_{f_n(x)>C} h(f_n)d\rho \le \varepsilon\int_X h(f_n)d\rho \le \varepsilon M.$$

This shows that the sequence f_n is uniformly equiintegrable, which is known to imply that f_n has a weakly convergent subsequence. □

Theorem 11. *Suppose that π_n is a sequence of probability measures, such that*

$$\limsup_{m,n\to\infty} H\left(\frac{\pi_n+\pi_m}{2}\middle|\rho\right) \ge \lim_{n\to\infty} H(\pi_n|\rho). \tag{8}$$

Then the Radon-Nikodym derivatives $f_n = \frac{d\pi_n}{d\rho}$ converge in L^1, or equivalently the π_n converge in total variation.

Proof. Denote $\pi_{n,m} = (\pi_n+\pi_m)/2, f_{n,m} = (f_n+f_m)/2$. Then we have

$$\begin{aligned}
2H(\pi_{n,m}|\rho) &= \mathbb{E}_\rho\left[f_n \ln\left(\frac{d\pi_{n,m}}{d\rho}\right)\right] + \mathbb{E}_\rho\left[f_m \ln\left(\frac{d\pi_{n,m}}{d\rho}\right)\right] \\
&= \mathbb{E}_\rho\left[f_n \ln\left(\frac{d\pi_{n,m}}{d\pi_n}\right)\right] + \mathbb{E}_\rho\left[f_n \ln\left(\frac{d\pi_n}{d\rho}\right)\right] \\
&\quad + \mathbb{E}_\rho\left[f_m \ln\left(\frac{d\pi_{n,m}}{d\pi_m}\right)\right] + \mathbb{E}_\rho\left[f_m \ln\left(\frac{d\pi_m}{d\rho}\right)\right] \\
&= -H(\pi_n|\pi_{n,m}) + H(\pi_n|\rho) - H(\pi_m|\pi_{n,m}) + H(\pi_m|\rho),
\end{aligned}$$

and after rearranging:

$$H(\pi_n|\rho) + H(\pi_m|\rho) - 2H(\pi_{n,m}|\rho) = H(\pi_n|\pi_{n,m}) + H(\pi_m|\pi_{n,m}).$$

Note that the left hand side converges to 0, by (8) and the fact that $H(\cdot|\rho)$ is convex. Then the terms on the right hand side also converge to 0. By Pinsker's inequality, we have

$$\begin{aligned}
d_{TV}(\pi_n,\pi_m) = \|f_n - f_m\|_{L^1(\rho)} &\le \|f_n - f_{n,m}\|_{L^1(\rho)} + \|f_{n,m} - f_m\|_{L^1(\rho)} \\
&\le \sqrt{2H(\pi_n|\pi_{n,m})} + \sqrt{2H(\pi_m|\pi_{n,m})} \to 0.
\end{aligned}$$

Therefore, π_n is a Cauchy sequence, so it converges. □

Theorem 12. *Suppose that K is a set of probability measures absolutely continuous with respect to ρ, such that the set of densities of measures in K is convex and closed in $L^1(\rho)$. Suppose also there exists $\mu \in K$ such that $H(\mu|\rho) < \infty$. Then*

a. *there exists a unique minimizer π^* for $H(\cdot|\rho)$ in K.*

b. *If π_0 is such that, denoting $f_0 = \frac{d\pi_0}{d\rho}$, we have $\mathbb{E}_\pi[\ln f_0] \geq H(\pi_0|\rho)$ for all $\pi \in K$, then $\pi_0 = \pi^*$.*

Proof. a. Let a be the infimum of $H(\pi|\rho)$ among $\pi \in K$, and let π_n be a sequence such that $\lim_{n\to\infty} H(\pi_n|\rho) = a$. Then $H(\frac{\pi_n+\pi_m}{2}|\rho) \geq a$ by assumption, but on the other hand by convexity $H(\frac{\pi_n+\pi_m}{2}|\rho) \leq \frac{1}{2}H(\pi_n|\rho) + \frac{1}{2}H(\pi_m|\rho) \to a$. Therefore the conditions for theorem 11 are fullfilled, and we have $\pi_n \to \pi^* \in K$, which must then be a minimizer. Uniqueness follows by the strict convexity of $h(x) = x \ln x$.

b. Let $\pi \in K$ such that $H(\pi|\rho) < \infty$. We calculate:

$$\begin{aligned} H(\pi|\rho) &= \mathbb{E}_\pi\left[\ln\left(\frac{d\pi}{d\rho}\right)\right] = \mathbb{E}_\pi\left[\ln\left(\frac{d\pi}{d\pi_0}\right)\right] + \mathbb{E}_\pi\left[\ln\left(\frac{d\pi_0}{d\rho}\right)\right] \\ &= H(\pi|\pi_0) + \mathbb{E}_\pi[\ln f_0], \end{aligned}$$

so

$$H(\pi|\rho) \geq H(\pi|\rho) - H(\pi|\pi_0) = \mathbb{E}_\pi[\ln f_0] \geq H(\pi_0|\rho).$$

This inequality certainly holds if $H(\pi|\rho) = \infty$, so we get that π_0 minimizes $H(\pi_0|\rho)$.

□

We state the following corollaries to the theorem.

Corollary 1. *Suppose that $H(\pi_0|\rho) < \infty$, and $\mathbb{E}_\pi[\ln f_0]$ is constant on the set of measures in K with finite entropy. Then $\pi_0 = \pi^*$.*

Proof. This follows from part (b) of the theorem. □

Corollary 2. *Let $\psi_1, \ldots, \psi_n : \mathbb{R}^n \to \mathbb{R}$, and let $K = \{\pi : \mathbb{E}_\pi[\psi_i] = 0 \ \forall 1 \leq i \leq n\}$. Then K has a unique minimizer of the form*

$$\frac{d\pi^*}{d\rho} = ae^{b_1\psi_1 + \ldots + b_n\psi_n}, \tag{9}$$

for any probability measure ρ. Also, π^* is the unique measure in K expressible this way.

Proof. If such an expression exists, by the previous corollary π^* is indeed the optimizer. The proof that such an expression exists involves Lagrange multipliers, see [9, Example 1.18] for more details. □

Returning to optimal transport, we now consider the case where $K = \Pi(\mu, \nu)$, for some probability measures $\mu \in X, \nu \in Y$. We assume that X, Y are polish spaces. We also write $\rho_1 \sim \rho_2$ to mean that ρ_1 is absolutely continuous with respect to ρ_2, and vice versa.

Theorem 13. *Suppose that $\pi^* \in \Pi(\mu,\nu)$ is an optimizer for $H(\pi|\rho)$, where ρ is a probability measure on $X \times Y$ which satisfies $\rho \sim \mu \otimes \nu$. Then π^* can be uniquely characterized as the unique probability measure in $\Pi(\mu,\nu)$ for which the exist functions $\varphi \in L^1(\mu), \psi \in L^1(\nu)$ such that*

$$\frac{d\pi^*}{d\rho} = e^{\varphi+\psi},$$

and moreover, φ, ψ are unique up to an additive constant.

Proof. Because X and Y are separable, the conditions defining $\Pi(\mu,\nu)$ can be summarized as

$$\int_X f_i d\pi = 0, \int_Y g_i d\pi = 0, \tag{10}$$

for suitable countable dense families f_i, g_i. Denote $K = \Pi(\mu,\nu)$, and let K_n be the set of all probability measures π obeying (10) for $i = 1, \dots, n$. Then K_n are closed and convex, and $K = \bigcap_{n=1}^{\infty} K_n$. For each K_n, by corollary 2, the minimizer π_n of $H(\pi|\rho)$ among $\pi \in K_n$ is of the form $\frac{d\pi_n}{d\rho} = e^{\varphi_n + \psi_n}$ for suitable φ_n, ψ_n. Now note that $\frac{\pi_n + \pi_m}{2} \in K_{\max(n,m)}$, so $\limsup_{n,m\to\infty} H\left(\frac{\pi_n+\pi_m}{2}|\rho\right) \geq H(\pi^*|\rho)$, which means the conditions of 11 are fulfilled, and the densities of π_n converge in L^1 to the density of π^*. Recalling that convergence in L^1 implies convergence almost everywhere of a subsequence, we can write

$$\frac{d\pi^*}{d\rho} = \lim_{n\to\infty} e^{\varphi_n + \psi_n}$$

ρ-almost everywhere. So, we have that $\varphi_n + \psi_n$ converges, and want to deduce that φ_n, ψ_n converge. Here a somewhat technical measure-theoretic argument is required; if the convergence holds everywhere, then we can force $\varphi_n(x^*) = 0$ for some $x^* \in X$, and now

$$\varphi_n(x) = \varphi_n(x) + \psi_n(y) - \varphi_n(x^*) - \psi_n(y)$$

And the right hand side obviously converges, which implies that φ_n converges to some measurable φ, and in that case $\psi_n = (\varphi_n + \psi_n) - \varphi_n$ also converges. However, this argument requires some modification in the case where the convergence holds merely ρ-almost everywhere.

Specifically, in this case, we have that the set A of (x,y) for which $\varphi_n(x) + \psi_n(y)$ converges has full measure. For each $x \in X$, let A_x denote the set of all y such that $(x,y) \in A$. Then by Fubini's theorem, almost all the sets A_x are of full measure, which proves the theorem. $\square$

Since ρ is absolutely continuous with respect to $\mu \otimes \nu$, we can write $e^{-c(x,y)} = \frac{d\rho}{d(\mu\otimes\nu)}$, where $c(x,y) : X \times X \to \mathbb{R} \cup \{\infty\}$.

If $\frac{d\pi}{d\rho} = e^{\varphi+\psi}$, then $\pi = e^{\varphi+\psi}\rho = e^{\varphi+\psi-c}(\mu \otimes \nu)$. The marginals of π are then

$$\pi_1(\pi) = e^{\varphi(x)} \int_X e^{\psi(y)-c(x,y)} d\nu(y) \cdot \mu,$$
$$\pi_2(\pi) = e^{\psi(y)} \int_X e^{\varphi(x)-c(x,y)} d\mu(x) \cdot \nu.$$

If $\pi \in \Pi(\mu,\nu)$, then this leads us to the following equations, called the Schrödinger equations:

$$\varphi(x) = -\ln \int_X e^{\psi(y)-c(x,y)} d\nu(y),$$
$$\psi(y) = -\ln \int_X e^{\varphi(x)-c(x,y)} d\mu(x).$$

2.6 Other results

We will need a couple of other results in our study. We recall that a measure μ over $\mathbb{R}^n$ is called log-concave if it is absolutely continuous, and its density is of the form $e^{-\psi}$, where $\psi : \mathbb{R}^n \to \mathbb{R}$ is a convex function. This can be seen to be equivalent to the density f satisfying $f(ty + (1-t)x) \geq f(y)^t f(x)^{1-t}$ for all $x, y \in \mathbb{R}^n, 0 \leq t \leq 1$.

Theorem 14 (Brascamp-Lieb Inequality)**.** *Let μ be a log-concave probability measure, with density $e^{-\psi}$ over $\mathbb{R}^n$. Let $f \in C^2 \cap L^1(\mu)$. Then*

$$\mathrm{Var}_\mu(f) \leq \mathbb{E}[\nabla f \cdot (D^2\psi)^{-1} \nabla f]. \tag{11}$$

Proof. We will first define a sort of Laplacian corresponding to μ. If $u, v \in C_c^\infty$, we have by integration in parts

$$\begin{aligned}\int_{\mathbb{R}^n} \partial_i u \cdot v d\mu &= \int_{\mathbb{R}^n} D_i u \cdot v \cdot e^{-\psi} dx \\ &= -\int_{\mathbb{R}^n} u \left(\partial_i v \cdot e^{-\psi} - v \cdot \partial_i \psi \cdot e^{-\psi}\right) \\ &= -\int_{\mathbb{R}^n} u \left(\partial_i v - \partial_i \psi \cdot v\right) d\mu.\end{aligned} \tag{12}$$

Therefore, if we define $\partial_i^* v = \partial_i v - \partial_i \psi \cdot v$, (12) gives us

$$\int_{\mathbb{R}^n} \partial_i u \cdot v d\mu = -\int_{\mathbb{R}^n} u \cdot \partial_i^* v d\mu. \tag{13}$$

We now define the Laplacian operator (with respect to μ) to be

$$L(u) = \sum_{k=1}^{n} \partial_i^*(\partial_i u).$$

Note that by 13, the Laplacian operator satisfies the identity

$$\begin{aligned}\int_{\mathbb{R}^n} L(u) \cdot v d\mu &= \sum_{i=1^n} \int_{\mathbb{R}^n} \partial_i^*(\partial_i u) \cdot v d\mu \\ &= -\sum_{i=1}^{n} \int_{\mathbb{R}^n} \partial_i(u) \cdot \partial_i(v) d\mu = -\int_{\mathbb{R}^n} \nabla u \cdot \nabla v d\mu.\end{aligned}$$

We want to compute the partial derivatives of $L(u)$. To do this, we first compute how ∂_i, ∂_k^* commute:

$$\begin{aligned}\partial_i(\partial_k^* u) &= \partial_i(\partial_k u - \partial_k \psi \cdot u) = \partial_{ki} u - \partial_{ik}\psi \cdot u - \partial_k \psi \cdot \partial_i u \\ &= \partial_k^*(\partial_i u) - \partial_{ik}\psi \cdot u.\end{aligned}$$

Therefore,

$$\begin{aligned}\partial_i L(u) &= \sum_{k=1}^{n} \partial_i(\partial_k^*(\partial_k u)) = \sum_{k=1}^{n} \partial_k^*(\partial_{ik} u) - \partial_{ik}\psi \cdot \partial_k u \\ &= L(\partial_i u) - \sum_{k=1}^{n} \partial_{ik}\psi \cdot \partial_k u.\end{aligned}$$

Now, we have enough tools to compute the L_2 norm of $L(u)$:

$$\begin{aligned}\int_{\mathbb{R}^n} L(u)^2 d\mu &= -\sum_{i=1}^n \int_{\mathbb{R}^n} \partial_i u \cdot \partial_i L(u) d\mu \\ &= -\sum_{i=1}^n \int_{\mathbb{R}^n} \partial_i u \cdot L(\partial_i u) d\mu + \sum_{i,k=1}^n \int_{\mathbb{R}^n} \partial_{ik}\psi \cdot \partial_i u \cdot \partial_k u d\mu \\ &= \int_{\mathbb{R}^n} (D^2\psi)\nabla u \cdot \nabla u d\mu + \sum_{i=1}^n \int_{\mathbb{R}^n} \|\partial_i u\|^2 d\mu.\end{aligned}$$

This implies in particular the following inequality:

$$\int_{\mathbb{R}^n} L(u)^2 d\mu \geq \int_{\mathbb{R}^n} (D^2\psi)\nabla u \cdot \nabla u d\mu. \tag{14}$$

To deduce (11) from (14), we use the density of functions of the form $L(u)$ in the subspace of zero-mean functions in $L^2(\mu)$. Specifically, if $f \in C^1 \cap L^2(\mu)$ and $\varepsilon > 0$, we can assume that $\int_{\mathbb{R}^n} f d\mu = 0$, and then there exists $u \in C_c^\infty$ such that $\|f - L(u)\|_{L^2(\mu)}^2 \leq \varepsilon$, or equivalently

$$\int f^2 d\mu \leq \varepsilon + 2\int fL(u)d\mu - \int L(u)^2 d\mu \leq \varepsilon - 2\int \nabla f \cdot \nabla u d\mu - \int (D^2\psi)\nabla u \cdot \nabla u d\mu. \tag{15}$$

Now we will use a general linear algebra fact. If $A > 0$, it has a square root $\sqrt{A} > 0$. If x, y are vectors, then by the Cauchy-Schwarz inequality and the arithmetic-geometric inequality:

$$\begin{aligned}x \cdot y &= \sqrt{A}x \cdot \sqrt{A^{-1}}y \leq (\sqrt{A}x \cdot \sqrt{A}x)^{1/2}(\sqrt{A^{-1}}y \cdot \sqrt{A^{-1}}y)^{1/2} \\ &= (Ax \cdot x)^{1/2}(A^{-1}y \cdot y)^{1/2} \leq \frac{1}{2}Ax \cdot x + \frac{1}{2}Ay \cdot y,\end{aligned}$$

And by swapping sides, we find

$$-2x \cdot y - Ax \cdot x \leq Ay \cdot y.$$

Applying this to (15) we get

$$\int f^2 d\mu \leq \varepsilon + \int (D^2\psi)\nabla f \cdot \nabla f d\mu,$$

and letting $\varepsilon \to 0$, we get the statement.

To finish the proof, we need to prove the density argument we used. To prove it, it is enough to show that the only function orthogonal to $L(u)$ for each u is 0. Let f be such a function, then

$$\int L(f) u d\mu = \int fL(u) d\mu = 0,$$

so $L(f) = 0$. The equation $L(f) = 0$ defines an elliptic partial differential equation, and f is a weak solution to it. By the regularity of weak solutions to elliptic equations, this implies that f is C^∞.

A direct computation shows

$$\begin{aligned}L(f^2) &= \Delta(f^2) - \nabla\psi \cdot \nabla(f^2) = 2|\nabla f|^2 + 2f \cdot \Delta f - 2f\nabla f \cdot \nabla\psi \\ &= 2|\nabla f|^2 + 2fL(f) = 2|\nabla f|^2,\end{aligned}$$

so if $\theta \in C_c^\infty$, then

$$\begin{aligned}\int |\nabla(\theta f)|^2 d\mu &= \int |f\nabla\theta + \theta\nabla f|^2 d\mu = \int (f^2|\nabla\theta|^2 + 2f\theta\nabla\theta\cdot\nabla f + \theta^2|\nabla f|^2)d\mu \\ &= \int \left(f^2|\nabla\theta|^2 + \frac{1}{2}\nabla(f^2)\cdot\nabla(\theta^2) + \theta^2|\nabla f|^2\right) d\mu \\ &= \int \left(f^2|\nabla\theta|^2 + \frac{1}{2}L(f^2)\theta^2 + \theta^2|\nabla f|^2\right) d\mu = \int f^2|\nabla\theta^2| d\mu.\end{aligned}$$

Recall we want to prove that f is constant, i.e. $\nabla f = 0$. If we were able to choose $\theta = 1$, this would be obvious, however we need θ to be compactly supported. Therefore, let θ_k be a sequence of smooth functions such that $\theta_k \equiv 1$ on $B_k(0)$, and $|\nabla\theta_k| \le 1/k$. Then

$$\int_{B_k(0)} |\nabla f|^2 d\mu \le \int |\nabla(\theta_k f)|^2 d\mu = \int f^2|\nabla\theta_k^2| d\mu \le \frac{1}{k^2}\|f\|_{L^2(\mu)}^2.$$

When $k \to \infty$, the right hand side tends to zero. This means that $|\nabla f|$ is zero almost everywhere, so f is constant. $\square$

Another inequality we will use is the Prékopa-Leindler inequality.

Theorem 15 (Prékopa-Leindler Inequality)**.** *Let μ be a log-concave probability measure over $\mathbb{R}^n$. Let $p \ge 1, 0 < t < 1, f, g, h : \mathbb{R}^n \to \mathbb{R}$ be such that, for all $x, y \in \mathbb{R}^n$,*

$$h(tx + (1-t)y) \ge f(x)^t g(x)^{1-t}.$$

Then

$$\|h\|_{L^1(\mu)} \ge \|f\|_{L^1(\mu)}^t \|g\|_{L^1(\mu)}^{1-t}.$$

The Prékopa-Leindler inequality can be derived from another transport related inequality, which we will now describe.

Proposition 14. *Let $p > 1$, μ_0, μ_1 measures, and T the optimal transport map from μ_0 to μ_1 for the p-cost. Let $T_t = tT + (1-t)\mathrm{Id})$, and $\mu_t = T_t\#\mu_0$. Then*

$$\mu_t(T_t(x)) \le \mu_0(x)^{1-t}\mu_1(T(x))^t. \tag{16}$$

The proof relies on a formula on the push-forward of a measure. Specifically, if $T\#\mu = \nu$, T is approximately differentiable almost everywhere and injective, and f, g are the densities of μ, ν respectively, then

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

This formula is essentially the change of variables formula: Specifically, the change of variables formula implies the integrals of the two sides on any measurable subset $A \subseteq \mathbb{R}^n$ coincide, since

$$\int_A g(T(x))|\det\tilde{\nabla}T(x)| = \int_{T(A)} g(x)dx = \nu(T(A)) = \mu(A).$$

Now, we prove the proposition.

Proof of proposition 14. By theorem 9, $\tilde{\nabla}T_t$ is diagonalizable with nonnegative eigenvalues, so $\det\tilde{\nabla}T_t$ is nonnegative. Also, by lemma 3, T_t is injective almost everywhere, since μ_t is absolutely continuous. Writing the push-forward equation in this case, we find that

$$\mu_0(x) = \mu_t(T_t(x))\det\tilde{\nabla}T_t(x) = \mu_t(T_t(x))\det(t\tilde{\nabla}T + (1-t)I). \tag{17}$$

Next, we use the fact that $\det(tA + (1-t)I)^{1/n}$ is concave in t, when A is diagonalizable and with nonnegative eigenvalues. This first gives us

$$\det(t\tilde{\nabla}T + (1-t)I) \geq \left(t\det\tilde{\nabla}T^{1/n} + (1-t)\det I^{1/n}\right)^n = \left(t\left(\frac{\mu_0(x)}{\mu_1(x)}\right)^{1/n} + (1-t)\right)^n.$$

Plugging into (17), we conclude

$$\begin{aligned}\mu_t(T_t(x)) &\leq \frac{\mu_0(x)}{\left(t\left(\frac{\mu_0(x)}{\mu_1(x)}\right)^{1/n} + (1-t)\right)^n} \\ &= \left(t\mu_1(x)^{-1/n} + (1-t)\mu_0(x)^{-1/n}\right)^{-n} \leq \mu_0(x)^{1-t}\mu_1(T(x))^t,\end{aligned}$$

where the second inequality is just the arithmetic-geometric inequality. □

Using this result, we can give a simple proof of the Prékopa-Leindler inequality.

Proof of the Prékopa-Leindler inequality. By normalizing, we can assume $\|f\|_{L^1(\mu)} = \|g\|_{L^1(\mu)} = 1$, so it is enough to prove that $\|h\|_{L^p(\mu)} \geq 1$. Define $\mu_0 = fd\mu, \mu_1 = gd\mu$. If T is the transport map between μ_0 and μ_1, the proposition and log-concavity of μ give us

$$\begin{aligned}\mu_t(T_t(x)) &\leq \mu_0(x)^{1-t}\mu_1(T(x))^t = f(x)^{1-t}\mu(x)^{1-t}\mu(T(x))^t g(T(x))^t \leq f(x)^{1-t}g(T(x))^t\mu(T_t(x)) \\ &\leq h(T_t(x))\mu(T_t(x))\end{aligned}$$

Therefore

$$\mu_t(x) \leq h(x)\mu(x).$$

Integrating, we get

$$1 = \int 1 d\mu_t \leq \int h(x)d\mu = \|h\|_{L^1(\mu)},$$

which is what we wanted to prove. □

When dealing with the $p = 1$ case, we will consider the finiteness of integrals of the form

$$\int_X d(x, \partial X)^{-\alpha}dx, \tag{18}$$

for $0 \leq \alpha < 1$. It turns out that such integrals converge when X has rectifiable boundary. We recall the definition of rectifiability in this case:

Definition 11. *Let $X \subseteq \mathbb{R}^n$ be bounded. The boundary of X is said to be rectifiable if there exists a bounded subset $C \subseteq \mathbb{R}^{n-1}$ and a Lipschitz function $f : \mathbb{R}^{n-1} \to \mathbb{R}^n$ such that $\partial X \subseteq f(C)$.*

Proposition 15. *Let $X \subseteq \mathbb{R}^n$ be a bounded set with rectifiable boundary, and let $0 \leq \alpha < 1$. Then the integral (18) is finite.*

For the proof, we will use the concept of the Minkowski content. We denote $X_t = \{x \in X : d(x, \partial X \geq t)\}$, so $\partial X_t = \{x \in X : d(x, \partial X) = t)$.

Definition 12. *Let $Y \subseteq \mathbb{R}^{n-1}$. We define the upper $(n-1)$-dimensional Minkowski content of Y by*

$$M^{n-1}(Y) = \limsup_{\varepsilon \to 0} \frac{|\{x \in \mathbb{R}^n : d(x, Y) \leq \varepsilon\}|}{\varepsilon}.$$

So, in case that $Y = \partial X$, we have

$$M^{n-1}(\partial X) = \limsup_{\varepsilon \to 0} \frac{|X \backslash X_\varepsilon|}{\varepsilon}.$$

Theorem 16. *Let X be a set with rectifiable boundary. Then*

$$M^{n-1}(\partial X) = \mathcal{H}^{n-1}(\partial X),$$

where M^{n-1} is the $n-1$-dimensional Minkowski content. In particular, the Minkowski dimension of ∂X is $n-1$.

For the proof, see [6, theorem 3.2.39]. In fact, we will only need a much weaker theorem: All we need is for the Minkowski content $M^{n-1}(\partial X)$ to not be infinite, which is much easier to prove.

Proof. By definition, there exists a bounded subset $A \subseteq \mathbb{R}^{n-1}$ and a surjective L-Lipschitz map $f : A \to \partial X$. For a given $\varepsilon > 0$, we can cover A by C/ε^{n-1} balls of radius ε, for some C depending on A. Now if $|x - y| \leq \varepsilon$ for some $y \in \partial X$, then we can write $y = f(z)$, and z lies in one of those ε balls, say $B_\varepsilon(z')$. Then

$$|x - f(z')| \leq |x - f(z)| + |f(z) - f(z')| \leq \varepsilon + L\varepsilon = (L + 1)\varepsilon.$$

Denoting $C' = L + 1$, we get that $X \backslash X_\varepsilon$ is covered by $C\varepsilon^{n-1}$ balls of radius $C'\varepsilon$. Since the volume of a ball in $\mathbb{R}^n$ is proportional to the n-th power of the radius, we get the total volume of $(\partial_X)_\varepsilon$ is bounded by

$$C''\varepsilon^n/\varepsilon^{n-1} = C''\varepsilon,$$

which shows that ∂X has Minkowski dimension at most $n - 1$. □

Proof of proposition 15. We know that $M^{n-1}(\partial X) < \infty$, so there exists a constant $C < \infty$ such that $|X \backslash X_\varepsilon| \leq C\varepsilon + \beta(\varepsilon)$, where $\beta(\varepsilon) = o(\varepsilon)$. Therefore,

$$\begin{aligned}
\int_X d(x, \partial X)^{-\alpha} dx &= \int_0^\infty |\{x \in X : d(x, \partial X)^{-\alpha} \geq s\}| ds \\
&\leq \int_0^1 |X| ds + \int_1^\infty |X \backslash X_{(1/s)^{1/\alpha}}| ds \\
&\leq |X| + C \int_1^\infty \frac{1}{s^{1/\alpha}} ds + \int_1^\infty \beta\left(\frac{1}{s^{1/\alpha}}\right) ds \\
&< \infty.
\end{aligned}$$

□

For C^2 domains, we can actually derive a better result, determining the speed of convergence to the limit defining the Minkowski content. We first note the following:

Proposition 16. *Let X be a C^2 domain. Then*

- *$(X_t)_s = X_{t+s}$ for all sufficiently small t, s,*
- *the quantity $\mathcal{H}^{n-1}(\partial X_t)$ stays bounded as $t \to 0$.*

We will use the theory of the geometry of tubes. We first prove the following.

Proposition 17. *Let M be a C^1 manifold, $y \in \mathbb{R}^n$, and let $x \in M$ be the closest point in M to y. Then $x - y$ is normal to M at the point x.*

Proof. Write M implicitly as $\{x : f(x) = 0\}$, for $f \in C^1$. Then the closest point to y is the minimizer of $|x-y|^2$, subject to $f(x) = 0$. The lagrange multiplier is

$$\mathcal{L}(x, \lambda) = |x-y|^2 + \lambda f(x)$$

Then we have, for the minimizer

$$2(x-y) + \lambda \nabla f(x) = 0$$

So, $x - y$ is parallel to $\nabla f(x)$, which is perpendicular to the manifold. □

Now, let $M \subseteq \mathbb{R}^n$ be a closed, oriented, $n-1$-dimensional, C^2 submanifold, and denote by $\nu(m)$ the unit normal vector to $m \in M$. Define a map $F : M \times \mathbb{R} \to \mathbb{R}^n$ by $F(m,t) = m + t\nu(m)$. Note that F is linear in t, and $\frac{d}{dt}F = \nu$.

Proposition 18. *$F(-,t)$ is a diffeomorphism for small enough $t > 0$.*

Proof. Because M is C^2, the map F is continuously differentiable. At each point $m \in M$, we can express the tangent space $(\mathbb{R}_n)_m$ to m in $\mathbb{R}^n$ as a direct sum $(\mathbb{R}_n)_m = di(M_m) \oplus \mathbb{R}\nu(m)$, where i is the inclusion map $i : M \to \mathbb{R}^n$.

Note that $F|_{M\times\{0\}}$ is just i, so $dF(m,0)|_{M_m}$ is equal to di. Also, $F|_{\{m\}\times\mathbb{R}}$ is linear in t, and the image $dF(m,0)|_{\mathbb{R}_0}$ is $\mathbb{R}\nu(m)$. This proves that dF is surjective, hence an isomorphism at each point $(m,0)$.

By the inverse function theorem, F is a local diffeomorphism in a neighborhood of $M \times \{0\}$. By compactness, there exists open sets $U_1, \dots, U_k$ such that $\bigcup_{i=1}^k U_i \subseteq M$, and such that $F|_{U_i}$ is a diffeomorphism for each i. Let d be the Lebesgue number of this covering; this means that $|m - m'| \geq d$ if m, m' do not belong to the same U_i for any i. Pick t_0 small enough so that $F(M \times [0, t_0] \subseteq \bigcup_{i=1}^k U_i$, and $2t_0 \leq d$. Then, if $m \neq m'$ do not belong to the same U_i, and $t, t' \leq t_0$, we get that

$$|F(m,t) - F(m',t')| \geq |m - m'| - |m - F(m,t)| - |m' - F(m',t')| \geq d - t - t' \geq d - 2t_0 > 0,$$

which shows that $F(m,t) \neq F(m',t')$. If they do belong to the same open set, then $F(m,t) \neq F(m',t')$ since $F|_{U_i}$ is a diffeomorphism for each i. This proves that $F|_{\bigcup_{i=1}^k U_i}$ is injective, so it is a diffeomorphism. □

Next, we introduce an object called the second fundamental form. The manifold M inherits a Riemannian metric from $\mathbb{R}^n$. Explicitly, if $\partial_1(m), \ldots, \partial_{n-1}(m)$ are local coordinates of M_m, g is equal to

$$g(\partial_i(m), \partial_j(m)) = \langle \partial_i F(m,0), \partial_j F(m,0) \rangle.$$

Define the second fundamental form at point $m \in M$ to be the unique A satisfying

$$g(A\partial_i, \partial_j) = -\langle \nu(m), \partial_{ij} F(m,0) \rangle.$$

Define M_t to be the image of $F(M,t)$ with g_t the induced Riemannnian metric from $\mathbb{R}^n$. By proposition 18, M_t is diffeomorphic to M for small enough t, so we can identify tangent vectors on M and M_t. Then

$$\partial_i \left\langle \frac{d}{dt} F(m,t), \partial_j F(m,t) \right\rangle = \left\langle \frac{d}{dt} \partial_i F(m,t), \partial_j F(m,t) \right\rangle + \left\langle \frac{d}{dt} \partial_i F(m,t), \partial_{ij} F(m,t) \right\rangle,$$

by the product rule. So:

$$\begin{aligned} \left\langle \frac{d}{dt} \partial_i F(m,t), \partial_j F(m,t) \right\rangle &= \partial_i \left\langle \frac{d}{dt} F(m,t), \partial_j F(m,t) \right\rangle - \left\langle \frac{d}{dt} F(m,t), \partial_{ij} F(m,t) \right\rangle \\ &= \partial_i \langle \nu(m), \partial_j F(m,t) \rangle - \langle \nu(m), \partial_{ij} F(m,t) \rangle \\ &= -\langle \nu(m), \partial_{ij} F(m,t) \rangle, \end{aligned} \tag{19}$$

where the last equality follows from the fact that ν is perpendicular to any tangent vector, so $\langle \nu(m), \partial_j F(m,t) \rangle = 0$. So, by the product rule again:

$$\begin{aligned} \frac{d}{dt} g_t(\partial_i, \partial_j) \Big|_{t=0} &= \frac{d}{dt} \langle \partial_i F(m,t), \partial_j F(m,t) \rangle \Big|_{t=0} \\ &= \left\langle \frac{d}{dt} \partial_i F(m,t), \partial_j F(m,t) \right\rangle \Big|_{t=0} + \left\langle \partial_i F(m,t), \frac{d}{dt} \partial_j F(m,t) \right\rangle \Big|_{t=0} \\ &= -2 \langle \nu(m), \partial_{ij} F(m,0) \rangle = 2g(A\partial_i, \partial_j). \end{aligned}$$

For t is small enough, we can define the second fundamental form also for the hypersurfaces M_t, and denote it by $A(t)$, so that $A = A(0)$. We want to prove the following:

Proposition 19. *There holds*

$$A'(t) = -A^2(t).$$

Proof. By (19), we have

$$\left\langle \frac{d}{dt} \partial_i F(m,t) \Big|_{t=t_0}, \partial_j F(m,t_0) \right\rangle = -\langle \nu(m), \partial_{ij} F(m,t_0) \rangle = g_{t_0}(A(t_0)\partial_i, \partial_j),$$

so

$$\left\langle \frac{d}{dt} \partial_i F(m,t) \Big|_{t=t_0}, \partial_j F(m,t_0) \right\rangle = \langle (A(t_0)\partial_i F(m,t_0)), \partial_j F(m,t_0) \rangle = \langle dF_m(A(t_0)\partial_i), \partial_j F(m,t_0) \rangle.$$

Because dF is an isomorphism, this implies that $dF_m(A(t_0))\partial_i) = \frac{d}{dt}\partial_i F(m,t)\big|_{t=t_0}$. Note that $A(t)$ is clearly self adjoint by definition, so we get

$$\begin{aligned} g_{t_0}(A^2(t_0)\partial_i, \partial_j) &= g_{t_0}(A(t_0)\partial_i, A(t_0)\partial_j) = \langle dF_m(A(t_0)\partial_i), dF_m(A(t_0)\partial_j)\rangle \\ &= \left\langle \frac{d}{dt}\partial_i F(m,t)\bigg|_{t=t_0}, \frac{d}{dt}\partial_j F(m,t)\bigg|_{t=t_0}\right\rangle. \end{aligned}$$

Since $\frac{d^2}{dt^2}F = \frac{d}{dt}\nu = 0$, we conclude

$$\begin{aligned} g_{t_0}(A^2(t_0)\partial_i, \partial_j) &= \left\langle \frac{d}{dt}\partial_i F(m,t)\bigg|_{t=t_0}, \frac{d}{dt}\partial_j F(m,t)\bigg|_{t=t_0}\right\rangle \\ &= \frac{d}{dt}\left\langle \partial_i F(m,t), \frac{d}{dt}\partial_j F(m,t)\right\rangle\bigg|_{t=t_0} - \left\langle \partial_i F(m,t_0), \frac{d^2}{dt^2}\partial_j F(m,t)\right\rangle\bigg|_{t=t_0} \\ &= \frac{d}{dt}\left\langle \partial_i F(m,t), \frac{d}{dt}\partial_j F(m,t)\right\rangle\bigg|_{t=t_0} \\ &= \frac{d}{dt} g_t(A(t)\partial_i, \partial_j)\bigg|_{t=t_0} \\ &= g'_{t_0}(A(t_0)\partial_i, \partial_j) + g_{t_0}(A'(t_0)\partial_i, \partial_j) \\ &= 2g_{t_0}(A^2(t_0)\partial_i, \partial_j) + g_{t_0}(A'(t_0)\partial_i, \partial_j), \end{aligned}$$

And by swapping sides, we get the result. □

As a consequence, consider $\frac{d}{dt}A^{-1}(t)$: It is equal to

$$\frac{d}{dt}A^{-1}(t) = -A^{-1}A'(t)A^{-1} = -A\cdot A^{-2}A'(t)A^{-1} = -A\cdot(A'(t))^{-1}A'(t)A^{-1} = -AA^{-1} = -I.$$

In other words, $A^{-1}(t)$ is a linear function. It is nonzero everywhere, so $A(t)$ is the inverse of a linear function.

As a consequence of this, $\mathrm{trace}(A(t))$ stays bounded as $t \to 0$. If $\omega_t = \sqrt{\det g_t}dx_1 \wedge \ldots \wedge dx_n$ is the volume form on M_t, then the derivative of the $(n-1)$-dimensional volume of M_t is given by

$$\begin{aligned} \frac{d}{dt}\mathrm{Vol}(M_t) &= \int \frac{d}{dt}\omega_t = \int \frac{\frac{d}{dt}\det g_t}{2\sqrt{\det g_t}}dx_1 \wedge \ldots \wedge x_n = \int \mathrm{trace}(A(t)) \cdot \sqrt{\det g_t}dx_1 \wedge \ldots \wedge x_n \\ &= \int \mathrm{trace}(A(t))\omega_t, \end{aligned}$$

since

$$\frac{d}{dt}\det g_t = \det g_t \cdot \mathrm{trace}(g_t^{-1}\cdot\frac{d}{dt}g_t) = \det g_t \cdot \mathrm{trace}(g_t^{-1}\cdot g_t \cdot A(t)) = \det g_t \cdot 2\mathrm{trace}A(t).$$

Since $\mathrm{trace}(A(t))$ is bounded and M is compact, we have that $\mathrm{Vol}(M_t)$ stays bounded as $t \to 0$, which is what we wanted to prove.

proof of proposition 16. Orient ∂X so ν is the inward pointing normal vector. Define the map $F : \partial X \times [0, t_0] \to \mathbb{R}^n$ as before, by $F(x,t) = x + t\nu(x)$ is a diffeomorphism for t_0 small enough.

We want to prove that $\partial X_t = F(\partial X, t)$. Let $y \in \partial X_t$, and let $x \in X$ be such that $|x - y| = t$. By proposition 17, $|x - y| = t\nu(x)$, so $y \in F(\partial X, t)$. On the other hand, if $y \in F(\partial X, t)$, then $y = x + t\nu(x)$ for some $x \in \partial X$, so $d(y, \partial X) \leq |x - y| = t$. If $t' := d(x, \partial X) < t$, then $y \in F(\partial X, t')$. However, F is a diffeomorphism, so y cannot belong to both $F(X,t)$ and $F(X,t')$, for $t' < t$. Therefore $d(y, \partial X)$, or in other words $y \in \partial X_t$.

Now, the first point follows easily, because

$$(\partial X_t)_s = F(F(\partial X, t), s) = F(\partial X, t + s) = \partial X_{t+s}.$$

The second part was already proven. $\square$

To evaluate (18), we use the coarea formula, shown below.

Proposition 20 (coarea Formula)**.** *Let u be a Lipschitz function, $f \in L^1(X)$. Then*

$$\int_X f(x) |\det u(x)| dx = \int_0^\infty \int_{u^{-1}(\{y\})} f(x) d\mathcal{H}^{n-1}(x) dy.$$

Proposition 21. *Let X be a bounded C^2 domain, $0 \leq \alpha < 1$. Define*

$$h(\alpha, t) = \int_{X_t} d(x, \partial X_t)^{-\alpha} dx.$$

Then $h(\alpha, t)$ stays bounded as $t \to 0$.

Proof. Let t_0 be small enough so that $(\partial X_t)_s = \partial X_{t+s}$ for every $t, s \leq t_0$. We use the coarea formula, with $u_t = d(x, \partial X_t)$. Then $|\det \nabla u| = 1$ if $d(x, \partial X) \leq t_0$. The level sets of u are $u^{-1}(\{s\}) = (\partial X_t)_s$, which is 1-Lipschitz. Calculating, we get

$$\int_{X_t} d(x, \partial X_t)^{-\alpha} dx \leq \int_{X_t \cap \{d(x, \partial X_t) > t_0\}} t_0^{-\alpha} dx + \int_{X_t \cap \{d(x, \partial X_t) \leq t_0\}} d(x, \partial X_t)^{-\alpha} dx.$$

The first integral is bounded since X, and hence X_t, is bounded. For the second integral, we calculate:

$$\begin{aligned}
\int_{X_t \cap \{d(x, \partial X_t) \leq t_0\}} d(x, \partial X_t)^{-\alpha} dx &= \int_0^{t_0} \int_{u^{-1}(s)} s^{-\alpha} d\mathcal{H}^{n-1}(x) ds \\
&= \int_0^{t_0} s^{-\alpha} \mathcal{H}^{n-1}((\partial X_t)_s) ds \\
&= \int_0^{t_0} s^{-\alpha} \mathcal{H}^{n-1}(X_{t+s}) ds,
\end{aligned}$$

and since $\alpha < 1$ and $\mathcal{H}^{n-1}(X_{t+s})$ is bounded as $t, s \to 0$, the last integral stays bounded as $t \to 0$. $\square$

3 Stability of Optimal Transport - The Case $p > 1$

In this section, we prove quantitative stability of optimal transport maps for $p > 1$. This section summarizes two articles: The first, by Delalande and Mérigot [5], deals with $p = 2$, and the second, by Mischler and Trevisan [8], generalizes Delalande and Mérigot's results to $p > 1$.

The main theorems we prove are the following.

Theorem 17. *Suppose μ is a probability density supported on a bounded convex set Ω, with $\infty > M \geq \mu \geq m > 0$ on Ω, and ν_1, ν_2 are contained in a bounded set X. Let T_1, T_2 be the Brenier maps from μ to ν_1, ν_2. Then*

$$\|T_1 - T_2\|_{L^2(\mu)} \leq C_{\Omega,X,\mu} W_1(\nu_1, \nu_2)^{1/6}.$$

Theorem 18. *Suppose μ is a log-concave probability measure supported on a bounded convex set Ω, and ν_1, ν_2 are contained in a bounded set X. Let T_1, T_2 be the optimal transport maps for the p-cost from μ to ν_1, ν_2. Then if $p > 2$, we have*

$$\|T_1 - T_2\|_{L^2(\mu)} \leq C_{\Omega,X,\mu} W_1(\nu_1, \nu_2)^{1/6(p-1)},$$

and if $1 < p < 2$, we have

$$\|T_1 - T_2\|_{L^2(\mu)} \leq C_{\Omega,X,\mu,\alpha} W_1(\nu_1, \nu_2)^{\alpha},$$

For every $0 < \alpha < \frac{(p-1)^2}{p(p-1)}$.

We start with the proof of the first theorem in the next subsection.

3.1 Stability of Potentials for $p = 2$

We first prove the following theorem.

Theorem 19. *Let μ be a probability density on a compact and convex set $\Omega \subseteq \mathbb{R}^n$, with $0 < m_\mu \leq \rho \leq M_\mu < \infty$. Let ν_0, ν_1 be two probability measures with support bounded by R, and let φ_0, φ_1 be the associated potentials transporting from μ to ν_0, ν_1 for the square cost. Suppose that there exist constants $m_\varphi, M_\varphi \in \mathbb{R}$ such that $m_\varphi \leq \varphi_0, \varphi_1 \leq M_\varphi$. Then denoting $\psi_0 = \varphi_0^*, \psi_1 = \varphi_1^*$, the following inequality holds:*

$$\mathrm{Var}_\mu(\varphi_1 - \varphi_0) \leq eR\mathrm{diam}(\Omega)\frac{M_\mu}{m_\mu}\int_{\mathbb{R}^n}(\psi_0 - \psi_1)d(\nu_1 - \nu_0). \tag{20}$$

The proof will consist of applying the Brascamp-Lieb inequality to an expression involving the Kantorovich functional, which is defined as $K(\psi) = \mathbb{E}_\mu[\psi^*]$. We first prove the following proposition.

Proposition 22. *Let $\varphi_0, \varphi_1 \in C^2(\mathbb{R}^n)$ be strongly convex functions. Define $\psi_0 = \varphi_0^*, \psi_1 = \varphi_1^*$. Now define $v = \psi_1 - \psi_0, \psi_t = \psi_0 + tv = t\psi_1 + (1-t)\psi_0$, and finally $\varphi_t = \psi_t^*$. Then φ_t are also in C^2, are strongly convex, and we have the following identities:*

$$\frac{d}{dt}K(\psi_t) = -\int_\Omega v(\nabla\varphi_t)d\mu, \tag{21}$$

$$\frac{d^2}{dt^2}K(\psi_t) = \int_\Omega \nabla v\,(\nabla\varphi_t)\cdot D^2\varphi_t(\nabla v(\nabla\varphi_t))d\mu. \tag{22}$$

Proof. We assume that φ_0, φ_1 are α-strongly convex and C^2. The inverse function theorem implies that ψ_0, ψ_1 are also C^2 with α-Lipschitz gradients, and $D^2\psi_i > 0$ for $i = 0, 1$. These properties carry over to ψ_t, and by the inverse function theorem again we have that φ_t is C^2 and α-strongly convex.

Define $F(t, x, y) = \nabla\psi_t(y) - x$, and consider the function $G(t, x) = \nabla\varphi_t(x)$; then G satisfies $F(t, x, G(t, x)) = 0$. $\partial_y F = D^2\psi_t(y)$ is positive definite, so the implicit function theorem applies and shows that G is a C^1 map. Moreover, by implicit differentiation, we get

$$\left(\frac{d}{dt}\nabla\psi_t\right)(\nabla\varphi_t(x)) + D^2\psi_t(\nabla\varphi_t(x)) \cdot \frac{d}{dt}\nabla\varphi_t(x) = 0,$$

and after rearranging:

$$\frac{d}{dt}\varphi_t(x) = -D^2\varphi_t(x) \cdot \nabla v(\nabla\varphi_t(x)). \tag{23}$$

Now, we turn toward computing the derivatives of φ_t with respect to time. Using the Fenchell-Young equality, we have

$$\varphi_t(x) = x \cdot \nabla\varphi_t(x) - \psi_t(\nabla\varphi_t(x)).$$

This equation shows that φ_t is differentiable, and we can find its derivative as

$$\frac{d}{dt}\varphi_t(x) = x \cdot \frac{d}{dt}\nabla\varphi_t(x) - v(\nabla\varphi_t(x)) - \nabla\psi_t(\nabla\varphi_t(x)) \cdot \frac{d}{dt}\nabla\varphi_t(x) = -v(\nabla\varphi_t(x)).$$

And now $\frac{d}{dt}\varphi_t$ can be differentiated again, and using (23) we can find the second derivative of φ_t to be

$$\frac{d^2}{dt^2}\varphi_t(x) = -\nabla v(\nabla\varphi_t(x)) \cdot \frac{d}{dt}\nabla\varphi_t(x) = \nabla v\left(\nabla\varphi_t(x)\right) \cdot D^2\varphi_t(x)\nabla v\left(\nabla\varphi_t(x)\right).$$

The proposition now follows from differentiation under the integral sign. □

Proof of Theorem 19. First, we can by standard approximation reduce to C^2, strongly convex functions. Using the fundamental theorem of calculus and equations (21), (22), we find

$$\begin{aligned}\int_{\mathbb{R}^n}(\psi_0 - \psi_1)d(\nu_1 - \nu_0) &= \frac{d}{dt}K(\psi_t)\big|_{t=1} - \frac{d}{dt}K(\psi_t)\big|_{t=0} = \int_0^1 \frac{d^2}{dt^2}K(\psi_t)dt \\ &= \int_0^1 \mathbb{E}_\mu\left[\nabla v(\nabla\varphi_t) \cdot D^2\varphi_t\nabla v(\nabla\varphi_t)\right]dt.\end{aligned}$$

To bring this equation to a form that is closer to the Brascamp-Lieb inequality, introduce the change of variables $\overline{v}_t := v(\nabla\varphi_t)$, so the expression inside the integral becomes

$$\frac{d^2}{dt^2}K(\psi_t) = \mathbb{E}_\mu[(D^2\varphi_t)^{-1}\nabla\overline{v}_t \cdot \nabla\overline{v}_t] \tag{24}$$

We want to apply the Brascamp-Lieb inequality, however note that we have not assumed ρ is log-concave. Therefore, we define a new measure $\overline{\mu}_t$ with density $\frac{1}{Z_t}e^{-\varphi_t}$, with Z_t chosen so that this results in a probability distribution. Then the Brascamp-Lieb inequality gives

$$\mathbb{E}_{\overline{\mu}_t}\left[(D^2\varphi_t)^{-1}\nabla\overline{v}_t \cdot \nabla\overline{v}_t\right] \geq \mathrm{Var}_{\overline{\mu}_t}(\overline{v}_t). \tag{25}$$

In order to relate this inequality to (24), we will show that the ratio of $\overline{\mu}_t$ and μ is bounded from above and below. First, we will show that $m_\varphi \leq \varphi_t \leq M_\varphi$. This holds for $t = 0, 1$ by assumption, and by convexity of the conjugation operator, we see that

$$\varphi_t = (t\psi_1 + (1-t)\psi_0)^* \leq t\varphi_1 + (1-t)\varphi_0 \leq M_\varphi.$$

On the other hand, we have $\psi_0(0) = \sup\{-\varphi_0\} \leq -m_\varphi$ and similarly for $\psi_1(0)$, so

$$\varphi_t(x) = \sup\{x \cdot y - \psi_t(y)\} \geq -\psi_t(0) = -t\psi_1(0) - (1-t)\psi_0 \geq m_\varphi.$$

Therefore, we have that $m_\varphi \leq \varphi_t \leq M_\varphi$ for all $t \in [0,1]$, so $\frac{e^{-M_\varphi}}{Z_t} \leq \overline{\mu}_t \leq \frac{e^{-m_\varphi}}{Z_t}$, and since we know that $m_\mu \leq \mu \leq M_\mu$, we get

$$Z_t e^{m_\varphi} m_\mu \overline{\mu}_t \leq \mu \leq Z_t e^{M_\varphi} M_\mu \overline{\mu}_t. \tag{26}$$

Now, the value inside the expectation in (24) is nonnegative, since $D^2\varphi_t$ is positive definite. Therefore, using (25) and (26) gives us

$$\begin{aligned}\mathbb{E}_\mu[(D^2\varphi_t)^{-1}\nabla\overline{v}_t \cdot \nabla\overline{v}_t] &\geq Z_t e^{m_\varphi} m_\rho \mathbb{E}_{\overline{\mu}_t}[(D^2\varphi_t)^{-1}\nabla\overline{v}_t \cdot \nabla\overline{v}_t] \\ &\geq Z_t e^{m_\varphi} m_\rho \mathrm{Var}_{\overline{\mu}_t}(\overline{v}_t) \geq e^{m_\varphi - M_\varphi}\frac{m_\rho}{M_\rho}\mathrm{Var}_\mu(\overline{v}_t).\end{aligned} \tag{27}$$

The last inequality follows from a general fact: If $\rho_1 \leq C\rho_2$, then

$$\mathrm{Var}_{\rho_1}(f) = \min_{c\in\mathbb{R}} \mathbb{E}_{\rho_1}[(f-c)^2] \leq \min_{c\in\mathbb{R}} C\mathbb{E}_{\rho_2}[(f-c)^2] = C\mathrm{Var}_{\rho_2}(f).$$

Returning to (27), recalling the definition of $\overline{v}_t$, we get

$$\mathrm{Var}_\mu(\overline{v}_t) = \mathrm{Var}_\mu(\psi_1(\nabla\varphi_t) - \psi_0(\nabla\varphi_t)) = \mathrm{Var}_{\nu_t}(\psi_1 - \psi_0),$$

where $\nu_t = \nabla\varphi_t \# \mu$ interpolates between μ_0 and μ_1.

To summarize, so far we have shown that

$$\int_0^1 \mathrm{Var}_{\nu_t}(\psi_1 - \psi_0)dt \leq \frac{M_\mu}{m_\mu} e^{M_\varphi - m_\varphi} \int_X (\psi_0 - \psi_1)d(\nu_1 - \nu_0). \tag{28}$$

Now, by convexity of the Variance and Jensen's inequality,

$$\int_0^1 \mathrm{Var}_{\nu_t}(\psi_1 - \psi_0)dt = \int_0^1 \mathrm{Var}_\mu((\psi_1 - \psi_0) \circ \nabla\varphi_t)dt \geq \mathrm{Var}_\mu\left(-\int_0^1 \frac{d}{dt}\varphi_t(x)\right) = \mathrm{Var}_\mu(\varphi_1 - \varphi_0).$$

Finally, we want to remove the exponential dependence on $M_\varphi - m_\varphi$. Given $\alpha \in \mathbb{R}$, denote $\varphi_i^\alpha = \alpha\varphi_i, \nu_i^\alpha = \nabla\varphi_i^\alpha \# \mu$ for $i = 0, 1$. Denoting $\psi_i^\alpha = (\varphi_i^\alpha)^* = \alpha\psi_i(\cdot/\alpha)$, we have the following relations:

$$\begin{aligned}&\mathrm{Var}_\mu(\varphi_1^\alpha - \varphi_0^\alpha) = \alpha^2 \mathrm{Var}_\mu(\varphi_1 - \varphi_0), \\ &\int_{\mathbb{R}^n} (\psi_0^\alpha - \psi_1^\alpha)d(\nu_1^\alpha - \nu_0^\alpha) = \alpha \int_{\mathbb{R}^n} (\psi_0 - \psi_1)d(\nu_1 - \nu_0), \\ &\alpha m_\varphi \leq \varphi_0^\alpha, \varphi_1^\alpha \leq \alpha M_\varphi,\end{aligned}$$

which give us the following inequality:

$$\mathrm{Var}_\mu(\varphi_1 - \varphi_0) \leq \frac{M_\mu}{m_\mu} \frac{e^{\alpha(M_\varphi - m_\varphi)}}{\alpha} \int_{\mathbb{R}^n} (\psi_0 - \psi_1)d(\nu_1 - \nu_0).$$

If we set $\alpha = (M_\varphi - m_\varphi)^{-1}$, we are left with

$$\mathrm{Var}_\mu(\varphi_1 - \varphi_0) \leq e\frac{M_\mu}{m_\mu}(M_\varphi - m_\varphi)\int_{\mathbb{R}^n}(\psi_0 - \psi_1)d(\nu_1 - \nu_0).$$

Since $\nabla\varphi_0, \nabla\varphi_1$ are bounded by R, the maximum difference $M_\varphi - m_\varphi$ can take is $R\mathrm{diam}(\Omega)$, which proves the theorem. □

Corollary 3. *Assume the same assumptions as in Theorem 19. Then*

$$\mathrm{Var}_\mu(\varphi_1 - \varphi_0) \leq eR\mathrm{diam}(\Omega)^2\frac{M_\mu}{m_\mu}W_1(\mu_0, \mu_1).$$

Proof. The functions ψ_0, ψ_1 are diam(Ω)-Lipschitz since their the image of their gradients is in X, so this corollary follows by Kantorovich duality for $p = 1$. □

3.2 Stability of Transport Maps for $p = 2$

In the previous section, we proved a stability result relating the potentials. In this part, we want to extend this to the maps themselves. Our main tool is the following Gagliardo-Nirenberg type inequality, which allows us to bound the L^2 norm of the transport maps using the potentials:

Theorem 20. *Let $K \subseteq \mathbb{R}^n$ be a compact domain with rectifiable boundary. Let u, v be L-Lipschitz functions which are convex on intervals. Then there exists a constant C_n depending only on n such that the following inequality holds:*

$$\|\nabla u - \nabla v\|_{L^2}^2 \leq C_n L^{4/3}\mathcal{H}^{n-1}(\partial K)^{2/3}\|u - v\|_{L^2}^{2/3}.$$

Now theorem 17 follows immediately from theorem 20 and corollary 3. We now prove Theorem 20. We will first prove the theorem for the $n = 1$ case.

Proposition 23. *Suppose that $u, v : I \to \mathbb{R}$ are convex with unformly bounded gradients. Then*

$$\|u' - v'\|_{L^2}^2 \leq 8(\|u'\|_{L^\infty} + \|v'\|_{L^\infty})^{4/3}\|u - v\|_{L^2}^{2/3}.$$

Proof. By approximating u, v with smooth functions, we can assume that u, v are both C^2. Also, by a change of variables we can assume $I = [0, 1]$. Now, using integration by parts, we get

$$\|u' - v'\|_{L^2}^2 = \int_0^1 (u' - v')^2 dt = (u - v)(u' - v')\big|_0^1 - \int_0^1 (u - v)(u'' - v'')dt.$$

The first term can be bounded as

$$|(u - v)(u' - v')(1) - (u - v)(u' - v')(0)| \leq 2(\|u'\|_{L^\infty} + \|v'\|_{L^\infty})\|u - v\|_{L^\infty},$$

and using convexity, the second term can be bounded by

$$\begin{aligned}\left|\int_0^1 (u - v)(u'' - v'')dt\right| &\leq \|u - v\|_{L^\infty}\int_0^1 |u'' - v''|dt \\ &\leq \|u - v\|_{L^\infty}\int_0^1 u'' + v''dt \\ &= (u' + v')\big|_0^1\|u - v\|_{L^\infty} \\ &\leq 2(\|u'\|_\infty + \|v'\|_\infty)\|u - v\|_{L^\infty}.\end{aligned}$$

Here convexity ensures that u'', v'' are positive. Therefore, we get:

$$\|u' - v'\|_{L^2}^2 \le 4(\|u'\|_{L^\infty} + \|v'\|_{L^\infty})\|u - v\|_{L^\infty}. \tag{29}$$

Denote $\varepsilon = \|u - v\|_{L^\infty}$. Since $u - v$ is continuous, this value is achieved on some point $x \in [0, 1]$, and we can assume without loss of generality that $\varepsilon = u(x) - v(x)$. $u - v$ is actually L-Lipschitz continuous, for $L \le \|u'\|_{L^\infty} + \|v'\|_{L^\infty}$, so we have $u(x') - v(x') \ge \varepsilon/2$ for $x' \in [x - a, x + a]$, with $a = \min\{1/2, \varepsilon/2L\}$. Therefore

$$\|u - v\|_{L^2}^2 \ge \frac{\varepsilon^2}{4} \cdot \min\left\{1, \frac{\varepsilon}{2L}\right\}. \tag{30}$$

If $\varepsilon \le 2L$, then equation (30) becomes

$$\|u - v\|_{L^2}^2 \ge \frac{\varepsilon^3}{4L} \ge \frac{\|u - v\|_{L^\infty}^3}{8(\|u'\|_{L^\infty} + \|v'\|_{L^\infty})}.$$

Combining this with (29), we get after simplification

$$\|u' - v'\|_{L^2}^2 \le 8(\|u'\|_{L^\infty} + \|v'\|_{L^\infty})^{4/3}\|u - v\|_{L^2}^{2/3}.$$

If $\varepsilon \ge 2L$, on the other hand, then $\|u - v\|_2 \ge \varepsilon/2$, so

$$L^2 = L^{4/3} \cdot L^{2/3} \le (\|u'\|_{L^\infty} + \|v'\|_{L^\infty})^{4/3} \cdot \|u - v\|_{L^2}^{2/3},$$

and now we finish by noting that $L \ge \|u' - v'\|_{L^2}$. □

The generalization to higher dimensions will involve two results from integral geometry. First, by substitution, we have the following two facts, for any vector $e \in \mathbb{S}^{n-1}$:

$$\int_{\mathbb{R}^n} f(x)^2 dx = \int_{\{e\}^\perp} \int_{-\infty}^{\infty} f(y + te)^2 dt dy,$$
$$\int_{\mathbb{R}^n} \langle F(x), e\rangle^2 dx = \int_{\{e\}^\perp} \int_{-\infty\infty} \langle F(y + te), e\rangle^2 dt dy.$$

Integrating over all e, we get:

$$\begin{aligned}\int_{\mathbb{R}^n} f(x)^2 &= \int_{\mathbb{S}^{n-1}} \int_{\{e\}^\perp} \int_{-\infty}^{\infty} f(y + te) dt dy de,\\ \int_{\mathbb{R}^n} |F(x)|^2 dx &= C_n \int_{\mathbb{S}^{n-1}} \int_{\{e\}^\perp} \int_{-\infty\infty} \langle F(y + te), e\rangle^2 dt dy de,\end{aligned} \tag{31}$$

for some constant C_n. The second part is a result of Fubini's theorem and the the identity

$$C_n \int_{\mathbb{S}^{n-1}} \langle V, e\rangle^2 de = |V|^2, \tag{32}$$

which can be easily seen from the fact that (32) is invariant under scaling and rotations.

The second fact we will need is Crofton's formula: Denote l_e^y to be the oriented line $y + \mathbb{R}e$. Then if S is a rectifiable subset of $\mathbb{R}^n$,

$$H^{n-1}(S) = C_n' \int_{\mathbb{S}^{n-1}} \int_{\{e\}^\perp} \#(l_e^y \cap S) dy de.$$

Proof of theorem 20. Set $F = \nabla u - \nabla v$. Then (31) implies

$$\|\nabla u - \nabla v\|_{L^2}^2 = C_n \int_{\mathbb{S}^{n-1}} \int_{\{e\}^\perp} \|u'_{l_e^y} - v'_{l_e^y}\|^2_{L^2(l_e^y \cap K)} dyde,$$

where u_l denotes the restriction of u to l. For $e \in \mathbb{S}^{n-1}, y \in \{e\}^\perp$, let $n_{l_e^y}$ denote the number of connected components in $l_e^y \cap K$. Clearly $n_{l_e^y} \leq \#\{l_e^y \cap \partial K\}$. Since we assumed ∂K is rectifiable, Crofton's formula applies and we have

$$\int_{\mathbb{S}^{n-1}} \int_{\{e\}^\perp} n_{l_e^y} dyde \leq \int_{\mathbb{S}^{n-1}} \int_{\{e\}^\perp} \#(l_e^y \cap \partial K) dyde = \frac{1}{C'_n} H^{n-1}(\partial K) < \infty.$$

Therefore, for almost all lines l_e^y, $l_e^y \cap K$ is a finite union of $n_{l_e^y}$ segments. On each such segment, with the help of proposition 23 and Jensen's inequality, we get

$$\|u'_{l_e^y} - v'_{l_e^y}\|^2_{L^2(l_e^y \cap K)} \leq 8(2L)^{4/3} \left(n_{l_e^y}\right)^{2/3} \|u_{l_e^y} - v_{l_e^y}\|^{2/3}_{L^2(l_e^y \cap K)}.$$

From here, we see

$$\|\nabla u - \nabla v\|_{L^2}^2 \leq (2L)^{4/3} C_n \int_{\mathbb{S}^{n-1}} \int_{\{e\}^\perp} \left(n_{l_e^y}\right)^{2/3} \|u_{l_e^y} - v_{l_e^y}\|^{2/3}_{L^2(l_e^y \cap K)} dyde.$$

Using Hölder's inequality with $p = 3/2, q = 3$ and (31), we get that the last part is bounded by

$$C_n L^{4/3} \left(\int_{\mathbb{S}^{n-1}} \int_{\{e\}^\perp} n_{l_e^y} dyde \right)^{2/3} \cdot \|u - v\|_{L^2}^{2/3}. \tag{33}$$

Finally, we use Crofton's formula again to conclude that the integral in (33) is bounded by $C_{n'} \mathcal{H}^{n-1}(\partial K)^{2/3}$ for some constant $C_{n'}$. Putting all this together, we get that

$$\|\nabla u - \nabla v\|_{L^2}^2 \leq C_n L^{4/3} \mathcal{H}^{n-1}(\partial K)^{2/3} \|u - v\|_{L^2}^{2/3},$$

which is what we wanted to prove. □

3.3 Stability of Potentials for $p > 1$

Next, we discuss the more general case of $p > 1$. To prove stability of transport potentials in this case, we utilize entropic optimal transport. We will need to differentiate between the cases $p \geq 2$ and $1 < p < 2$. The main difference is that for $p \geq 2$, we can shift the cost $|x - y|^p$ to be concave, which turns out to be relevant for the proof. For $p < 2$, the second derivative of $|x - y|^p$ explodes near $x = y$, which prevents us from doing so and means the proof in this case will be more involved.

We first recall some results about entropic optimal transport. The main problem is minimizing

$$\min_{\pi \in \Pi(\mu, \nu)} \left\{ \int_{X \times Y} c(x, y) d\pi + \varepsilon H(\pi | \mu \otimes \nu) \right\}.$$

We can reduce this problem to the case $\varepsilon = 1$ by setting $c_\varepsilon := c/\varepsilon$. Then we can reduce this to the problem studied in section 2.4 by setting $\rho := \alpha e^{-c} \mu \otimes \nu$, for a suitable $\alpha > 0$ that turns this into a probability measure, and then minimizing $H(\pi|\rho)$. By introducing a reference measure $\sigma >> \nu$, we can write

$$\frac{d\pi}{d(\mu \otimes \sigma)} = \frac{d\pi}{d\mu \otimes \nu}\frac{d\nu}{d\sigma},$$

so

$$\int_{X\times Y} c(x,y)d\pi + \varepsilon H(\pi|\mu\otimes\nu) = \int_{X\times Y} c(x,y)d\pi + \varepsilon H(\pi|\mu\otimes\sigma) - \varepsilon H(\nu|\sigma).$$

The dual formulation is

$$\sup_{\psi\in L^1(\sigma)} \left\{ \int_X \psi^{c,\varepsilon} d\rho + \int_Y \psi d\sigma - \varepsilon H(\nu|\sigma) \right\},$$

where

$$\psi^{c,\varepsilon}(x) = -\varepsilon \ln \int_Y e^{(\psi(y)-c(x,y))/\varepsilon} d\sigma. \tag{34}$$

Note that this definition is independent of the target measure ν considered.

For convenience, in what follows we will often reduce to the case where the target measure is discrete. Since every measure is a weak limit of discrete measures, this will not hurt the generality of the arguments. If $Y = \{y_1, \ldots, y_m\}$, denote $\psi_i = \psi(y_i)$, so we can identify ψ as an element of $\mathbb{R}^m$. Then (34) becomes

$$\psi^{c,\varepsilon}(x) = -\varepsilon \ln \left(\sum_{i=1}^{n} \exp\left(\frac{\psi_i - c(x,y_i)}{\varepsilon} \right) \sigma_i \right).$$

We can also differentiate $\psi^{c,\varepsilon}(x)$ with respect to its coordinates, and get

$$\frac{\partial(\psi^{c,\varepsilon}(x))}{\partial\psi_i} = -\frac{\exp\left(\frac{\psi_i - c(x,y_i)}{\varepsilon}\right)\sigma_i}{\sum_{j=1}^m \exp\left(\frac{\psi_j - c(x,y_j)}{\varepsilon}\right)\sigma_j},$$

and the second derivatives are, for $i \neq j$,

$$\begin{aligned}
\frac{\partial^2(\psi^{c,\varepsilon}(x))}{\partial\psi_i^2} &= \frac{1}{\varepsilon}\left(-\frac{\exp\left(\frac{\psi_i - c(x,y_i)}{\varepsilon}\right)\sigma_i}{\sum_{j=1}^m \exp\left(\frac{\psi_j - c(x,y_j)}{\varepsilon}\right)\sigma_j} + \frac{\exp\left(\frac{\psi_i - c(x,y_i)}{\varepsilon}\right)^2\sigma_i^2}{\left(\sum_{j=1}^m \exp\left(\frac{\psi_j - c(x,y_j)}{\varepsilon}\right)\sigma_j\right)^2} \right), \\
\frac{\partial(\psi^{c,\varepsilon}(x)}{\partial\psi_i\partial\psi_j} &= \frac{1}{\varepsilon}\frac{\exp\left(\frac{\psi_i - c(x,y_i)}{\varepsilon}\right)}{\sum_{j=1}^m \exp\left(\frac{\psi_j - c(x,y_j)}{\varepsilon}\right)\sigma_j}\frac{\exp\left(\frac{\psi_j - c(x,y_j)}{\varepsilon}\right)}{\sum_{j=1}^m \exp\left(\frac{\psi_j - c(x,y_j)}{\varepsilon}\right)\sigma_j}\sigma_i\sigma_j.
\end{aligned}$$

Let $\pi_\psi^\varepsilon(x)$ be the discrete probability distribution which equals i with probability $-\frac{\partial(\psi^{c,\varepsilon}(x))}{\partial\psi_i}$. As in the $p = 2$ case, denote the Kantorovich functional to be $\mathcal{K}^\varepsilon(\psi) := -\mathbb{E}_\mu[\psi^{c,\varepsilon}]$. Then the above calculations give us, for $v \in \mathbb{R}^m$:

$$\begin{aligned}
\nabla\mathcal{K}^\varepsilon(\psi)\cdot v &= \int_\Omega \mathbb{E}_{\pi_\psi^\varepsilon(x)}[v]d\mu, \\
\langle v, D^2\mathcal{K}^\varepsilon(\psi)v\rangle &= \frac{1}{\varepsilon}\int_\Omega \mathrm{Var}_{\pi_\psi^\varepsilon(x)}(v)d\mu.
\end{aligned} \tag{35}$$

In the course of the proof, we will also use the function $Z_\beta(\psi)$, defined as

$$Z_\beta(\psi) = \int_\Omega e^{\beta \psi^{c,\varepsilon}} d\mu. \tag{36}$$

We can compute the Hessian of $\ln Z_\beta$ to be

$$\begin{aligned}
\frac{\partial \ln Z_\beta(\psi)}{\partial \psi_i} &= \frac{1}{Z_\beta(\psi)} \int_\Omega \beta e^{\beta\psi^{c,\varepsilon}} \cdot \frac{\partial \psi^{c,\varepsilon}}{\partial \psi_i} d\mu, \\
\frac{\partial^2 \ln Z_\beta(\psi)}{\partial \psi_i \partial \psi_j} &= \frac{1}{Z_\beta(\psi)} \int_\Omega \beta^2 e^{\beta\psi^{c,\varepsilon}} \cdot \frac{\partial \psi^{c,\varepsilon}}{\partial \psi_i} \frac{\partial \psi^{c,\varepsilon}}{\partial \psi_j} + \beta e^{\beta\psi^{c,\varepsilon}} \frac{\partial^2 \psi^{c,\varepsilon}}{\partial \psi_i \psi_j} d\mu,
\end{aligned}$$

and so, we can write

$$\langle v, D^2 \ln Z_\beta(\psi) v\rangle = -\frac{\beta}{\varepsilon} \int_\Omega \mathrm{Var}_{\pi^\varepsilon_\psi}(v) d\mu^{\beta,\varepsilon}_\psi + \beta^2 \mathrm{Var}_{\mu^\varepsilon_\psi}(\mathbb{E}_{\pi^\varepsilon_\psi}[v]), \tag{37}$$

where

$$\mu^{\beta,\varepsilon}_\psi = \frac{1}{Z_\beta(\psi)} e^{\beta\psi^{c,\varepsilon}} \mu.$$

Now, we can proceed to the proof. We first prove that Z_β is log-concave.

Proposition 24 (Log-Concavity)**.** *Let μ be a log-concave probability measure supported on a compact and convex set Ω, let $Y = \{y_1, \dots, y_m\}$ be discrete, and let $c : \Omega \times Y \to \mathbb{R}$ be a cost function that is concave on its first coordinate. Then the function $Z_\beta(\psi)$ defined as in (36) is log-concave, so*

$$Z_\beta(\psi_t) \geq Z_\beta(\psi_0)^{1-t} Z_\beta(\psi_1)^t,$$

Whenever $\psi_t = t\psi_1 + (1-t)\psi_0, 0 \leq t \leq 1$.

Proof. Since $Z_\beta(\psi) = Z_1(\beta\psi)$, we can assume for simplicity that $\beta = 1$. Let $x_0, x_1 \in X$, and $x_t = tx_1 + (1-t)x_0$. Then by concavity of c, and hence convexity of $-c$:

$$\begin{aligned}
&-\psi_t^{c,\varepsilon}(x_t) = \varepsilon \ln \left(\sum_{i=1}^n \exp\left(\frac{(\psi_t)_i - c(x_t, y_i)}{\varepsilon} \right) \sigma_i \right) \\
&\leq \varepsilon \ln \left(\left(\sum_{i=1}^n \exp\left(\frac{(\psi_1)_i - c(x_1, y_i)}{\varepsilon} \right) \cdot \sigma_i \right)^t \cdot \left(\sum_{i=1}^n \exp\left(\frac{(\psi_0)_i - c(x_0, y_i)}{\varepsilon} \right) \cdot \sigma_i \right)^{1-t} \right) \\
&\leq -t\psi_1^{c,\varepsilon}(x_1) - (1-t)\psi_0^{c,\varepsilon}(x_0),
\end{aligned} \tag{38}$$

where we used Hölder's inequality with exponentials $1/t, 1/(1-t)$. Denoting $h_t(x) = e^{\psi_t^{c,\varepsilon}(x)}$, (38) becomes

$$h_t(x_t) \geq h_0(x_0)^{1-t} h_1(x_1)^t.$$

Since $Z_1(\psi) = \|h_t\|_{L^1(\mu)}$, the statement now follows from the Prékopa-Leindler inequality. ☐

Proposition 25. *Under the assumptions of of proposition 24, if $\psi_0, \psi_1 \in C(Y)$, $\psi_t = t\psi_1 + (1-t)\psi_0$, and $\varphi_t^\varepsilon = \psi_t^{c,\varepsilon}$ for $0 \leq t \leq 1$, then*

$$\mathrm{Var}_\mu(\varphi_1^\varepsilon - \varphi_0^\varepsilon) \leq 2eM_{c,\varepsilon} \langle \psi_1 - \psi_0, \nabla \mathcal{K}^\varepsilon(\psi_1) - \nabla \mathcal{K}^\varepsilon(\psi_0) \rangle,$$

where

$$M_{c,\varepsilon} = \sup_{t\in[0,1]} \sup_{x,y\in\Omega} \{\varphi_t^\varepsilon(x) - \varphi_t^\varepsilon(y)\}.$$

Proof. By the previous proposition, we get that $D^2 \ln Z_\beta(\psi)$ is negative semidefinite for every $\psi \in C(Y)$; using (37), this equates to, for all $v \in \mathbb{R}^m, 0 \le t \le 1$:

$$\beta \mathrm{Var}_{\mu^{\beta,\varepsilon}_{\psi_t}} \left(\mathbb{E}_{\pi^\varepsilon_{\psi_t}}[v]\right) \le \frac{1}{\varepsilon} \int_\Omega \mathrm{Var}_{\pi^\varepsilon_{\psi_t}}(v) d\mu^{\beta,\varepsilon}_{\psi_t}. \tag{39}$$

Recalling that $\mu^{\beta,\varepsilon}_{\psi_t} = \frac{1}{Z_\beta(\psi_t)} e^{\beta\psi_t^{c,\varepsilon}}\mu$, we get the inequalities

$$e^{-\beta M_{c,\varepsilon}} \mu^{\beta,\varepsilon}_{\psi_t} \le \mu \le e^{\beta M_{c,\varepsilon}} \mu^{\beta,\varepsilon}_{\psi_t}. \tag{40}$$

Plugging these inequalities into (39), we get

$$\mathrm{Var}_\mu \left(\mathbb{E}_{\pi^\varepsilon_{\psi_t}}[v]\right) \le \frac{e^{2\beta M_{c,\varepsilon}}}{\beta\varepsilon} \int_\Omega \mathrm{Var}_{\pi^\varepsilon_{\psi_t}}(v) d\mu = \frac{e^{2\beta M_{c,\varepsilon}}}{\beta} \langle v, D^2\mathcal{K}^\varepsilon(\psi_t) v\rangle. \tag{41}$$

To proceed, we use the chain rule to get

$$\frac{d}{dt}\psi_t^{c,\varepsilon} = \nabla\psi_t^{c,\varepsilon} \cdot (\psi_1 - \psi_0) = \mathbb{E}_{\pi^\varepsilon_{\psi_t}}[\psi_1 - \psi_0].$$

Using convexity of the variance, we can bound the left hand side of (41) from below as

$$\mathrm{Var}_\mu(\varphi_1^\varepsilon - \varphi_0^\varepsilon) = \mathrm{Var}_\mu \left(\int_0^1 \frac{d}{dt}\psi_t^{c,\varepsilon} dt\right) \le \int_0^1 \mathrm{Var}_\mu \left(\mathbb{E}_{\pi^{c,\varepsilon}_{\psi_t}}[\psi_1 - \psi_0]\right) dt.$$

Continuing, using $v := \psi_1 - \psi_0$, we get

$$\begin{aligned}
\int_0^1 \mathrm{Var}_\mu \left(\mathbb{E}_{\pi^{c,\varepsilon}_{\psi_t}}[\psi_1 - \psi_0]\right) dt &\le \int_0^1 \frac{e^{2\beta M_{c,\varepsilon}}}{\beta} \langle \psi_1 - \psi_0, D^2\mathcal{K}^\varepsilon(\psi_t)(\psi_1 - \psi_0)\rangle dt \\
&\le \frac{e^{2\beta M_{c,\varepsilon}}}{\beta} \left\langle \psi_1 - \psi_0, \int_0^1 D^2\mathcal{K}^\varepsilon(\psi_t)(\psi_1 - \psi_0) dt \right\rangle \\
&= \frac{e^{2\beta M_{c,\varepsilon}}}{\beta} \left\langle \psi_1 - \psi_0, \int_0^1 \frac{d}{dt}\nabla\mathcal{K}^\varepsilon(\psi_t) dt \right\rangle \\
&= \frac{e^{2\beta M_{c,\varepsilon}}}{\beta} \langle \psi_1 - \psi_0, \nabla\mathcal{K}^\varepsilon(\psi_1) - \nabla\mathcal{K}^\varepsilon(\psi_0)\rangle.
\end{aligned}$$

Finally, we optimize for β by setting $\beta := 1/2M_{c,\varepsilon}$, and get

$$\mathrm{Var}_\mu(\varphi_1^\varepsilon - \varphi_0^\varepsilon) \le 2eM_{c,\varepsilon} \langle \psi_1 - \psi_0, \nabla\mathcal{K}^\varepsilon(\psi_1) - \nabla\mathcal{K}^\varepsilon(\psi_0)\rangle.$$

□

Now we can prove the main result.

Theorem 21. *Suppose μ is a log-concave probability measure supported on a bounded convex set Ω, and ν_0, ν_1 are contained in a bounded set X. Let the maximal magnitudes of Ω, X be denoted by R_Ω, R_X. Let φ_0, φ_1 the potentials associated with transporting from μ to ν_0, ν_1 for the p-cost. Then*

$$\mathrm{Var}_\mu(\varphi_1 - \varphi_0) \leq 2e\mathrm{diam}(\Omega)(R_\Omega + R_X)^{p-1} \int_X (\psi_1 - \psi_0) d(\nu_0 - \nu_1).$$

Proof. We will first assume that Y is discrete. If c is as in the previous proposition, then normalizing the φ_i^ε and letting $\varepsilon \to 0$, we get

$$\mathrm{Var}_\mu(\varphi_1 - \varphi_0) \leq 2eM_c \langle \psi_1 - \psi_0, \nabla\mathcal{K}(\psi_1) - \nabla\mathcal{K}(\psi_0)\rangle,$$

where

$$\mathcal{K}(\psi) = -\mathbb{E}_\mu[\psi^c],$$
$$M_c = \sup_{t\in[0,1]} \sup_{x,y\in\Omega} \{\varphi_t(x) - \varphi_t(y)\},$$

and here $\varphi_t = \psi_t^c$. By definition, this means that

$$\varphi_t(y) = \sup_{x\in X} \frac{|x-y|^p}{p} - \psi_t(x).$$

This is a supremum of $(R_\Omega + R_X)^{p-1}$-Lipschitz continuous functions, and is therefore also Lipschitz continuous with the same Lipschitz constant. Therefore

$$|\varphi_t(x) - \varphi_t(y)| \leq (R_\Omega + R_X)^{p-1}|x-y| \leq 2\mathrm{diam}(\Omega)(R_\Omega + R_X)^{p-1},$$

so

$$M_c \leq 2\mathrm{diam}(\Omega)(R_\Omega + R_X)^{p-1},$$

and this inequality carries over to non-discrete probability measures by approximation.

Let $\tilde{c}(x,y) = |x-y|^p - \gamma|x|^2$, where γ is equal to

$$\gamma = \frac{1}{2} \sup_{x\in\Omega, y\in X} \left\{ \left\| D_x^2 c(x,y) \right\|_{\mathrm{op}} \right\}.$$

Then $\tilde{c}$ is concave in the first variable, so we have

$$\mathrm{Var}_\rho(\tilde{\varphi}_1 - \tilde{\varphi}_0) \leq 2eM_{\tilde{c}} \langle \psi_1 - \psi_0, \nabla\mathcal{K}(\psi_1) - \nabla\mathcal{K}(\psi_0)\rangle.$$

By the definition, for $i = 0,1$ we can see that $\tilde{\varphi}_i = \varphi_i^c + \gamma|x|^2$, so that $\tilde{\varphi}_1 - \tilde{\varphi}_0 = \varphi_1 - \varphi_0$. Therefore

$$\begin{aligned}\mathrm{Var}_\rho(\varphi_1 - \varphi_0) &\leq 2M_{\tilde{c}} \langle \psi_1 - \psi_0, \nabla\mathcal{K}(\psi_1) - \nabla\mathcal{K}(\psi_0)\rangle \\ &= 2e\mathrm{diam}(\Omega)(R_\Omega + R_X)^{p-1} \int_X (\psi_1 - \psi_0) d(\nu_0 - \nu_1),\end{aligned}$$

where the last equality follows from (35). The theorem for non discrete measures μ_0, μ_1 follows from the fact that every compactly supported measure is the weak limit of discrete measures. □

We now discuss $1 < p < 2$. Unlike in the $p \geq 2$ case, we cannot reduce to a concave cost, since the second derivatives of $|x-y|^p$ explode near the diagonal. Therefore our proofs will be more complicated, and the resulting bounds will be worse. We will rely on the following weaker version of concavity.

Lemma 5. *Let $1 < p < 2$, and $x, y \in \mathbb{R}^n, 0 < t < 1$. Then*

$$|ty + (1-t)x|^p \geq t|y|^p + (1-t)|x|^p - \gamma t(1-t)|x-y|^p,$$

for some $\gamma > 0$ depending only on p.

Proof. According to [7, chapter 12], we have the following inequality:

$$p\langle (a-z)^{(p-1)} - (b-z)^{(p-1)}, a-b\rangle \leq \gamma |a-b|^p, \tag{42}$$

for all $a, b, z \in \mathbb{R}^n$, and for some $\gamma > 0$ depending only on p. By convexity of $|\cdot|^p$ and the fact that convex functions are above their gradient, we have

$$\begin{aligned}
|ty + (1-t)x|^p &\geq |x|^p + \langle px^{(p-1)}, t(y-x)\rangle, \\
|ty + (1-t)x|^p &\geq |y|^p + \langle py^{(p-1)}, (1-t)(x-y)\rangle.
\end{aligned}$$

Linearly combining these inequalities, we get

$$|ty + (1-t)x|^p \geq t|y|^p + (1-t)|x|^p + pt(1-t)\langle x^{(p-1)} - y^{(p-1)}, y-x\rangle,$$

and finally using (42), we prove the result. $\square$

For $1 < p$, there exists $\gamma \in \mathbb{R}$ depending only on p such that

$$|tx_1 + (1-t)x_0|^p \geq t|x_1|^p + (1-t)|x_0|^p - \gamma t(1-t)|x_0 - x_1|^p. \tag{43}$$

We also cite the following fact which we will use in the proof.

Lemma 6. *Let ρ be a log-concave probability measure, and let f, g be nonnegative smooth functions. Define $\mu = f\rho, \nu = g\rho$. Then there exists a constant C depending only on ρ such that*

$$W_2(\mu,\nu)^2 \leq \frac{C}{\inf f}\|f-g\|^2_{L^2(\rho)}. \tag{44}$$

Proposition 26. *Denote $I_\beta(\psi) = \ln Z_\beta(\psi)$, and $\psi_t = (1-t)\psi_0 + t\psi_1$. Then*

$$I_\beta(\psi_t) \geq (1-t)I_\beta(\psi_0) + tI_\beta(\psi_1) - \beta\gamma t(1-t)W_p\left(\mu^{\beta,\varepsilon}_{\psi_0^{c,\varepsilon}}, \mu^{\beta,\varepsilon}_{\psi_1^{c,\varepsilon}}\right)^p.$$

Proof. Again, for simplicity, assume $\beta = 1$. Using the bound (43), we can proceed similarly to proposition 24, and get

$$\begin{aligned}
\psi_t^{c,\varepsilon}(x_t) &= -\varepsilon \ln\left(\sum_{i=1}^n \exp\left(\frac{(\psi_t)_i - c(x_t, y_i)}{\varepsilon}\right)\sigma_i\right) \\
&\geq t\psi_1^{c,\varepsilon}(x_1) + (1-t)\psi_0^{c,\varepsilon}(x_0) - \gamma t(1-t)|x_0 - x_1|^p.
\end{aligned}$$

Let $h_t = e^{\psi_t^{c,\varepsilon}}, \mu_i = h_i\mu$ for $i = 0, 1$, and normalize ψ_0, ψ_1 so $\|h_0\|_{L^1(\mu)}, \|h_1\|_{L^1(\mu)} = 1$. Then

$$h_t(x_t) \geq h_1(x_1)^t h_0(x_0)^{1-t} \exp\left(-\gamma t(1-t)|x_0 - x_1|^p\right).$$

Now let T be the optimal transport map from μ_0 to μ_1, and let $T_t = tT + (1-t)\mathrm{id}, \mu_t = T_t\#\mu_0$. Then according to (16):

$$\begin{aligned}\mu_t(T_t(x)) &\leq \mu_0(x)^{1-t}\mu_1(T(x))^t \\ &= h_0(x)^{1-t}\mu(x)^{1-t}h_1(T(x))^t\mu(T(x))^t \\ &\leq h_t(T_t(x))\exp\left(\gamma t(1-t)|x-T(x)|^p\right)\mu(T_t(x)),\end{aligned}$$

which means for μ_t-almost every x, we have

$$\mu_t(x) \leq h_t(x)\mu(x)\exp\left(\gamma t(1-t)|T_t^{-1}(x) - T(T_t^{-1}(x))\right). \tag{45}$$

Integrating the exponential term with respect to ρ_t and using Jensen's inequality, we have

$$\begin{aligned}&\int_\Omega \exp\left(-\gamma t(1-t)|T_t^{-1}(x) - T(T_t^{-1}(x))\right)d\mu_t \\ &\geq \exp\left(-\gamma t(1-t)\int_\Omega |T_t^{-1}(x) - T(T_t^{-1}(x))d\mu_t\right) \\ &= \exp\left(-\gamma t(1-t)\int_\Omega |x-T(x)|^p)h_0 d\mu\right) = \exp\left(-\gamma t(1-t)W_p(\mu_0,\mu_1)^p\right).\end{aligned}$$

Integrating the rest of (45), we get

$$\begin{aligned}\|h_t\|_{L^1(\mu)} &\geq \exp\left(-\gamma t(1-t)W_p(\mu_0,\mu_1)^p\right) \\ &= \exp\left(-\gamma t(1-t)W_p\left(\mu^{1,\varepsilon}_{\psi_0^{c,\varepsilon}}, \mu^{1,\varepsilon}_{\psi_1^{c,\varepsilon}}\right)^p\right)\|h_0\|^{1-t}_{L^1(\mu)}\|h_1\|^t_{L^1(\mu)}.\end{aligned}$$

Recalling that $Z_1(\psi_t) = \|h_t\|_{L^1(\rho)}$, we get the desired conclusion. □

Proposition 27. *Under the same assumptions as theorem 18, we have*

$$\mathrm{Var}_\mu(\varphi_1 - \varphi_0)^{p/(p-1)} \leq C_{X,\mu,p}\int_X (\psi_1 - \psi_0)d(\nu_0 - \nu_1) \leq C_{X,\mu,p}W_1(\nu_0,\nu_1).$$

Proof. By the previous proposition and using Using standard facts about convexity, we get that

$$\langle \psi_1 - \psi_0, \nabla I_\beta(\psi_1) - \nabla I_\beta(\psi_0)\rangle \leq 2\beta\gamma W_p(\mu_{\beta\psi_0^{c,\varepsilon}}, \mu_{\beta\psi_1^{c,\varepsilon}})^p. \tag{46}$$

Now, we retrace the steps of the proof of proposition 25 with the extra term. First define

$$\langle D^2 I_\beta(\psi_t)\cdot(\psi_1-\psi_0), \psi_1-\psi_0\rangle = b_t.$$

Then by (46) we have that

$$\int_0^1 b_t dt \leq 2\beta\gamma W_p(\mu_{\beta\psi_0^{c,\varepsilon}}, \mu_{\beta\psi_1^{c,\varepsilon}})^p. \tag{47}$$

Using (37) we also have that

$$\frac{b_t}{\beta} + \frac{1}{\varepsilon}\int_\Omega \mathrm{Var}_{\pi^\varepsilon_{\psi_t}}(\psi_1-\psi_0)d\mu_t = \beta\mathrm{Var}_{\mu_{\psi_t}}(\mathbb{E}_{\pi^\varepsilon_{\psi_t}}[\psi_1-\psi_0]).$$

And using (40) we conclude

$$\beta e^{-\beta M_{c,\varepsilon}} \mathrm{Var}_\mu \left(\mathbb{E}_{\pi^{c,\varepsilon}_{\psi_t}}[v] \right) \leq \frac{e^{\beta M_{c,\varepsilon}}}{\varepsilon} \int_\Omega \mathrm{Var}_{\pi^{c,\varepsilon}_{\psi_t}}(v) d\mu + \frac{b_t}{\beta}$$
$$= e^{\beta M_{c,\varepsilon}} \langle v, D^2 K^{c,\varepsilon}(\psi_t) v \rangle + \frac{b_t}{\beta}.$$

Now if we follow the rest of the steps in proposition 25 and use 47, we find that

$$\beta e^{-\beta M_{c,\varepsilon}} \mathrm{Var}_\mu(\varphi_1 - \varphi_0) \leq e^{\beta M_{c,\varepsilon}} \langle \psi_1 - \psi_0, \nabla K^{c,\varepsilon}(\psi_1) - \nabla K^{c,\varepsilon}(\psi_1) \rangle + 2\gamma W_p \left(\mu^{\beta,\varepsilon}_{\psi_0^{\beta,\varepsilon}}, \mu^{\beta,\varepsilon}_{\psi_1^{c,\varepsilon}} \right)^p. \quad (48)$$

Denote

$$\mu_\varphi = \frac{1}{Z(\varphi)} e^\varphi \mu.$$

Then letting $\varepsilon \to 0$ in (48), we finally get

$$\beta e^{-\beta M_c} \mathrm{Var}_\mu(\varphi_1 - \varphi_0) \leq e^{\beta M_c} \int_X (\psi_1 - \psi_0) d(\nu_0 - \nu_1) + 2\gamma W_p(\mu_{\beta\varphi_0}, \mu_{\beta\varphi_1})^p. \quad (49)$$

We want to bound the last term on the right hand side. For simplicity we again assume $\beta = 1$. Using (44), we see that

$$W_p(\mu_{\varphi_0}, \mu_{\varphi_1}) \leq W_2(\mu_{\varphi_0}, \mu_{\varphi_1}) \leq C e^{M_c} \left\| \frac{e^{\varphi_1}}{\int_\Omega e^{\varphi_1} d\mu} - \frac{e^{\varphi_0}}{\int_\Omega e^{\varphi_0} d\mu} \right\|^2_{L^2(\mu)}. \quad (50)$$

Going further, we see

$$\left| \frac{e^{\varphi_1(x)}}{\int_\Omega e^{\varphi_1} d\mu} - \frac{e^{\varphi_0(x)}}{\int_\Omega e^{\varphi_0} d\mu} \right| \leq \left| \frac{e^{\varphi_1(x)} - e^{\varphi_0(x)}}{\int_\Omega e^{\varphi_1} d\mu} \right| + e^{\varphi_0(x)} \left| \frac{1}{\int_\Omega e^{\varphi_1} d\mu} - \frac{1}{\int_\Omega e^{\varphi_0} d\mu} \right|$$
$$\leq e^{M_c} |e^{\varphi_1(x)} - e^{\varphi_0(x)}| + e^{2M_c} \left| \int_\Omega (e^{\varphi_1} - e^{\varphi_0}) d\mu \right|$$
$$\leq e^{2M_c} |\varphi_1(x) - \varphi_0(x)| + e^{2M_c} \|\varphi_1 - \varphi_0\|_{L^1(\mu)},$$

and integrating, we get

$$\left\| \frac{e^{\varphi_1(x)}}{\int_\Omega e^{\varphi_1} d\mu} - \frac{e^{\varphi_0(x)}}{\int_\Omega e^{\varphi_0} d\mu} \right\|^2_{L^2(\mu)} \leq e^{4M_c} \left(\|\varphi_1 - \varphi_0\|_{L^2(\mu)} + \|\varphi_1 - \varphi_0\|_{L^1(\mu)} \right)^2$$
$$\leq 4 e^{4M_c} \|\varphi_1 - \varphi_0\|^2_{L^2(\mu)}. \quad (51)$$

Now we can combine (49), (50) and (51) to get

$$\beta e^{-\beta M_c} \mathrm{Var}_\mu(\varphi_1 - \varphi_0) \leq e^{\beta M_c} \int_X (\psi_1 - \psi_0) d(\nu_0 - \nu_1) + C e^{5p\beta M_c/2} \beta^p \|\varphi_1 - \varphi_0\|^p_{L^2(\mu)},$$

and changing sides, we find

$$\beta e^{-2\beta M_c} \|\varphi_1 - \varphi_0\|^2_{L^2(\mu)} - C e^{\frac{(5p-2)\beta}{2} M_c} \beta^p \|\varphi_1 - \varphi_0\|^p_{L^2(\mu)} \leq \int_X (\psi_1 - \psi_0) d(\mu_0 - \nu_1)$$

We want to choose β which maximizes the left hand side. If we ignore the exponential terms, the optimal β is given by

$$\beta := \left(\frac{\|\varphi_1 - \varphi_0\|_{L^2(\mu)}^{2-p}}{Cp} \right)^{1/(p-1)}.$$

Denote $\alpha := M_c, \alpha' := \frac{5p-2}{2} M_c$, and $\Phi := \|\varphi_1 - \varphi_0\|_{L^2(\mu)}^{(2-p)/(p-1)}$. Now, we define

$$\begin{aligned}
f(\beta) &:= \beta e^{-\alpha\beta} \|\varphi_1 - \varphi_0\|_{L^2(\mu)}^2 - C\beta^p e^{\alpha'\beta} \|\varphi_1 - \varphi_0\|_{L^2(\mu)}^p, \\
g(\varepsilon) &= \varepsilon e^{-\alpha\Phi\varepsilon} - C\varepsilon^p e^{\alpha'\Phi\varepsilon}, \\
h(\varepsilon) &= \varepsilon e^{-\alpha\varepsilon} - C\varepsilon^p e^{\alpha'\varepsilon}.
\end{aligned}$$

Since the values of φ_0, φ_1 are bounded, we can assume for simplicity that $\|\varphi_1 - \varphi_0\|_{L^2(\mu)} \leq 1$, so that $\Phi \leq 1$ as well. In that case, we have $g(\varepsilon) \geq h(\varepsilon)$, and we also have $f(\beta) = \|\varphi_1 - \varphi_0\|_{L^2(\mu)}^{p/(p-1)} g(\varepsilon)$ if $\varepsilon = \beta/\Phi$. Therefore

$$\int_X (\psi_1 - \psi_0) d(\nu_0 - \nu_1) \geq f(\beta) = \|\varphi_1 - \varphi_0\|_{L^2(\mu)}^{p/(p-1)} g(\varepsilon) \geq \|\varphi_1 - \varphi_0\|_{L^2(\mu)}^{p/(p-1)} h(\varepsilon).$$

The maximal value of $h(\varepsilon)$ is some quantity independent of φ_1, φ_0, which finishes the proof. $\square$

3.4 Stability of the Transport Maps for $p > 1$

To obtain stability of the transport maps, we recall that the transport plans can be expressed as

$$T(x) = x - (\nabla\varphi)^{1/(p-1)},$$

where φ is the potential associated with μ, ν. We will also use the following inequality:

$$\|v^{(\alpha)} - w^{(\alpha)}\| \leq C\|v - w\|^\alpha,$$

for some constant C depending only on α and the dimension n. Using these two relations and the reverse Gagliardo-Nirenberg inequality from theorem 20, for $p > 2$ we can simply write

$$\begin{aligned}
\|T_2 - T_1\|_{L^2(\mu)}^2 &= \int_\Omega \|(\nabla\varphi_2)^{1/(p-1)} - (\nabla\varphi_1)^{1/(p-1)}\|^2 d\mu \\
&\leq C \int_\Omega \|\nabla\varphi_2 - \nabla\varphi_1\|^{2/(p-1)} d\mu \\
&\leq C\|\nabla\varphi_2 - \nabla\varphi_1\|_{L^2(\mu)}^{2/(p-1)} \\
&\leq C\|\varphi_2 - \varphi_1\|_{L^2(\mu)}^{2/(p-1)},
\end{aligned}$$

where the last inequality follows from the fact that by following reasoning similar to the proof of theorem 9, we see that by adding a factor of the form $\gamma\frac{|x|^2}{2}$, we can transform the functions φ_1, φ_2 to be concave, and then apply theorem 20.

For $1 < p < 2$, the situation is again different. We will use a fractional Sobolev inequality, but first let us define fractional Sobolev spaces. Let $u : \Omega \to \mathbb{R}$, for an open domain $\Omega \subseteq \mathbb{R}^n$. Also let $0 < \alpha < 1$, $1 \leq p < \infty$. Define

$$[u]_{\alpha,p} := \left(\int_{\Omega\times\Omega} \frac{|u(x) - u(y)|^p}{|x - y|^{n+\alpha p}} dxdy \right)^{1/p},$$

and

$$\|u\|_{W^{\alpha,p}(\Omega)} = \|u\|_{L^p(\Omega)} + [u]_{\alpha,p}.$$

Define $W^{\alpha,p}(\Omega)$ to be the space of all functions u for which $\|u\|_{W^{\alpha,p}(\Omega)} < \infty$. For $p = \infty$, define $W^{s,\infty}(\Omega)$ to be the Hölder space $C^s(\Omega)$, with the regular norm.

For $s > 1$, we define the fractional Sobolev space $W^{s,p}(\Omega)$ as follows. Express $s = m + \alpha$, where $m \in \mathbb{N}, 0 \leq \alpha < 1$. Then we can define

$$\|u\|_{W^{s,p}(\Omega)} := \|u\|_{W^{m,p}(\Omega)} + \sup_{|\alpha|=m} \|D^\alpha u\|_{W^{\alpha,p}(\Omega)},$$

and define $W^{s,p}$ to be the space of functions for which this norm is less than ∞.

Proposition 28. *Let $1 < r < 2$. Let $u \in L^2(\Omega) \cap W^{1,\infty}(\Omega) \cap W^{r,1}(\Omega)$. Then $u \in H_1(\Omega)$, and we have*

$$\|u\|_{H_1(\Omega)} \leq C\|u\|_{L^2(\Omega)}^{1-\frac{2}{1+r}} \|u\|_{W^{1,\infty}(\Omega)}^{\frac{1}{1+r}} \|u\|_{W^{r,1}(\Omega)}^{\frac{1}{1+r}}.$$

Proof. The inequality follows from the more standard fractional Gagliardo-Nirenberg inequality, which we cite from [3]. It states that for $0 \leq s_1 < s_2, 1 \leq p_1, p_2 \leq \infty$ and $0 < \theta < 1$, defining $s := \theta s_1 + (1-\theta)s_2$ and $\frac{1}{p} := \theta\frac{1}{p_1} + (1-\theta)\frac{1}{p_2}$, we have

$$\|u\|_{W^{s,p}(\Omega)} \leq \|u\|_{W^{s_1,p_1}(\Omega)}^{\theta} \|u\|_{W^{s_2,p_2}(\Omega)}^{1-\theta}, \tag{52}$$

provided that the following does not hold: $s_2 \in \mathbb{N}$, $p_2 = 1$, and $s_1 - \frac{1}{p_1} \geq s_2 - \frac{1}{p_2}$.

We can iterate on the inequality, so that if $0 < \theta_1, \theta_2, \theta_3 < 1$ with $\theta_1 + \theta_2 + \theta_3 = 1, 0 \leq s_1 < s_2 < s_3$, $1 \leq p_1, p_2, p_3 \leq \infty$, and we define $s = \theta_1 s_1 + \theta_2 s_2 + \theta_3 s_3$, and $\frac{1}{p} = \theta_1\frac{1}{p_1} + \theta_2\frac{1}{p_2} + \theta_3\frac{1}{p_3}$, then we have

$$\|u\|_{W^{s,p}(\Omega)} \leq \|u\|_{W^{s_1,p_1}(\Omega)}^{\theta_1} \|u\|_{W^{s_2,p_2}(\Omega)}^{\theta_2} \|u\|_{W^{s_3,p_3}(\Omega)}^{\theta_3},$$

provided that in addition to the previous condition, the following does not hold: $s_3 \in \mathbb{N}, p_3 = 1$ and $s - \frac{1}{p} \geq s_3 - \frac{1}{p_3}$. For our purposes, we set:

$$\begin{aligned}
\theta_1 &= 1 - \frac{2}{1+r}, \\
\theta_2 = \theta_3 &= \frac{1}{1+r}, \\
p_1 &= 2, \\
p_2 &= \infty, \\
p_3 &= 1, \\
s_1 &= 0, \\
s_2 &= 1, \\
s_3 &= r,
\end{aligned}$$

Then the conditions for the inequality are fulfilled, and we obtain (52). $\square$

We also use the following inequality.

Lemma 7. *The following inequality holds, where φ is a potential:*

$$\langle \nabla\varphi(x) - \nabla\varphi(y), x - y\rangle \leq \lambda|x-y|^p. \tag{53}$$

Proof. Suppose that T is the optimal transport map. By c-cyclical monotonicity, we have for all $x, y \in X$

$$|x - T(x)|^p - |y - T(x)|^p \leq |x - T(y)|^p - |y - T(y)|^p. \tag{54}$$

By (42), we have, for $v, h, z \in \mathbb{R}^n$:

$$\begin{aligned}
|v + h - z|^p - |v - z|^p - p\langle (v-z)^{(p-1)}, h\rangle \\
&= p\int_0^1 \langle (v + th - z)^{(p-1)} - (v-z)^{(p-1)}, h\rangle dt \\
&\leq p\int_0^1 \lambda t^{p-1}|h|^p dt = \lambda|h|^p.
\end{aligned} \tag{55}$$

Next, we plug (54) into (55) twice, first using $v = x, z = T(x), h = y - x$ on the left hand side, and then using $v = y, z = T(y), h = x - y$ on the right hand side to get

$$-\lambda|x-y|^p - p\langle (x - T(x))^{(p-1)}, y - x\rangle \leq \lambda|x-y|^p + p\langle (y - T(y))^{(p-1)}, x - y\rangle.$$

Finally, by (9) we have $(x - T(x))^{(p-1)} = \nabla\varphi(x)$ and changing sides, we get

$$\langle \nabla\varphi(x) - \nabla\varphi(y), x - y\rangle \leq \lambda|x-y|^p.$$

□

We now prove a variant of the Gagliardo-Nierenberg type inequality.

Theorem 22. *Let u, v be c-concave function for the p-cost, on some compact domain K with rectifiable boundary. Then*

$$\|\nabla u - \nabla v\|_{L^2} \leq C\|u - v\|_{L^2}^{\theta},$$

for each $\theta < (p-1)/(p+1)$, and for C depending only on θ and K.

With this theorem and the stability of potentials we proved, 18 follows.

Proof of theorem 22. Similarly to the proof of theorem 20, it is enough to prove the 1-dimensional case, and then generalize to higher dimensions using integral geometry. For one dimension, we can exploit (53) to get

$$\begin{aligned}
\frac{|u'(x) - u'(y)|}{|x-y|} &= \frac{|(u'(x) - u'(y)(x-y)|}{(x-y)^2} \\
&\leq \frac{|(u'(x) - u'(y))(x-y) - \lambda(x-y)^p| + \lambda|x-y|^p}{(x-y)^2} \\
&= \frac{(u'(y) - u'(x))}{x-y} + \frac{2\lambda}{|x-y|^{2-p}}.
\end{aligned}$$

Next, pick $\alpha < p - 1$. Computing:

$$\begin{aligned}
[u']_{\alpha,1} &= \int_0^1 \int_0^1 \frac{|u'(x) - u'(y)|}{|x-y|^{\alpha+1}} dxdy \\
&\leq -\int_0^1 \int_0^1 \frac{u'(x) - u'(y)}{(x-y)|x-y|^{\alpha}} dxdy + \int_0^1 \int_0^1 \frac{2\lambda}{|x-y|^{2-p+\alpha}} dxdy \\
&\leq -\int_0^1 \int_{-1}^1 1_{\{0 \leq x+h \leq 1\}} \frac{u'(x+h) - u'(x)}{h|h|^{\alpha}} dhdx + \int_0^1 \int_{-x}^{1-x} \frac{2\lambda}{|h|^{2-p+\alpha}} dhdx
\end{aligned}$$

The second integral converges, since $2 - p + \alpha < 1$. For the first integral, changing the order of integration, we get

$$\begin{aligned}
&\int_0^1 \int_0^1 1_{\{0 \leq x+h \leq 1\}} \frac{u'(x+h) - u'(x)}{h|h|^{\alpha}} dxdh \\
&= \int_0^1 \frac{u(1) - u(h) - u(1-h) + u(0)}{h^{\alpha+1}} dh \leq 2L \int_0^1 \frac{1}{h^{\alpha}} dh < \infty.
\end{aligned}$$

Similarly, the integral for $h < 0$ converges as well. Therefore, we have that $[u']_{\alpha,1}$ is finite and bounded by a constant C, depending only on L, α and p. Therefore, u belongs to the fractional Sobolev space $W^{r,p}$, where $r = 1 + \alpha$. We can take u, v to have mean zero, and the fractional Gagliardo-Nirenberg inequality gives us, for $\theta = 1 - 2/(1+r)$:

$$\|u - v\|_{H^1} \leq C\|u - v\|_{L^2}^{\theta} \cdot \|u - v\|_{W^{1,1}}^{1/(1-r)} \cdot \|u - v\|_{W^{r,1}}^{1/(1+r)} \leq C'\|u - v\|_{L^2}^{\theta}.$$

Since $r < p$, we therefore have the inequality for any $\theta < (p-1)/(p-1)$. $\square$

4 Stability of Optimal Transport - The Case $p = 1$

In this section, we prove stability of potentials in the case $p = 1$. This section contains novel work.

Let μ, ν be probability measures over $\mathbb{R}^n$. We recall the definition of the transport density σ as

$$\sigma = \int_0^1 \pi_t\#(|y - x|\pi)dt,$$

We will prove the following theorems.

Theorem 23. *Suppose μ, ν are probability measures with disjoint supports contained in $X := B_d(0)$, and in addition μ is absolutely continuous, supported on a set Ω with rectifiable boundary, and there exist constants $0 < m, M < \infty$ such that $m \leq \mu \leq M$ on Ω. Let u be the Kantorovich potential associated with (μ, ν), and v be any 1-Lipschitz function. Then*

$$\|\nabla u - \nabla v\|_{L^1(\mu)} \leq C_{d,\mu,\alpha} \left(\int_X (u - v)d(\mu - \nu) \right)^{\alpha},$$

for every $0 < \alpha < 1/3$.

The next theorems have a worse exponent, but apply to probability measures which may not be bounded from below or have unbounded support.

Theorem 24. *Suppose μ, ν are probability measures with disjoint supports contained in $X := B_d(0)$, and in addition μ has a log-concave density. Let u be the Kantorovich potential associated with (μ, ν), and v be any 1-Lipschitz function. Then*

$$\|\nabla u - \nabla v\|_{L^1(\mu)} \leq C_{d,\mu,\alpha} \left(\int_X (u - v) d(\mu - \nu) \right)^{\alpha},$$

for every $0 < \alpha < 1/4$.

Theorem 25. *Suppose μ, ν are probability measures with disjoint supports. Suppose in addition ν has support contained in $X := B_d(0)$, and μ is κ-uniformly log-concave. Let u be the Kantorovich potential associated with (μ, ν), and v be any 1-Lipschitz function. Then*

$$\|\nabla u - \nabla v\|_{L^1(\mu)} \leq C_{d,\mu,\alpha} \left(\int_X (u - v) d(\mu - \nu) \right)^{\alpha},$$

for every $0 < \alpha < 1/4$.

Theorem 26. *Suppose μ, ν are probability measures with disjoint supports contained in $X := B_d(0)$, and in addition μ is absolutely continuous, supported on a set Ω with C^2 boundary, and there exist constants $0 < m, M < \infty$, and $\delta > 0$ such that $md(x, \partial\Omega)^{\delta} \leq \mu \leq M$ on Ω. Let u be the Kantorovich potential associated with (μ, ν), and v be any 1-Lipschitz function. Then*

$$\|\nabla u - \nabla v\|_{L^1(\mu)} \leq C_{d,\mu,\alpha} \left(\int_X (u - v) d(\mu - \nu) \right)^{\alpha},$$

for every $0 < \alpha < 1/4$.

Our method of proof relies on an intermediate step using the transport density. The usefulness of the transport density in the proof is demonstrated by the following lemma.

Lemma 8. *Suppose μ, ν are probability measures with supports contained in $X \subseteq \mathbb{R}^n$, and that μ is absolutely continuous. Let v be a 1-Lipschitz function. Denoting σ, u to be the transport density and potential associated with (μ, ν), we have*

$$\|\nabla u - \nabla v\|^2_{L^2(\sigma)} \leq 2 \int_X (u - v) d(\mu - \nu).$$

Proof. Let π be any optimal transport plan. Recalling that $\pi_t(x,y) = ty + (1-t)x$, then

$$\begin{aligned}
\int_X (u-v)d(\mu-\nu) &= \int_{X\times X} (|x-y| - v(x) + v(y))d\pi \\
&= \int_{X\times X}\int_0^1 (|x-y| - \nabla v(\pi_t(x,y))\cdot(x-y))dtd\pi \\
&= \int_{X\times X}\int_0^1 |x-y|\left(1 - \nabla v(\pi_t(x,y))\cdot \frac{x-y}{|x-y|}\right)dtd\pi \\
&\geq \int_{X\times X}\int_0^1 \frac{|x-y|}{2}\left|\nabla v(\pi_t(x,y)) - \frac{x-y}{|x-y|}\right|^2 dtd\pi \\
&= \frac{1}{2}\int_0^1\int_{X\times X} |x-y||\nabla v(\pi_t(x,y)) - \nabla u(\pi_t(x,y))|^2 d\pi dt \\
&= \frac{1}{2}\int_0^1\int_X |\nabla u - \nabla v|^2 d|E_t|dt = \frac{1}{2}\int_X |\nabla u - \nabla v|^2 d\sigma \\
&= \frac{1}{2}\|\nabla u - \nabla v\|^2_{L^2(\sigma)}.
\end{aligned}$$

□

Next, we aim to obtain a bound in terms of μ. We use proposition 15 to obtain a relation between the measure μ and the transport density σ, namely an upper bound on the Rényi divergence $D_\alpha(\mu||\sigma)$. Recall that for arbitrary probability measures ρ_1, ρ_2 in $\mathbb{R}^n$, the Rényi divergence is given by

$$D_\alpha(\rho_1||\rho_2) := \frac{1}{\alpha-1}\ln\left(\int_{\mathbb{R}^n}\left(\frac{d\rho_1}{d\rho_2}\right)^{\alpha-1} d\rho_1\right).$$

Lemma 9. *Assume the same assumptions as Theorem 23, and in addition that ν is discrete and $n \geq 2$. Then*

$$D_\alpha(\mu||\sigma) \leq 2\ln(M) - \ln(m) + C_{\Omega,\alpha},$$

for every $\alpha < 3/2$.

Proof. Let T be the monotone optimal transport map between μ and ν, which induces an optimal tranport plan π, and let $T_t = tT + (1-t)\mathrm{id}$. We can write

$$\sigma(x) = \int_0^1 \mu_t'(x)dt,$$

where $\mu_t' = \pi_t\#(|y-x|\pi) = T_t\#(|T(x)-x|\mu)$. As a byproduct of the proof of proposition 12, the maps T_t are injective. The determinant approximate differential of $\det T$ is also easily seen to be $(1-t)^n$, since T is piecewise constant. Therefore, writing the push-forward equation, we see

$$\mu_t'(T_t(x)) = \frac{1}{|\det\tilde{\nabla}T|}|T(x)-x|\mu(x) = |T(x)-x|\mu(x)(1-t)^{-n} \geq |T(x)-x|\mu(x).$$

If $x, T_t^{-1}(x) \in \Omega$, having $T(T_t^{-1}(x)) \neq T(x)$ implies that x lies on the line passing through two points in the support of ν; since ν is discrete and $n \geq 2$, it holds that $T(T_t^{-1}(x)) = T(x)$ for almost all $x \in \Omega$ for which $T_t^{-1}(x) \in \Omega$. Therefore, $T_t^{-1}(x)$ is linear in t for every $t < t_0$, where t_0 is the largest value for which $T_{t_0}^{-1}$ is defined. Integrating, we get

$$\sigma(x) \geq |T(x) - x| \int_0^1 \mu(T_t^{-1}(x))dt \geq |T(x) - x| \cdot \frac{m}{d} d(x, \partial\Omega) \geq \frac{m}{d} d(x, \partial\Omega)^2,$$

where here we used the fact that $T(x)$ is outside of Ω, since the supports of μ, ν are disjoint. Therefore

$$\frac{\mu(x)}{\sigma(x)} \leq \frac{dM}{m} \frac{1}{d(x, \partial\Omega)^2}.$$

By proposition (15), the Rényi divergence is therefore bounded by

$$D_\alpha(\mu||\sigma) \leq \frac{1}{\alpha - 1} \ln \left(\int_\Omega \left(\frac{M}{m} \frac{1}{d(x, \partial\Omega)^2} \right)^{\alpha - 1} d\mu \right) \leq 2\ln(M) - \ln(m) + C_{\Omega, X, \alpha},$$

as long as $2(\alpha - 1) < 1$, or equivalently $\alpha < 3/2$. ☐

Now we can prove the result which allows us to transition from a bound in terms of the transport density to a bound in terms of the measure μ.

Lemma 10. *Assume the same assumptions as Theorem 23. Then*

$$\|f\|_{L^1(\mu)} \leq C_{d,\mu,p} \left(\frac{M}{m} \right)^{1/p} \|f\|_{L^p(\sigma)},$$

for every $f \in L^\infty(X)$, $p > 3$.

Proof. We first assume that ν is discrete and $n \geq 2$. Let q be such that $1/p + 1/q = 1$. By Hölder's inequality, we see that

$$\begin{aligned} \|f\|_{L^1(\mu)} &\leq \|f\|_{L^p(\sigma)} \left\| \frac{d\mu}{d\sigma} \right\|_{L^q(\sigma)} = \|f\|_{L^p(\sigma)} \left(\int_\Omega \left(\frac{d\mu}{d\sigma} \right)^{q-1} d\mu \right)^{1/q} = \|f\|_{L^p(\sigma)} \left(e^{(q-1)D_q(\mu||\sigma)} \right)^{1/q} \\ &\leq C_{\Omega, X, p} \|f\|_{L^p(\sigma)} \frac{M^{2/p}}{m^{1/p}}. \end{aligned}$$

For general ν, the result is obtained by approximation by discrete measures. Finally, for $n = 1$, denoting m to be the Lebesgue measure on $[0, 1]$, the lemma follows by replacing μ, ν with $\mu \otimes m, \nu \otimes m$. ☐

Proof of theorem 23. We apply the previous result to $f = |\nabla u - \nabla v|$, which is bounded by 2. We get that

$$\begin{aligned} \|\nabla u - \nabla v\|_{L^1(\mu)} &\leq C_{X,\mu,p} \|\nabla u - \nabla v\|_{L^p(\sigma)} \leq C_{X,\mu} \|\nabla u - \nabla v\|_{L^2(\sigma)}^{2/p} \\ &\leq C_{X,\mu} \left(\int_X (u - v) d(\mu - \nu) \right)^{1/p}. \end{aligned}$$

This inequality holds for any $p > 3$, and hence for any exponent less than 1/3. ☐

Proof of theorems 24 and 25. We will prove a similar result to lemma 10, which will imply the theorems. For any $f \in L^\infty$, we can write

$$\|f\|_{L^1(\mu)} = \|f\|_{L^1(\mu,B_{r,m})} + \|f\|_{L^1(\mu,B_r\setminus B_{r,m})} + \|f\|_{L^1(\mu,\mathbb{R}^n\setminus B_r)},$$

where $B_{r,m} = \{x \in B_r : \mu(x) \geq m\}$. For theorem 24, we can ignore the third term for large enough r. On the first term, we do a calculation similar to the one done in lemma 10:

$$\begin{aligned}
\|f\|_{L^1(\mu,B_{r,m})} &\leq \|f\|_{L^p(\sigma)} \left\| \left(\frac{d\mu}{d\sigma}\bigg|_{B_{r,m}} \right) \right\|_{L^q(\sigma)} \\
&= \|f\|_{L^p(\sigma)} \left(\int_{B_{r,m}} \left(\frac{d\mu}{d\sigma} \right)^{q-1} d\mu \right)^{1/q} \\
&\leq \|f\|_{L^p(\sigma)} M \left(\frac{d}{m} \right)^{1/p} \left(\int_{B_{r,m}} \frac{1}{d(x,\partial B_{r,m})^{2(q-1)}} dx \right)^{1/q} \\
&\leq C_{p,\mu} r^{n/q} \|f\|_{L^p(\sigma)} \left(\frac{d}{m} \right)^{1/p},
\end{aligned}$$

because, by [5, proposition 4.6], the last integral is bounded by r^n times some constant value (independent of r, m for small enough m and big enough r).

The second term is clearly bounded by Cmr^n. The last term is bounded by a tail of a uniformly log-concave distribution, and so bounded by $C_\mu r^{n-2} e^{-\frac{1}{2}\kappa r^2}$. All in all we see that

$$\|f\|_{L^1(\mu)} \leq C_{p,\mu,d} \left(\frac{r^n \|f\|_{L^p(\sigma)}}{m^{1/p}} + mr^n + r^{n-2} e^{-\frac{1}{2}\kappa r^2} \right).$$

Optimizing for m, we get

$$\|f\|_{L^1(\mu)} \leq C_{p,\mu,d} \left(r^n \|f\|_{L^p(\sigma)}^{p/(p+1)} + r^{n-2} e^{-\frac{1}{2}\kappa r^2} \right).$$

Now set $r = (2\kappa^{-1} |\log \|f\|_{L^p(\sigma)}|)^{1/2n}$. We get

$$\|f\|_{L^1(\mu)} \leq C_{p,\mu,d} \|f\|_{L^p(\sigma)}^{p/(p+1)} |\log \|f\|_{L^p(\sigma)}|^{n/2}.$$

Because the only condition on p is that it is bigger than 3, we can ignore the log term, and get

$$\begin{aligned}
\|\nabla u - \nabla v\|_{L^1(\mu)} &\leq C_{p,\mu,d} \|\nabla u - \nabla v\|_{L^2(\sigma)}^{2/(p+1)} \\
&\leq C_{p,\mu,d} \left(\int_X (u-v) d(\mu - \nu) \right)^{1/(p+1)}.
\end{aligned}$$

$\square$

Proof of theorem 26. By proposition 21 we conclude that the integral

$$\int_{\Omega_t} d(x, \partial\Omega_t)^{-\alpha} dx,$$

stays bounded as $t \to 0$, for a fixed $0 \le \alpha < 1$. On Ω_t, μ is bounded from below by mt^δ; outside of Ω_t, μ is bounded from above by Mt^δ. Therefore, using similar reasoning to the previous proofs we get that for any $f \in L^\infty$, we have

$$\|f\|_{L^1(\mu)} = \|f\|_{L^1(\mu,\Omega_t)} + \|f\|_{L^1(\mu,\Omega\setminus\Omega_t)} \le C_{p,\mu}\left(\frac{\|f\|_{L^p(\sigma)}}{t^{\delta/p}} + t^\delta\right).$$

And optimizing for $s := t^\delta$ and substituting $f := |\nabla u - \nabla v|$ as in the previous proofs, we get the result. $\square$

5 An Application of Stability

In this section, we discuss an application of the stability for potentials in the $p = 1$ case to what is called the convolution inequality. Let μ, ν be two probability measures, and ρ some convolution kernel. We define the convolution $\mu * \rho$ as

$$\mu * \rho(A) = \int_{\mathbb{R}^n} \mu^z(A)\rho(z)dz,$$

where μ^z is defined as

$$\mu^z(A) := \mu(A - z).$$

We will define $\rho_\varepsilon = \rho(\cdot/\varepsilon)$, so that as $\varepsilon \to 0$, ρ_ε turns into a Dirac mass. We define $\mu_\varepsilon := \mu * \rho_\varepsilon$. The convolution inequality arises from studying the Wasserstein distance between $\mu_\varepsilon, \nu_\varepsilon$ - see, for example, [4], where the author discusses the relation between similar concepts on a Riemannian manifold and the geometry of the Riemannian manifold. It turns out, that we have $W_p(\mu_\varepsilon, \nu_\varepsilon) \le W_p(\mu, \nu)$. For $p = 1$, this is because, if u is a Kantorovich potential associated with μ, ν, then $u^z = u(\cdot - z)$ is a Kantorovich potential associated with μ^z, ν^z, which implies that $W_1(\mu^z, \nu^z) = W_1(\mu, \nu)$. Letting u^ε be the Kantorovich potential associated with $\mu_\varepsilon, \nu_\varepsilon$, using Kantorovich duality, we get

$$\begin{aligned} W_1(\mu,\nu) &= \int_{\mathbb{R}^n} W_1(\mu^z,\nu^z)\rho(z)dz \ge \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} u_\varepsilon d(\mu^z - \nu^z)\rho_\varepsilon(z)dz \\ &= \int_{\mathbb{R}^n} u_\varepsilon d(\mu_\varepsilon - \nu_\varepsilon) = W_1(\mu_\varepsilon,\nu_\varepsilon), \end{aligned} \tag{56}$$

and in fact, a similar logic holds for $p > 1$ as well, though we will focus on the case $p = 1$ from now on. There are two natural questions that arise here - first, what are the equality cases, and second, whether we can then prove some stable version of this theorem, so that if the difference $W_1(\mu,\nu) - W_1(\mu_\varepsilon,\nu_\varepsilon)$ is small, then μ, ν are close to the aforementioned equality cases. Note that as $\varepsilon \to 0$, and ρ_ε converges do a Dirac mass, $\mu_\varepsilon, \nu_\varepsilon$ get closer to μ, ν, so we would expect stability to degenerate as $\varepsilon \to 0$. The question one might ask is, how fast does it degenerate in this case?

Suppose μ, ν achieve equality in (56). This means that u is a Kantorovich potential for μ^z, ν^z for almost all z, so that $u(x - z) = u(x) - C_z$ for almost all x, z. This implies $\nabla u(x - z) = \nabla u(x)$, i.e. ∇u is constant. In this case u can be chosen to be of the form $u(x) = \langle x, e\rangle$ for some $e \in \mathbb{R}^n, \|e\| = 1$. On the other hand, if $u(x) = \langle x, e\rangle$, then $u(x - z) = u(x) - \langle z, e\rangle$, so that we have $W_1(\mu,\nu) = W_1(\mu_\varepsilon,\nu_\varepsilon)$ and (μ,ν) achieve equality. For an equivalent definition, define p_e to be the projection to the hyperplane orthogonal to e.

Proposition 29. *(μ, ν) achieve equality if and only if for some $e \in \mathbb{R}^n$, we have $\eta = p_e\#\mu = p_e\#\nu$, and writing the disintegration $\mu = \mu_y \otimes \eta, \nu = \nu_y \otimes \eta$, μ_y is stochastically dominated by ν_y for almosy all y.*

We note that the concept of stochastic dominance has some various applications to economics, see for example [10].

Proof. First suppose that (μ, ν) are an equality-achieving pair. Let T be an optimal transport map from μ to ν. By definition of potentials, we have $\langle T(x) - x, e\rangle = u(T(x)) - u(x) = |T(x) - x|$, i.e. $T(x) - x$ is a positive multiple of e. and therefore $p_e(T(x) - x) = p_e(e) = 0$, or $p_e(x) = p_e(T(x))$, which can be written as $p_e = p_e \circ T$.

We can pick T to be increasing on each $p_e^{-1}(y)$, in which case $T|_{p^{-1}(y)}$ is an optimal transport map from μ_y to ν_y for almost all y. We also have $T(x) \geq x$ for all x. For any $a \in p^{-1}(y)$:

$$\begin{aligned}\mathbb{E}_{\nu_y}[1_{x\leq a}] &= \int_{p^{-1}(y)} 1_{x\leq a} d\nu_y = \int_{p^{-1}(y)} 1_{T(x)\leq a} d\mu_y \\ &\leq \int_{p^{-1}(y)} 1_{x\leq a} d\mu_y = \mathbb{E}_{\mu_y}[1_{x\leq a}],\end{aligned}$$

so that μ_y is stochastically dominated by ν_y.

For the other direction, let T be glued from monotone maps T_y form μ_y to ν_y. By stochastic domination we have $T(x) \geq x$, so $T(x) - x$ is pointing in direction e. This implies that $u(x) = \langle x, e\rangle$ is a potential, so (μ, ν) is a an equality-achieving pair. ☐

An equivalent characterization to the disintegration condition is the following: $\mathbb{E}_\mu[\varphi] \leq \mathbb{E}_\nu[\varphi]$, whenever φ is monotone along direction e, i.e. $\varphi(x + te) \geq \varphi(x)$ for every $x \in \mathbb{R}^n, t \geq 0$. If μ_y is stochastically dominated by ν_y for all y, then this is clear because we would have $\mathbb{E}_{\mu_y}[\varphi] \leq \mathbb{E}_{\nu_y}[\varphi]$ for almost all y. On the other hand, if μ_y is not stochastically dominated by ν_y for almost all y, we can pick monotone φ_y so that $\mathbb{E}_{\mu_y}[\varphi_y] > \mathbb{E}_{\nu_y}[\varphi_y]$ if possible, and $\varphi_y = 0$ otherwise; then gluing the φ_y-s together into φ, we find that $\mathbb{E}_\mu[\varphi] > \mathbb{E}_\nu[\varphi]$.

Using a result by Michael Goldman's, and the stability result proved in the previous section, we can get the following corollary.

Corollary 4. *Assume the same assumptions as theorem 23, and also assume that the convolution kernel ρ satisfies $\rho \geq a > 0$ on $B_1(0)$. If we write $\delta := W_1(\mu, \nu) - W_1(\mu_\varepsilon, \nu_\varepsilon)$, then there exists $e \in \mathbb{R}^n, \|e\| = 1$ such that*

$$\int_X |\nabla u - e| d\mu \leq C_{X,\mu,\alpha,\rho}\varepsilon^{-1}\delta^\alpha,$$

for any $0 < \alpha < 1/3$.

Proof. We will provide a sketch of the proof, omitting some technical details. We begin by expanding the definition of δ. Recalling that u^z is the potential between μ^z and ν^z, we find

$$\begin{aligned}
\delta = W_1(\mu,\nu) - W_1(\mu_\varepsilon,\nu_\varepsilon) &= \int_{\mathbb{R}^n} W_1(\mu^z,\nu^z)\rho_\varepsilon(z)dz - W_1(\mu_\varepsilon,\nu_\varepsilon) \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} u^z d(\mu^z-\nu^z)\rho_\varepsilon(z)dz - \int_{\mathbb{R}^n} u_\varepsilon d(\mu_\varepsilon - \nu_\varepsilon) \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} u^z d(\mu^z-\nu^z)\rho_\varepsilon(z)dz - \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} u_\varepsilon d(\mu^z-\nu^z)\rho_\varepsilon(z)dz \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} (u^z - u_\varepsilon) d(\mu^z-\nu^z)\rho_\varepsilon(z)dz.
\end{aligned}$$

Let $0 \le \alpha < 1/3$. We can use theorem 23 on the inner integral, and combining with Jensen's inequality, we conclude that

$$\begin{aligned}
C_{X,\mu,\alpha}\delta^\alpha &\ge \int_{\mathbb{R}^n}\left(C_{X,\mu,\alpha}\int_{\mathbb{R}^n}(u^z-u_\varepsilon)d(\mu^z-\nu^z)\right)^\alpha \rho_\varepsilon(z)dz \\
&\ge \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |\nabla u(x-z) - \nabla u_\varepsilon(x)| d\mu^z(x)\rho_\varepsilon(z)dz \qquad (57) \\
&= \int_{\mathbb{R}^n\times\mathbb{R}^n} |\nabla u(x) - \nabla u_\varepsilon(y)|\rho_\varepsilon(x-y)d\mu(x)dy.
\end{aligned}$$

To make sense of the last integral, recall that we assumed that $\rho(x) \ge a$ for $|x| \le 1$. Therefore, $\rho_\varepsilon(x-y) \ge a$ when $|x-y| \le \varepsilon$. Careful analysis of the integral then the leads to the following bound:

$$C_{\mu,\rho}\int_{\mathbb{R}^n\times\mathbb{R}^n} |\nabla u(x) - \nabla u_\varepsilon(y)|\rho_\varepsilon(x-y)d\mu(x)dy \ge \varepsilon\int_{\mathbb{R}^n\times\mathbb{R}^n}|\nabla u(x) - \nabla u(y)|d\mu(x)d\mu(y),$$

and combined with (57), we get

$$\int_{\mathbb{R}^n\times\mathbb{R}^n}|\nabla u(x) - \nabla u(y)|d\mu(x)d\mu(y) \le C_{X,\mu,\alpha,\rho}\varepsilon^{-1}\delta^\alpha,$$

As a final step, this implies the existence of $y \in \mathbb{R}^n$ such that

$$\int_{\mathbb{R}^n}|\nabla u(x) - \nabla u(y)|d\mu(x) \le C_{X,\mu,\alpha,\rho}\varepsilon^{-1}\delta^\alpha,$$

by basic properties of the integral. Denoting $e := \nabla u(y)$, we are done. □

With that, we can prove the following.

Proposition 30. *Suppose the Kantorovich potential u for (μ,ν) satisfies $\|\nabla u - e\|_{L^1(\mu)} \le \delta$. Then $W_1(p_e\#\mu, p_e\#\nu) \le Cd\delta$, where $d = \operatorname{diam}(X)$ as before.*

Proof. Let X' be the union of all transport rays perpendicular to e. Then if $x \in X'$, we have $|\nabla u(x) - e| = \sqrt{2}$, and hence by Markov's inequality $\mu(X') \le \delta/\sqrt{2}$, and in addition

$$\nu(X') = \mu(T^{-1}(X')) = \mu(X') \le \frac{\delta}{\sqrt{2}}.$$

On $X\backslash X'$, we let $p_u(x)$ be the point of intersection of the line passing through x in direction ∇u with H. If θ is the angle between ∇u and e, then

$$\cos\theta = \langle \nabla u, e\rangle = 1 - \frac{|\nabla u - e|^2}{2},$$

so

$$\begin{aligned}|p_u(x) - p_e(x)| &= |p_u(x) - x|\sin\theta \le d\sin\theta = d\sqrt{1-\cos^2\theta} \le d\sqrt{2(1-\cos\theta)} \\ &= d|\nabla u(x) - e|.\end{aligned}$$

We will switch to a probabilistic interpretation. Let M, N be random variables with laws $\mu|_{X\backslash X'}, \nu|_{X\backslash X'}$, and coupled according to an optimal transport plan, so that $p_u(M) = p_u(N)$. We have that

$$\begin{aligned}|p_e(M) - p_e(N)| &\le |p_u(M) - p_e(M)| + |p_u(N) - p_e(N)| \\ &\le d(|\nabla u(M) - e| + |\nabla u(N) - e|),\end{aligned}$$

and so

$$W_1(p_e\#\mu|_{X\backslash X'}, p_e\#\nu|_{X\backslash X'}) \le d(\|\nabla u(x) - e\|_{L^1(\mu)} + \|\nabla u(x) - e\|_{L^1(\nu)}) = 2d\delta.$$

Combining with what we saw about X', we end up with

$$W_1(p_e\#\mu, p_e\#\nu) \le Cd\delta.$$

□

We next prove the following

Proposition 31. *Given the same assumption as the previous proposition, then*

$$\mathbb{E}_\mu[\varphi] \le \mathbb{E}_\nu[\varphi] - Cd\delta,$$

whenever φ is 1-Lipschitz and monotone along direction e.

Proof. Let $\tilde{T}(x)$ denote the orthogonal projection of $T(x)$ onto the line spanned by e. If $|\nabla u(x) - e| \le \sqrt{2}$, then $\tilde{T}(x) - x$ is a positive multiple of e, so $\varphi(\tilde{T}(x)) \ge \varphi(x)$. As in the previous section, the distance between $T(x)$ and $\tilde{T}(x)$ is bounded by $Cd|\nabla u(x) - e|$, so we get

$$\varphi(T(x)) \ge \varphi(x) - Cd|\nabla u(x) - e|.$$

The set $A = \{x : |\nabla u(x) - e| \ge \sqrt{2}\}$ satisfies $\mu(A) = \nu(A) \le \frac{\delta}{\sqrt{2}}$ as in the previous proposition, so altogether we can write

$$\mathbb{E}_\nu[\varphi] = \mathbb{E}_\mu[\varphi \circ T] \ge \mathbb{E}_\mu[\varphi - Cd|\nabla u - e] = \mathbb{E}[\varphi(M)] - Cd\delta.$$

□

Combining all the we proved, we have the following corollary.

Corollary 5. *Assume the same assumptions as in Corollary 4, and write $\delta = W_1(\mu,\nu) - W_1(\mu_\varepsilon, \nu_\varepsilon)$. Then for every $\alpha < 1$, there exists a constant $C = C_{X,\mu,\alpha}$ such that*

$$W_1(p_e\#\mu, p_e\#\nu) \le C\varepsilon^{-1}\delta^\alpha,$$

and

$$\mathbb{E}_\mu[\varphi] \le \mathbb{E}_\nu[\varphi] - C\varepsilon^{-1}\delta^\alpha,$$

whenever φ is 1-Lipschitz and monotone along e.

References

[1] Luigi Ambrosio. *Lecture Notes on Optimal Transport Problems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.

[3] Haïm Brezis and Petru Mironescu. Gagliardo–nirenberg inequalities and non-inequalities: The full story. *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 35(5):1355–1376, 2018.

[4] Hong-Bin Chen and Jonathan Niles-Weed. Asymptotics of smoothed wasserstein distances. *Potential Analysis*, pages 1–25, 2021.

[5] Alex Delalande and Quentin Mérigot. Quantitative stability of optimal transport maps under variations of the target measure. *Duke Mathematical Journal*, 172(17):3321 – 3357, 2023.

[6] H. Federer. *Geometric Measure Theory*. Classics in Mathematics. Springer Berlin Heidelberg, 2014.

[7] Peter Lindqvist. *Notes on the stationary p-Laplace equation*. Springer, 2019.

[8] Octave Mischler and Dario Trevisan. Quantitative stability in optimal transport for general power costs, 2025.

[9] Marcel Nutz. Introduction to entropic optimal transport. *Lecture notes, Columbia University*, 2021.

[10] Luciano Pomatto, Philipp Strack, and Omer Tamuz. Stochastic dominance under independent noise. *Journal of Political Economy*, 128(5):1877–1900, May 2020.

[11] Serik Sagitov. Weak convergence of probability measures. *arXiv preprint arXiv:2007.10293*, 2020.

[12] Filippo Santambrogio. *Optimal transport for applied mathematicians*, volume 55. Springer, 2015.

[13] C. Villani. *Topics in Optimal Transportation*. Graduate studies in mathematics. American Mathematical Society, 2003.