UNIVERSITAT DE BARCELONA

ADVANCED MATHEMATICS
MASTER'S FINAL PROJECT

The Optimal Transport Problem And Its Applications

Author:
Pablo García Arias

Supervisor:
Joaquim Ortega Cerdà

Facultat de Matemàtiques i Informàtica

June 28, 2023

Acknowledgments

I would like to start by thanking Joaquim Ortega Cerdà for his great help, without which I would not have completed this project. I would like to mention too the professors of *Universitat de Barcelona*, from whom I had the pleasure to learn during this Master.

A special thanks to my family and friends, for their continuous support and encouragement during all these years.

Finally, to Pedro, who had my back at the difficult times.

Summary

The objective of this project is to present the base of Optimal Transport Theory and some of its applications. The Optimal Transport Problem was first studied by Monge in the 18th century, and later reformulated by Kantorovich during the 20th century, being this second version the main object of study. One of the key results relating Monge's formulation is Brenier's theorem, which we will prove and apply to prove the Isoperimetric inequality and the Sobolev inequality. By employing a different method we will prove another classical result, the Brunn-Minkowski inequality. This essay concludes with some conditions for the two problems to have the same optimal value. The other main topic studied during this work are the Wasserstein spaces. They are a family of probability measures spaces where we use Optimal transport to construct a metric, the Wasserstein distance. A key result is that it metrizes the weak topology of these spaces.

Contents

Introduction

The study of the Optimal Transport Problem is a topic in Analysis that recently is earning the interest of more mathematicians. This is partly due to the deep connections that are being found between this theory and topics like Partial Differential Equations and the Ricci flow in Differential Geometry. It is also starting to be applied to problems like machine learning or image interpolating.

In this work I will present the basics of the Optimal Transport theory, how it can be applied to easily prove some classical theorems, and the concept of Wasserstein distance. The main references have been [1], [4] and [11]. The structure is as follows:

Chapters 1 and 2 provide the theoretical backbone of this essay. In the first chapter I introduce the Optimal Transport Problem and its different formulations, including its duality. The main theorems will be about characterizing the solutions of the problem. In Chapter 2 it is developed the Wasserstein distance, how it is a metric in space of probability distributions, and that it metrizes the weak topology.

Regarding the applications, I have included three distinct flavors. In Chapter 3 it is used the theory of the first chapter, with a notable mention to Brenier's theorem, to prove the Isoperimetric inequality and the Sobolev inequality. In Chapter 4 it is proved the Brunn-Minkowski inequality, for which is needed a brief introduction to the study of convex functionals over the Wasserstein space W_2.

Chapter 5 is devoted to more recent applications of the Wasserstein to Kronecker sequences, for which I have followed the paper of [12].

Finally, Chapter 6 proves how Kantorovich's formulation of the Optimal Transport problem is a relaxation of Monge's in compact subsets of $\mathbb{R}^d$.

Chapter 1
The Optimal Transport Problem

We start by considering (X, d) a Polish space, this is, a metric space that is both complete and separable. The used notation is as follows:

- $\mathcal{P}(X)$ is the set of Borel probability measures on X.
- For $\mu \in \mathcal{P}(X)$, its support $\operatorname{supp} \mu$ is the set of all points $x \in X$ such that for any open neighborhood N_x we have that $\mu(N_x) > 0$.

During this whole essay X, Y will be two Polish spaces. For a Borel map $T : X \to Y$, we define the **push forward** of $\mu \in \mathcal{P}(X)$ as $T_{\#\mu} \in \mathcal{P}(Y)$ given by $T_{\#\mu}(E) = \mu(T^{-1}(E))$ for each Borel set $E \subseteq Y$. This is characterized by

$$\int f \, dT_{\#\mu} = \int f \circ T \, d\mu, \qquad \text{for all } f : Y \to \mathbb{R} \cup \{\pm\infty\} \text{ measurable Borel.}$$

Monge published "*Mémoire sur la théorie des déblais et des remblais*" in 1781, where he first proposed a problem that has led to the field of Optimal Transport Theory. He was interested in minimizing the effort of moving some resources that has te be extracted in some places and transported to some destinations. We will consider a more general version here.

Problem 1.1 **Monge's Optimal Transport Problem**

Given a cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ measurable Borel, and measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, we want to minimize

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

for all T **transport maps**, this is $T : X \to Y$ with $T_{\#\mu} = \nu$.

It is worth commenting that most of the time we are not only interested on the minimal value, but also on the map that optimizes it (if it exists) and its properties.

This problem has some phenomenons that are not desirable for its study. For example, there may not exist any transport map T. To illustrate this it is enough to consider $\mu = \delta_x$ some Dirac delta and ν not being one. Then we have that

$$T_{\#\mu}(B) = \delta_x\left(T^{-1}(B)\right) = \left\{\begin{array}{ll} 1 & \text{if } x \in T^{-1}(B) \\ 0 & \text{if } x \notin T^{-1}(B) \end{array}\right\} = \delta_{T(x)}(B) \not\equiv \nu$$

Another issue is that the constraint $T_{\#\mu} = \nu$ is not closed with respect to any reasonable topology. Take for example $f : \mathbb{R} \to \mathbb{R}$ one periodic, with $f\big|_{[0,1/2)} = 1$ and $f\big|_{[1/2,1)} = -1$. We define then

$$\mu = \mathcal{L}\big|_{[0,1]}, \qquad \nu = \frac{\delta_1 + \delta_{-1}}{2}, \qquad f_n(x) = f(nx).$$

It immediately follows that $(f_n)_{\#\mu} = \nu$. We are going to prove that $f_n \rightharpoonup 0$ weakly, which, as $0_{\#\mu} = \delta_0 \neq \nu$ finishes the example.

Consider first any function $g \in C_c(\mathbb{R})$, we want to prove that $\int f_n g \to 0$. As the support is compact, we have $\operatorname{supp} g \subseteq [-M, M]$. Note that we can write

$$f_n|_{[-M,M]} = \begin{cases} 1 & \text{on } \left[\frac{k}{n}, \frac{k+\frac{1}{2}}{n}\right) \\ -1 & \text{on } \left[\frac{k+\frac{1}{2}}{n}, \frac{k+1}{n}\right) \end{cases} \quad k = -Mn, -Mn+1, \dots, -1, 0, 1, \dots, Mn-1.$$

As the function g is continous and with compact support, it is uniformly continous, so there is some $\delta > 0$ such that $|x - y| < \delta \Rightarrow |g(x) - g(y)| < \frac{\varepsilon}{M}$. Then we can find a N such that for any $n \geq N$, $\frac{1}{2n} < \delta$ and therefore

$$\left|\int_{-M}^{M} f_n g\right| \leq \sum_{k=-Mn}^{Mn-1} \int_{x=\frac{k}{n}}^{\frac{k+\frac{1}{2}}{n}} \left|g(x) - g(x + \frac{1}{2n})\right| < 2Mn \cdot \frac{1}{2n} \cdot \frac{\varepsilon}{M} = \varepsilon.$$

Kantorovich was a 20th century mathematician who contributed to many different research topics. He is most known as being one of the founder of the modern linear optimization field and being one of the Nobel Economy Prize winner of 1975. In the field of Optimal Transport, he proposed a relaxation of the problem that avoids the complications previous discussed.

Problem 1.2 **Kantorovich's Optimal Transport Problem**

For $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$, and measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, we want to minimize

$$\gamma \longmapsto \int_{X\times Y} c(x,y)\, d\gamma$$

for γ a **transport plan** in

$$\mathrm{Adm}(\mu,\nu) = \left\{ \gamma \in \mathcal{P}(X\times Y) : \begin{array}{l} \gamma(A\times Y) = \mu(A)\ \forall A \subseteq X \text{ Borel} \\ \gamma(X\times B) = \nu(B)\ \forall B \subseteq Y \text{ Borel} \end{array} \right\}.$$

Note that the conditions for $\gamma \in \mathrm{Adm}(\mu,\nu)$ are equivalent to the first marginal being $\pi^X_{\#\gamma} = \mu$ and the second $\pi^Y_{\#\gamma} = \nu$.

Transport plans are a generalization of transport maps, in the sense that if we have $T : X \to Y$ a transport map from μ to ν, then

$$\begin{aligned} (Id, T) : X &\longrightarrow X \times Y \\ x &\longmapsto (x, T(x)) \end{aligned}$$

induces the transport plan $\gamma = \gamma_T := (Id, T)_{\#\mu}$:

- For $A \subseteq X$ Borel, $(Id,T)^{-1}(A\times Y) = A$, so $\gamma(A\times Y) = \int_{A\times Y} d\gamma = \int_A d\mu = \mu(A)$.
- For $B \subseteq Y$ Borel, $T^{-1}(B) = (Id,T)^{-1}(X\times B)$, so using that $T_{\#\mu} = \nu$

$$\gamma(X\times B) = \int_{X\times B} d\gamma = \int_{T^{-1}(B)} d\mu = \int_B d\nu = \nu(B).$$

During this section we are going to see that Kantorovich's formulation has several advantages over Monge's that makes its study more suitable. Notably:

- There is always at least one transport map, as $\mu \times \nu \in \mathrm{Adm}(\mu,\nu)$.
- The set $\mathrm{Adm}(\mu,\nu)$ is convex and compact with respect to the narrow topology, defined in the following section.
- Under some mild regularity hypothesis over the cost function, there are minimizers and they are nicely characterized.
- When the cost function is continuous and the first measure μ is atomless, $\inf(\text{Monge}) = \min(\text{Kantorovich})$[1], so it makes sense to study this nicer problem.

[1] This theorem can be consulted on [10].

1.1 Conditions For Optimality

Definition 1 **Narrow Topology**

We say a sequence of measures $(\mu_n) \subseteq \mathcal{P}(X)$ narrowly converges to $\mu \in \mathcal{P}(X)$ when

$$\int \varphi \, d\mu_n \to \int \varphi \, d\mu \qquad \forall\varphi \text{ continuous and bounded.}$$

This is sometimes written as $\mu_n \xrightarrow{\text{nrw}} \mu$.
A set $\mathcal{A} \subseteq \mathcal{P}(X)$ is called **tight** when for each $\varepsilon > 0$ exists some compact set $K_\varepsilon \subseteq X$ such that $\mu(X \backslash K_\varepsilon) \leq \varepsilon$ for all $\mu \in \mathcal{A}$.

It can be shown that this topology is metrizable. Moreover, the following two results are well known in the study of Polish spaces. Let X be a complete separable metric space (this is, a Polish space), then:

- (*Prokhorov's theorem*) A family $\mathcal{K} \subseteq \mathcal{P}(X)$ is relatively compact w.r.t. the narrow topology if and only if $\mathcal{K}$ is tight.
- (*Ulam's tightness theorem*) A finite Borel measure μ on X is tight, in the sense that $\{\mu\}$ is a tight set. This means that

$$\forall\varepsilon > 0 \; \exists K \text{ compact set such that } \mu(X \backslash K) < \varepsilon.$$

More details about this results are available at [13], theorems 5.2 and 2.6.

Remark 1.3. *Because any $\gamma \in \mathrm{Adm}(\mu, \nu)$ satisfies the inequality*

$$\gamma\left(X \times Y \backslash K_1 \times K_2\right) \leq \gamma\left((X \backslash K_1) \times Y\right) + \gamma\left(X \times (Y \backslash K_2)\right) = \mu(X \backslash K_1) + \nu(Y \backslash K_2)$$

we know that, if $\mathcal{K}_1 \subseteq \mathcal{P}(X)$ and $\mathcal{K}_2 \subseteq \mathcal{P}(Y)$ are tight, then so is the set

$$\left\{\gamma \in \mathcal{P}(X \times Y) \;:\; \pi^X_{\#\gamma} \in \mathcal{K}_1, \; \pi^Y_{\#\gamma} = \nu\right\}.$$

Remark 1.4. We are going to use extensively the properties of lower and upper semicontinous functions, so it will be useful for the reader to recall the following basic properties of them.

Given M a metric space and a function $f : X \to \overline{\mathbb{R}}$ we say that f is lower semicontinous (l.s.c.) at the point $x_0 \in X$ if for every $y \in \mathbb{R}$ with $y < f(x_0)$ exists $U = U(y)$ a neighborhood of x_0 such that $f(x) > y\ \forall x \in U$. The function f is l.s.c. if it is l.s.c. at every point.

The concept of upper semicontinous (u.s.c.) is defined in a complete symmetric way and is characterized by f is u.s.c. if and only if $-f$ is l.s.c.

- If the functions $(f_i)_{i\in I}$ are all l.s.c. then the pointwise supremum $\sup_{i\in I} f_i$ is l.s.c.
- Notably, for a monotone increasing sequence $f_n : X \to \mathbb{R}$ of continuous functions then $\lim_n f_n = \sup_n f_n$ is l.s.c.
- (*Theorem of Baire*) If $f : X \to \overline{\mathbb{R}}$ is l.s.c. then exists a monotone increasing sequence $f_n : X \to \overline{\mathbb{R}}$ of continuous functions with $f_n \to f$. If $f : X \to \mathbb{R} \cup \{+\infty\}$ then $f_n : X \to \mathbb{R}$.

Theorem 1.5 Existence of minimizers for the Kantorovich's formulation

If c is lower semicontinous and bounded from below, then there is a minimizer for the Kantorovich's Problem 1.2.

Proof. First, by Ulam's Theorem we know that the sets $\{\mu\}$ and $\{\nu\}$ are tight, so the previous remark tell us that $\mathrm{Adm}(\mu,\nu)$ is tight. Using Prokhorov's theorem, $\mathrm{Adm}(\mu,\nu)$ is relatively compact (in the narrow topology). We claim that it is in fact compact. Let $(\gamma_n) \subseteq \mathrm{Adm}(\mu,\nu)$ and $\gamma_n \to \gamma$ narrowly, we want to prove that $\gamma \in \mathrm{Adm}(\mu,\nu)$. For any $\varphi \in C(X)$ bounded, we know that $(x,y) \mapsto \varphi(x)$ is a continuous and bounded function on $X \times Y \to \mathbb{R}$, so using the narrow convergence

$$\int \varphi\, d\pi^X_{\#\gamma} = \int \varphi(x)\, d\gamma = \lim_{n\to\infty} \int \varphi(x)\, d\gamma_n = \lim_{n\to\infty} \int \gamma\, d\pi^X_{\#\gamma_n} = \int \varphi\, d\mu.$$

As φ was arbitrary, we get that $\pi^X_{\#\gamma} = \mu$. The same idea works for ν, so $\gamma \in \mathrm{Adm}(\mu,\nu)$.

Now, we want to prove that the functional $\gamma \mapsto \int c\ d\gamma$ is lower semicontinous in the narrow topology. We know there is an increasing sequence of continuous functions c_n with $c = \sup_n c_n$. Because there is a lower bound $c \geq L > -\infty$ we can assume that $c_n \geq L$ for all n too. We can also assume that every c_n is bounded, as long as we don't require the upper bound to be the same for each c_n.

Using the monotone convergence theorem, we get $\int c - L\ d\gamma = \sup_n \int c_n - L\ d\gamma$. But because γ is a probability, $\int L\ d\gamma = L$ and we get

$$\int c\ d\gamma = \sup_n \int c_n\ d\gamma.$$

Note that because each c_n is bounded, then $\gamma \mapsto \int c_n\ d\gamma$ is narrowly continuous. This makes then $\gamma \mapsto \int c\ d\gamma$ a supremum of continuous functions, so it is l.s.c.

Finally, as $\mathrm{Adm}(\mu,\nu)$ is compact and $\gamma \mapsto \int c d\gamma$ is l.s.c. it attains a minimum. $\square$

Definition 2 **Optimal plans**

The set $\mathrm{Opt}(\mu, \nu)$ denotes set of all **optimal plans** from μ to ν, this is, the set of plans that minimize the Kantorovich Problem. We will also say that a plan is optimal if it is optimal with respect to its marginals.

Note that this notation does not specify which is the cost function. This choice of c will be clear from the context, as this function is usually fixed.

To tackle the problem of when a plan is optimal, we need first to introduce some notions.

Definition 3 **c-transforms, c-concavity and c-convexity**

For $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ we define its $c+$ **transform** as

$$\psi^{c+} : X \to \mathbb{R} \cup \{-\infty\} \qquad \psi^{c+}(x) = \inf_{y \in Y} c(x, y) - \psi(y).$$

Analogously, for $\varphi : X \to \mathbb{R} \to \cup\{\pm\infty\}$ we define

$$\varphi^{c+} : Y \to \mathbb{R} \cup \{-\infty\} \qquad \varphi^{c+}(y) = \inf_{x \in X} c(x, y) - \varphi(x).$$

The $c-$ **transforms** are given by

$$\begin{aligned} \psi^{c-} &: X \to \mathbb{R} \cup \{-\infty\} & \psi^{c-}(x) &= \sup_{y \in Y} -c(x, y) - \psi(y) = -\inf c(x, y) + \psi(y) \\ \varphi^{c-} &: Y \to \mathbb{R} \cup \{-\infty\} & \varphi^{c-}(y) &= \sup_{x \in X} -c(x, y) - \varphi(x) = -\inf c(x, y) + \varphi(x) \end{aligned}$$

- $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is c-**concave** if there is some $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ with $\varphi = \psi^{c+}$.
- $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ is c-**concave** if there is some $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ with $\psi = \varphi^{c+}$.
- $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ is c-**convex** if there is some $\psi : Y \to \mathbb{R} \cup \{+\infty\}$ with $\varphi = \psi^{c-}$.
- $\psi : Y \to \mathbb{R} \cup \{+\infty\}$ is c-**convex** if there is some $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ with $\psi = \varphi^{c-}$.

For $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ c-concave, we define its c-**superdifferential** as

$$\partial^{c+}\varphi = \left\{(x, y) \in X \times Y \ : \ \varphi(x) + \varphi^{c+}(y) = c(x, y)\right\}.$$

For $x \in X$, we also denote $\partial^{c+}\varphi(x) = \{y \in Y \ : \ (x, y) \in \partial^{c+}\varphi\}$.

Given $\varphi : X \to \mathbb{R} \cup \{+\infty\}$ c-convex, we define its **c-subdifferential** as

$$\partial^{c-}\varphi = \left\{(x,y) \in X \times Y \ : \ \varphi(x) + \varphi^{c-}(y) = -c(x,y)\right\}.$$

In the case of c-convex and c-concave functions on Y we use the natural symmetric definitions.

I want first to highlight that there are some trivial connections between these concepts, like that a function φ is c-convex if and only if $-\varphi$ is c-concave or that $-\varphi^{c+} = (-\varphi)^{c-}$. This means that most statements regarding one concept have an analogous one for the other. To get rid off this redundancy, in this work I will focus mainly on the c-concave results.

One interesting property is that for $\psi : Y \to \mathbb{R} \cup \{\pm\infty\}$ we have that $\psi^{c+} = \psi^{c+c+c+}$. Indeed,

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

so taking $\tilde{x} = x$ we get $\psi^{c+c+c+} \geq \psi^{c+}$, and taking $y = \tilde{y}$ gives the other inequality.
From this follows that $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ is c-concave if and only if $\varphi^{c+c+} = \varphi$. Naturally, this also works for functions ψ on Y.

This has a nice implication. Consider a c-concave function $\varphi : X \to \mathbb{R}$, then

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

means that $\varphi(x) + \varphi^{c+}(y) \leq c(x,y)$ for all $x \in X$, $y \in Y$.

If we apply this to the c-supperdiferential, then

$$\varphi(x) + \varphi^{c+}(y) = c(x,y) \xLeftrightarrow{\varphi(z)+\varphi^{c+}(y)\leq c(z,y)} \varphi(z) - c(z,y) \leq \varphi(x) - c(x,y) \quad \forall z \in X.$$

The following proposition collects all these basic properties.

Proposition 1.6.

1. *A function is c-concave iff and only if $\varphi = \varphi^{c+c+}$.*
2. *For a c-concave function,*
$$\varphi(x) + \varphi^{c+}(y) \leq c(x,y) \qquad\qquad \forall x \in X,\ \forall y \in Y.$$
3. *For a c-concave function,*
$$(x,y) \in \partial^{c+}\varphi \iff \varphi(z) - c(z,y) \leq \varphi(x) - c(x,y) \qquad \forall z \in X.$$

Example 1.7. *If we consider $X = Y = \mathbb{R}^d$ and $c(x, y) = -\langle x, y\rangle$, we get that this notions are the standard ones in convex analysis:*

- *A function is c-convex (or c-concave) when it is convex (resp. concave).*
- *The c-superdifferential and c-subdifferential are the classical superdifferential and subdifferential.*
- *The $c-$ transform is the Legendre transform.*

Indeed, for the first part we need to recall the following not so well-known theorem:

Theorem 1.8. *A function $f : A \subseteq \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and l.s.c. if and only if $f = \sup_\alpha f_\alpha$ for a family f_α of affine functions.*

If φ is c-convex then exists a function with $\varphi(x) = \psi^{c-}(x) = \sup_y \langle x, y\rangle - \psi(y)$. But for any fixed y the function $\langle x, y\rangle - \psi(y)$ is affine, so φ is convex.

For the other implication, a convex function is continuous in the interior of its domain, so if $\varphi(x) = \sup_\alpha f_\alpha(x)$ for f_α affine functions we want to see that it is c-convex. Because the functions are affine, we can write them as $f_\alpha(x) = \langle x, y_\alpha\rangle + K_\alpha$ with $y_\alpha, K_\alpha \in \mathbb{R}^d$. Note that it can happen that $y_\alpha = y_{\tilde{\alpha}}$ but $K_\alpha \neq K_{\tilde{\alpha}}$. In this case we also have the inequality for $\sup\{K_\alpha, K_{\tilde{\alpha}}\}$ where the suppremum is taken coordinate by coordinate. We can define

$$-\psi(y) = \begin{cases} \sup\{K_\alpha \ : \ y_\alpha = y\} & \text{if exists } \alpha \text{ with } y_\alpha = \alpha \\ -\infty & \text{if } \nexists \alpha \text{ with } y_\alpha = y \end{cases}$$

Then it is clear that $\varphi(x) = \sup_y \langle x, y\rangle - \psi(y) = \psi^{c-}(x)$ is c-convex.

For the last part, the Legendre transform is the $c-$ transform *in the region where it is finite.*

Example 1.9 (Discrete measures)**.** Let's consider $X = Y = \mathbb{R}^d$ and $c(x, y) = |x - y|^2/2$ the cost function. Let's also assume that $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ are discrete measures of the form

$$\mu = \sum_{i=1}^{M} m_i \delta_{x_i}, \qquad\qquad \nu = \sum_{j=1}^{N} n_j \delta_{y_j}.$$

As any measure in $\gamma \in \mathrm{Adm}(\mu, \nu)$ is concentrated on the product of the supports, we can represent the admisible set as

$$\mathrm{Adm}(\mu, \nu) = \left\{ \sum_{\substack{i=M \\ j=N \\ i,j=1}} a_{i,j} \delta_{(x_i, y_j)} \ : \ a_{i,j} \geq 0, \ \sum_{i=1}^{M} a_{i,j} = n_j, \ \sum_{j=1}^{N} a_{i,j} = m_i \right\}$$

Let's consider now a plan $\gamma \in \mathrm{Opt}(\mu, \nu)$. Let's consider K points in $\operatorname{supp} \gamma$ and a permutation σ of $\{1, \dots, N\}$. Then we have that

$$\sum_{k=1}^{K} \frac{|x_k - y_k|^2}{2} \leq \sum_{k=1}^{K} \frac{|x_k - y_{\sigma(k)}|^2}{2}. \tag{1.1}$$

Indeed, if this is not true we can consider for $\varepsilon > 0$ small enough

$$\tilde{\gamma} = \gamma - \varepsilon \sum_{k=1}^{K} \left(\delta_{(x_k, y_{\sigma(k)})} - \delta_{(x_k, y_k)} \right)$$

would have a strictly smaller cost, contradicting the optimality of γ.

In fact, if for $\gamma \in \mathrm{Adm}(\mu, \nu)$ this condition holds true for any $N \in \mathbb{N}$, any points in the support and any permutation, it can be proved that γ is optimal. This result will also be true as a consequence of Theorem 1.10.

Note that if in (1.1) we expand the squares we get $\sum \langle x_k, y_k \rangle \geq \sum \langle x_k, y_{\sigma(k)} \rangle$. This means that $\gamma \in \mathrm{Adm}(\mu, \nu)$ is optimal when its support is a *ciclically monotone* set. This motivates us the generalize the concept using this equation.

Definition 4 **c-ciclically monotone**

A set $\Gamma \subseteq X \times Y$ is c-ciclically monotone if for any $N \in \mathbb{N}$ points $(x_i, y_i) \in \operatorname{supp}(\gamma)$ and any permutation $\sigma : \{1, \dots, N\} \to \{1, \dots, N\}$ we have

$$\sum_{n=1}^{N} c(x_i, y_i) \leq \sum_{n=1}^{N} c(x_i, y_{\sigma(i)}).$$

The previous example showed a situation where being an optimal plan was characterized by its support being c-ciclically monotone. This was not a special case, but is true in general, under some assumptions on the cost function.

Theorem 1.10 **Fundamental Theorem of Optimal Transport**

Let's assume the cost function c is continuous and bounded from below and with the upper bound

$$c(x, y) \leq a(x) + b(y)$$

for some functions $a \in L^1(X, \mu)$, $b \in L^1(Y, \nu)$. For any $\gamma \in \mathrm{Adm}(\mu, \nu)$ the following statements are equivalent:

1. γ is an optimal plan.
2. $\operatorname{supp} \gamma$ is c-cyclically monotone.
3. there exists $\varphi : X \to \mathbb{R}$, c-concave, with $\varphi \in L^1(X, \mu)$ and $\operatorname{supp} \gamma \subseteq \partial^{c+} \varphi$.
4. there exists $\varphi : X \to \mathbb{R}$, c-concave, with $\max\{\varphi, 0\} \in L^1(X, \mu)$ and $\operatorname{supp} \gamma \subseteq \partial^{c+} \varphi$.

Proof. Because of the inequality

$$\int c(x, y)\, d\gamma \leq \int a(x) + b(y)\, d\gamma = \int a(x)\, d\mu + \int b(y)\, d\nu < \infty$$

we know that $c \in L^1(\gamma)$ for any $\gamma \in \mathrm{Adm}(\mu, \nu)$. Moreover, $\max\{0, c\} \in L^1(\gamma)$ too.

1⇒2 Let's argue by contradiction. For γ a optimal plan, if $\operatorname{supp} \gamma$ is not c-ciclically monotone there are N points $(x_i, y_i) \in \operatorname{supp} \gamma$ and a permutation σ of them with

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

Because the function c is continuous, we can actually find disjoint neighborhoods with $x_i \in U_i$, $y_i \in U_i$ that keep the inequality true:

$$\sum_{i=1}^{N} c(u_i, v_{\sigma(i)}) - c(u_i, v_i) < 0, \qquad\qquad \forall u_i \in U_i,\ v_i \in V_i.$$

Note that as each $(x_i, y_i) \in \operatorname{supp} \gamma$, we can assume this sets to have non-zero γ measure.

We now want to construct a measure $\tilde{\gamma} = \gamma + \eta$ so that the optimality of γ is contradictory. For this, we want η to be a signed measure satisfying:

- $\eta^- \leq \gamma$, so $\tilde{\gamma}$ is a (positive) measure.
- null marginals and $\eta(X \times Y) = 0$, so we have $\tilde{\gamma} \in \mathrm{Adm}(\mu, \nu)$.
- $\int c\, d\eta < 0$, making γ not optimal.

To construct this measure, let's denote by $\Omega = \prod_{i=1}^N U_i \times V_i \subset (X \times Y)^N$ and the probability measure $\mathbf{P} \in \mathcal{P}(\Omega)$ given by the product of $\frac{1}{m_i}\gamma_i$ where $\gamma_i = \gamma\big|_{U_i \times V_i}$ and $m_i = \gamma(U_i \times V_i)$. With these notations, we consider

$$\eta = \frac{\min m_i}{N} \sum_{i=1}^N \left(\pi^{U_i}, \pi^{V_{\sigma(i)}}\right)_{\#\mathbf{P}} - \left(\pi^{U_i}, \pi^{V_i}\right)_{\#\mathbf{P}}$$

Note that for a set $A \times B \subseteq X \times Y$ we can give a more explicit expression,

$$\eta(A \times B) = \frac{\min m_i}{N} \sum_{i=1}^N \frac{1}{m_i m_{\sigma(i)}} \gamma_i\left((U_i \cap A) \times V_i\right) \gamma_{\sigma(i)}\left(U_{\sigma(i)} \times (V_\sigma(i) \cap B)\right) - \frac{1}{m_i}\gamma_i\left((U_i \cap A) \times (V_i \cap B)\right)$$

With both of these expression it is trivial to check that η satisfies all the wanted conditions.

4⇒1 Because $\operatorname{supp}\gamma \subseteq \partial^{c+}\varphi$ we have $\varphi(x) + \varphi^{c+}(y) = c(x,y)$ for $(x,y) \in \operatorname{supp}\gamma$. Recall that because φ is c-convex,

$$\varphi(x) + \varphi^{c+}(y) \leq c(x,y), \qquad x \in X,\ y \in Y.$$

For another admisible plan $\tilde{\gamma} \in \operatorname{Adm}(\mu,\nu)$ we want to see that $\int c\ d\gamma \leq \int c\ d\tilde{\gamma}$. But this follows from the previous relationships:

$$\int c\ d\gamma = \int \varphi(x) + \varphi^{c+}(y)\ d\gamma = \int \varphi\ d\mu + \int \varphi^{c+}\ d\nu = \int \varphi(x) + \varphi^{c+}(y)\ d\tilde{\gamma} \leq \int c(x,y)\ d\tilde{\gamma}.$$

3⇒4 This implication is clear.

2⇒3 Assuming that $\Gamma = \operatorname{supp}\gamma$ is c-ciclically monotone, a function φ with the conditions of 3 would have to satisfy the following: fix a point $(\overline{x},\overline{y}) \in \Gamma$, and consider any N points $(x_i, y_i) \in \Gamma$, then

$$\begin{aligned}
\varphi(x) &\leq c(x,y_1) - \varphi^{c+}(y_1) = c(x,y_1) - c(x_1,y_1) + \varphi(x_1) \\
&\leq c(x,y_1) - c(x_1,y_1) + c(x_1,y_2) - c(x_2,y_2) + \varphi(x_2) \\
&\vdots \\
\varphi(x) &\leq c(x,y_1) - c(x_1,y_2) + \cdots + c(x_N,\overline{y}) - c(\overline{x},\overline{y}) + \varphi(\overline{x}).
\end{aligned}$$

If φ satisfy the wanted conditions, $\varphi + k$ does too, so we can assume $\varphi(\overline{x}) = 0$, by replacing the function φ if necessary. This motivates us to define the function the following way:

$$\varphi(x) := \inf\left\{c(x,y_1) - c(x_1,y_2) \cdots + c(x_N,\overline{y}) - c(\overline{x},\overline{y})\ :\ N \in \mathbb{Z}_+,\ (x_i,y_i) \in \Gamma\ i = 1,\ldots,N\right\}$$

This is a $L^1(\mu)$ function because of the upper bound from taking $N = 1$ and $(x_1, y_1) = (\overline{x}, \overline{y})$:

$$\varphi(x) \leq c(x, \overline{y}) - c(\overline{x}, \overline{y}) + c(\overline{x}, \overline{y}) - c(\overline{x}, \overline{y}) \leq a(x) + b(\overline{y}) - c(\overline{x}, \overline{y}) \in L^1(\mu).$$

The c-concavity is easy to see, as we can write $\varphi(x) = \psi^{c+}(y) = \inf_{y \in Y} c(x, y) - \psi(y)$ for the function given by

$$-\psi(y) = \inf \{-c(x_1, y_2) \cdots + c(x_N, \overline{y}) - c(\overline{x}, \overline{y}) \ : \ N \in \mathbb{Z}_+, \ (x_1, y) \in \Gamma, \ (x_i, y_i) \in \Gamma \ i = 2, \ldots, N\}$$

Given $(x, y) \in \Gamma$, consider $N \geq 2$ and make $(x_1, y_1) = (x, y)$:

$$\varphi(z) \leq c(z, y) - c(x, y) + \inf \{c(x, y_2) - c(x_2, y_2) \cdots \} = c(z, y) - c(x, y) + \varphi(x),$$

which is equivalent to $(x, y) \in \partial^{c+}\varphi$. $\square$

Note that in the implication 1⇒2 we have no used the upper bound on c, so we have this nice corollary.

Corollary 1.11. *Given a continuous cost function bounded from below, if a plan is optimal then it has a c-cyclically monotone support.*

As a consequence of the Fundamental theorem we have the following result, that gives sufficient conditions for the stability of maps when we change the probabilities μ, ν and the cost c.

Theorem 1.12 (Stability of optimal plans). *Consider the following hypothesis:*

- *A cost function $c : X \times Y \to \mathbb{R}$ continuous and bounded from below.*
- *A sequence (c_k) of continuous cost functions with $c_k \to c$ uniformly.*
- *The measures $(\mu_k) \subseteq \mathcal{P}(X)$ with $\mu_k \xrightarrow{\text{nrw}} \mu$.*
- *$(\nu_k) \in \mathcal{P}(Y)$ with $\nu_k \xrightarrow{\text{nrw}} \nu$.*
- *The optimal plans $\gamma_k \in \mathrm{Opt}(\mu_k, \nu_k)$ such that $\int c_k \, d\gamma_k < \infty$.*

Then exists a subsequence $\gamma_{k'} \xrightarrow{\text{nrw}} \gamma \in \mathrm{Adm}(\mu, \nu)$ with c-cyclically monotone support.

Moreover, if $c(x, y) \leq a(x) + b(y)$ with $a \in L^1(X, \mu)$ and $b \in L^1(Y, \nu)$, then $\gamma \in \mathrm{Opt}(\mu, \nu)$.

Remark 1.13. We have not used it until now, but it is useful to note that a probability $\gamma \in \mathcal{P}(X \times Y)$ has c-ciclically monotone support if and only if

$$(\operatorname{supp} \gamma)^{\otimes N} \subseteq \left\{ ((x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)) \in (X \times Y)^N \ : \ \sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{\sigma(i)}) \right\}$$

for any $N \geq 1$ and any permutation $\sigma : \{1, \cdots, N\} \to \{1, \cdots, N\}$.

We are using the notation $(\operatorname{supp} \gamma)^{\otimes N} = \underbrace{\operatorname{supp} \gamma \times \cdots \times \operatorname{supp} \gamma}_{N \text{ times}}$.

Proof. Because the sequences $\{\mu_k\}$ and $\{\nu_k\}$ are convergent, they are relatively compact sets, and using Prokhorov theorem they are tight sets. Then, thanks to Remark 1.3, the sequence $\{\gamma_k\}$ is tight too and has a convergent subsequence $\gamma_k \xrightarrow{\text{nrw}} \gamma \in \operatorname{Adm}(\mu, \nu)$.

Using the Corollary 1.11 we know that $\operatorname{supp} \gamma_k$ is c_k-ciclically monotone. For a fix N and permutation σ we have

$$(\operatorname{supp} \gamma_k)^{\otimes N} \subseteq \mathcal{C}_k = \left\{ ((x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)) \ : \ \sum_{i=1}^{N} c_k(x_i, y_i) \leq \sum_{i=1}^{N} c_k(x_i, y_{\sigma(i)}) \right\}.$$

For k big enough, the uniform convergence gives the inequalities $c - \frac{\varepsilon}{2N} \leq c_k \leq c + \frac{\varepsilon}{2N}$, so we have the inclusions:

$$\mathcal{C}_k \subseteq \mathcal{C}_\varepsilon := \left\{ ((x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)) \ : \ \sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{\sigma(i)}) + \varepsilon \right\}.$$

As c is continuous, $\mathcal{C}_\varepsilon$ is a closed set, so $(\operatorname{supp} \gamma)^{\otimes N} \subseteq \mathcal{C}_\varepsilon$. This is true for any $\varepsilon > 0$, so it is also true that $(\operatorname{supp} \gamma)^{\otimes N} \subseteq \mathcal{C}_0$, which proves that γ is c-cyclically monotone.

The final part is a direct consequence of the Fundamental Theorem 1.10. □

1.2 Duality

In the Kantorovich problem, we look to minimize $\gamma \mapsto \int c \, d\gamma$ when $\gamma \in \operatorname{Adm}(\mu, \nu)$. In some sense, this can be seen as minimizing a linear functional with the constraints $\pi^X_{\#\gamma} = \mu$ and $\pi^Y_{\#\gamma} = \nu$. This kind of optimization problems admit a natural dual problem, whose optimal solution relates to the optimal solution of the original one.

Consider the following problem, which we will prove that is the dual of the Optimal transport problem.

Problem 1.14 **Dual Problem**

Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and a cost function $c : X \times Y \to \mathbb{R}$, we want to maximize

$$(\varphi, \psi) \longmapsto \int \varphi d \, \mu + \int \psi \, d\nu$$

for $\varphi \in L^1(\mu)$, $\psi \in L^1(\nu)$ subject to

$$\forall (x, y) \in X \times Y, \qquad \varphi(x) + \psi(y) \leq c(x, y).$$

Theorem 1.15 (Duality theorem). *Assuming that the cost function c is continuous, bounded from below and satisfying*

$$c(x,y) \leq a(x) + b(y), \qquad a \in L^1(\mu),\ b \in L^1(\nu),$$

Then,

- *The supremum of the Dual problem has the same value as the minimum of the Kantorovich Problem 1.2.*
- *The supremum of the Dual problem is attained by a pair of the form (φ, φ^{c+}) for some $\varphi \in L^1(\mu)$ c-concave.*

Proof. First, for any plan $\gamma \in \mathrm{Adm}(\mu,\nu)$ and any (φ,ψ) solution (not necessarily optimal) of the dual problem,

$$\int c\, d\gamma \geq \int \varphi(x) + \psi(y)\, d\gamma = \int \varphi\, d\mu + \int \psi\, d\nu. \tag{1.2}$$

Consider now an optimal plan $\gamma \in \mathrm{Opt}(\mu,\nu)$. By the fundamental theorem 1.10 we know there is some c-concave function $\varphi : X \to \mathbb{R}$ with $\operatorname{supp}\gamma \subseteq \partial^{c+}\varphi$. Then, we get

$$c(x,y) = \varphi(x) + \varphi^{c+}(y), \qquad \forall (x,y) \in \operatorname{supp}\gamma.$$

Taking $\psi = \varphi^{c+}$, this makes the inequality of (1.2) an equality, so the proof is completed. □

Definition 5 **Kantorovich potential**

A Kantorovich potential is a c-concave function φ such that (φ, φ^{c+}) is a optimal solution of the Dual Problem 1.14.

Proposition 1.16. *Under the assumptions of Theorem 1.15, consider any Kantorovich potential φ and any optimal plan γ, then we have the inclusion $\operatorname{supp}\gamma \subseteq \partial^{c+}\varphi$.*

Proof. For any plan $\gamma \in \mathrm{Opt}(\mu,\nu)$, because of the Fundamental Theorem 1.10 and the proof of theorem 1.15, we know there is a Kantorovich potencial φ with $\operatorname{supp}\gamma \subseteq \partial^{c+}\varphi$.

For any other $\gamma' \in \mathrm{Opt}(\mu,\nu)$ we have that

$$\int \varphi\, d\mu + \int \varphi^{c+}\, d\nu = \int \varphi(x) + \varphi^{c+}(y)\, d\gamma' \leq \int c\, d\gamma' = \int c\, d\gamma = \int \varphi\, d\mu + \int \varphi^{c+}\, d\nu.$$

To make the inequality an equality, it must be that γ'-a.e. $(x,y) \in \partial^{c+}\varphi$. Using that c is continuous we get $\operatorname{supp}\gamma' \subseteq \partial^{c+}\varphi$.

On the other hand, if we consider another Kantorovich potential $(\varphi', \varphi'^{c+})$ we get

$$\int \varphi \, d\mu + \int \varphi^{c+} \, d\nu = \int c \, d\gamma \geq \int \varphi'(x) + \varphi'^{c+}(y) \, d\gamma = \int \varphi' \, d\mu + \int \varphi'^{c+} \, d\nu$$

As both sides are the same value (because they are optimal solutions of the dual problem), the inequality is an equality and, just like before, this implies $\operatorname{supp}\gamma \subseteq \partial^{c+}\varphi'$. □

1.3 Optimal plans induced by maps

Lemma 1.17 (Characterization of optimal maps). *Let $\gamma \in Adm(\mu,\nu)$. The following are equivalent,*

i. γ is induced by a map.

ii. There is a γ-measurable set $\Gamma \subseteq X \times Y$ where γ is concentrated and such that for μ-a.e. $x \in X$ there is only a point $y =: T(x) \in Y$ such that $(x, T(x)) \in \Gamma$.

If this happens, γ is induced by the transport map T.

Proof.

i⇒ii Let's assume that exists $T : X \to Y$ with $\gamma = (Id, T)_{\#\mu}$ (this is, γ is induced by T). Define $\Gamma = \{(x, T(x)) \ : \ x \in X\}$. We get that

$$\gamma(\Gamma) = \int_\Gamma 1 \, d\gamma = \int_X 1 \, d\mu = 1$$

because Γ is the image of X by the map $(Id, T) : X \to X \times Y$.

ii⇒i The conditions of Γ allows us to consider $T : X \to Y$, a priori well defined only in a set of μ-measure 1 but that we can extend with any value in the problematic set of null measure.

For any $f : X \times Y \to \mathbb{R}$ continuous with compact support

$$\begin{aligned}\int f(x,y) \, d\gamma &= \int_\Gamma f(x,y) \, d\gamma = \int_\Gamma f(x,T(x)) \, d\gamma = \int f(x,T(x)) \, d\gamma = \int_X f(x,T(x)) \, d\mu \\ &= \int f(x,y) \, d(Id,T)_{\#\mu}\end{aligned}$$

so indeed $\gamma = (Id, T)_{\#\mu}$.

To see the final part that T is a transport plan we have to see that $T_{\#\mu} = \nu$. For this, consider any Borel set $B \subseteq Y$,

$$T_{\#\mu}(B) = \mu(T^{-1}(B)) = \int \chi_{T^{-1}(B)}(x)\, d\mu = \int \chi_B(T(x))\, d\gamma = \int \chi_B(y)\, d\gamma = \nu(B).$$

□

Proposition 1.18. *Consider the cost function $c(x,y) = |x-y|^2$ and $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$. $\varphi(x)$ is c-concave if and only if the function $\overline{\varphi}(x) := |x|^2/2 - \varphi(x)$ is convex and lower semicontinous. Moreover, $y \in \partial^{c+}\varphi(x)$ if and only if $y \in \partial^- \overline{\varphi}(x)$.*

Proof. We know φ is c-concave if and only if $\exists\psi$ with $\varphi(x) = \inf_y \left\{ \frac{|x-y|^2}{2} - \psi(y) \right\}$, but expanding the norm we get

$$\varphi(x) = \inf_y \left\{ \frac{|x|^2}{2} + \frac{|y|^2}{2} - \langle x, y\rangle - \psi(y) \right\} \iff \underbrace{\varphi(x) - \frac{|x|^2}{2}}_{-\overline{\varphi}(x)} = \inf \left\{ \frac{|y|^2}{2} - \langle x, y\rangle - \psi(y) \right\}.$$

So $\overline{\varphi}(x) = \sup_y \left\{ \langle x, y\rangle - \left(\frac{|y|^2}{2} - \psi(y) \right) \right\} = \left(\frac{|y|^2}{2} - \psi(y) \right)^-$, so $\overline{\varphi}$ is convex for the cost function $\langle x, y\rangle$, but because of the Remark 1.7 this means that it is convex (in the standard sense).

For the second part, we apply Proposition 1.6 and by expanding again the norm we get a c-superdifferential for $c(x,y) = \langle x, y\rangle$:

$$\begin{aligned} y \in \partial^{c+}\varphi(x) &\iff \varphi(z) - \frac{|z-y|^2}{2} \leq \varphi(x) - \frac{|x-y|^2}{2} \quad \forall z \in X \\ &\iff \varphi(z) - \frac{|z|^2}{2} + \langle z, y\rangle \leq \varphi(x) - \frac{|x|^2}{2} + \langle x, y\rangle \quad \forall z \in X \\ &\iff \overline{\varphi}(z) + \langle x, y\rangle \leq \overline{\varphi}(x) + \langle x, y\rangle \quad \forall z \in X \\ &\iff y \in \partial\overline{\varphi}(x). \end{aligned}$$

□

Definition 6 **Regular Measures**

A measure on $\mathbb{R}^d$ is said to be **regular** if it gives 0 mass to any c-c hypersurface, with the meaning that $E \subseteq \mathbb{R}^d$ is a c-c hypersurface (*convex minus convex hypersurface*) if, in some system of coordinates, we can write the set as the graph

$$E = \left\{ (y,t) \; : \; y \in \mathbb{R}^{d-1}, t \in \mathbb{R}, t = g(y) - f(y) \right\}$$

with $f, g : \mathbb{R}^{d-1} \to \mathbb{R}$ convex functions.

A clear example of regular measures are absolutely continuous ones. This is because they give no mass to Lipschitz hypersurfaces, and a c-c hypersurface is locally Lipschitz.

The differentiability of convex functions has a connection with c-c hypersurfaces, as the theorem below shows. The proof can be consulted on [14, Thereom 2].

Theorem 1.19. *$A \subseteq \mathbb{R}^d$ can be covered by countable many c-c hypersurfaces if and only if there is some convex function $f : \mathbb{R}^d \to \mathbb{R}$ such that the non-differenciable points of f cover A.*

Theorem 1.20 **Brenier's theorem**

Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ be a measure with finite second moment and the cost function $c(x, y) = |x-y|^2$, then the following are equivalent:

1. For each $\nu \in \mathcal{P}(\mathbb{R}^d)$ with finite second moment, there exists only one optimal transport plan from μ to ν, and it is induced by a map T.
2. μ is a regular measure.

If any of these cases happens, then the optimal map T is the gradient of some convex function.

Remark 1.21. Before the proof, it will be useful to recall some basic properties of the subdifferential. Given $f : \mathbb{R}^n \to \mathbb{R}$ defined in $dom(f)$ we say that $v \in \mathbb{R}^n$ is a subgradient of f at $x \in dom(f)$ if

$$f(z) \geq f(x) + v \cdot (z - x) \qquad \forall z \in dom(f).$$

The subdifferential at x is $\partial^- f(x)$ the set of al subgradients at x.

- It is a closed convex set.
- If f is convex and $x \in int\, dom(f)$ then $\partial^- f(x)$ is non-empty and bounded.
- If f is convex and differentiable at x then $\partial^- f(x) = \{\nabla f(x)\}$.

Proof.

1⇒2: In order to find a contradiction, we assume that exists a c-c hypersurface E' with positive measure. Then by Theorem 1.19 exists $f : \mathbb{R}^d \to \mathbb{R}$ such that the set E of non-differenciable points covers E' and therefore $\mu(E) > 0$.

By possibly modifying f outside of a compact set, we can also assume it has linear growth at infinity. We consider the maps

$$\begin{aligned} T(x) &= \text{ the element of smallest norm in } \partial^- f \\ S(x) &= \text{ the element of biggest norm in } \partial^- f \end{aligned}$$

Because $\partial^- f(x)$ is a closed convex bounded set it makes sense to consider the element of smallest norm and T is well defined. For the element of biggest norm, it may happen that there are several of them. In this case pick one following some criteria (the closest to $(1,0)$ for example)

Note that because of the definition of E, we have that $T \not\equiv S$ inside E and $T \equiv S$ outside. Consider the measure

$$\gamma = \frac{1}{2}\left((Id,T)_{\#\mu} + (Id,S)_{\#\mu}\right) \in \mathcal{P}(\mathbb{R}^{2d})$$

The linear growth means that exists $k \geq 0$ such that $f(x) = k|x| + O(|x|)$. For a fixed $x \in \mathbb{R}^d$, consider any $v \in \partial^- f(x)$, $v \neq 0$. This satisfies

$$f(z) \geq f(x) + v \cdot (z - x) = f(x) - v \cdot x + v \cdot z, \qquad\qquad \forall z \in \mathbb{R}^d.$$

Then $f(z) = k|z| + O(|z|) \geq f(x) - v \cdot x + v \cdot z$. We can choose z parallel to v so that $v \cdot z = |v| \cdot |z|$. Then

$$k|z| + O(|z|) \geq f(x) - v \cdot x + |v| \cdot |z| \qquad \Rightarrow \qquad (k - |v|)\,|z| + O(|z|) \geq f(x) - v \cdot x$$

As the right hand size is fixed, by making $|z| \to \infty$ we get that $|v| \leq k$.

This means that all the subgradients are bounded by k, so notably $T(\mathbb{R}^d) \subseteq [0,k]$ and $S(\mathbb{R}^d) \subseteq [0,k]$. This implies that $\nu = \pi^Y_{\#\gamma}$ is concentrated on $\left\{x \in \mathbb{R}^d \,:\, |x| \leq k\right\}$. Because of the compact support $\int |x|^2 \, d\nu < \infty$.

To see that γ is optimal, we want to check that $\operatorname{supp}\gamma \subseteq \partial^{c+}\varphi$ for some c-concave function. Note that if in Proposition 1.18 we take $f = \overline{\varphi}$ and $\varphi(x) = \frac{|x|^2}{2} - f(x)$ then φ is c-concave. Moreover, as $\int |x|^2 d\mu < \infty$, we can use the linear growth and continuity of f to get that $\varphi \in L^1(\mathbb{R}^d, \mu)$. To conclude that γ is then optimal we need only to check that $\operatorname{supp}\gamma \subseteq \partial^{c+}\varphi = \partial^- f$. But because

$$(Id,T)_{\#\mu}(U \times V) = \mu\left(\{x \in U \,:\, T(x) \in V\}\right), \quad (Id,S)_{\#\mu}(U \times V) = \mu\left(\{x \in U \,:\, S(x) \in V\}\right),$$

if $(x,y) \notin \{(x, S(x))\} \subseteq \partial^- f$ we can find a ball around it outside of $\partial^{c+}\varphi = \partial^- f$ and therefore with γ-measure zero. Note that this ball exists because the set

$$\partial^{c+}\varphi = \left\{ (x,y) \in \mathbb{R}^{2d} \ : \ \varphi(x) + \varphi^{c+}(y) = \frac{|x-y|}{2} \right\}$$

is closed as the functions φ, φ^{c+} are continuous.

We can apply the hypothesis and γ is induced by a map, but this is a contradiction because $T \neq S$ on E with $\mu(E) > 0$.

2⇒1: As both μ and ν have finite second moment we get that the bound

$$|x-y|^2 \leq 2|x|^2 + 2|y|^2$$

works for the Fundamental Theorem of OT 1.10 and the duality Theorem 1.15.

Moreover, by the Proposition 1.16 we know that for any c-concave Kantorovich potential φ and any $\gamma \in \mathrm{Opt}(\mu,\nu)$ we have the inclusion $\operatorname{supp}\gamma \subset \partial^{c+}\varphi$. But the cost function $|x-y|^2$ implies that $\overline{\varphi}(x) = \frac{|x|^2}{2} - \varphi(x)$ is convex and $\partial^{c+}\varphi = \partial^{-}\overline{\varphi}$. Because μ is a regular measure, the set of non-differentiability of $\overline{\varphi}$ has measure 0. This means that the $\nabla\overline{\varphi}$ is well defined μ-almost everywhere, and because $\operatorname{supp}\gamma \subseteq \partial^{-}\overline{\varphi}$ the plans are concentrated on the graph of $\nabla\overline{\varphi}$ and therefore induced by it. This also proves that there is only one *optimal* plan, the one induced by the map $\nabla\overline{\varphi} = T$.

□

Many of the use of Brenier's theorem comes from the following corollary, which follows mostly from applying Brenier's theorem to both measures and the lemma 1.17.

Corollary 1.22

If $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$ are regular measures with finite second moment, then the optimal transport plan from μ to ν is induced by a map T, which is the gradient of a convex function and is a bijection $T : A \to B$ for $A, B \subseteq \mathbb{R}^d$ with $\mu(A) = \nu(B) = 1$.

Indeed, if T, S are the optimal maps given by the Brenier's theorem with $T_{\#\mu} = \nu$ and $S_{\#\nu} = \mu$ then for any measurable function $F : \mathbb{R}^d \to \mathbb{R}$

$$\int F \, d\nu = \int F \circ T \, d\mu = \int F \circ T \circ S \, d\nu$$

So $T \circ S = Id$ a.e. in ν. The same reasoning shows that $S \circ T = Id$ a.e. in μ, which gives that T is a bijection in the sense explained in the statement.

Remark 1.23. We will usually use the map $T : A \to B$ as change of variables, as in this situation is a bijection. This map was the derivative of a convex function, so we can consider the gradient of T because the second derivative of a convex function is well-defined almost everywhere, thanks to Aleksandrov's Theorem (check [6, Theorem A.2] for the proof).

Remark 1.24. Note that when f, g are the probability densities of μ, ν the condition $T_{\#\mu} = \nu$ is equivalent that for any $F : \mathbb{R}^d \to \mathbb{R}$ measurable

$$\int F(T(x))f(x)\, dx = \int F(x)g(x)\, dx = \int F(T(x))g(T(x)) \det(\nabla T(x))\, dx$$

So by taking a sequence approximating a Dirac delta we get that

$$g(T(x)) \det(\nabla T(x)) = f(x) \qquad \xRightarrow{T=\nabla u} \qquad g(\nabla u(x)) \det(D^2 u(x)) = f(x).$$

And this is a particular case of the Monge-Ampère equation, a highly non-linear elliptic PDE. Brenier's theorem proves then the existence of a solution when f, g satisfy some conditions.

Chapter 2
Wasserstein distance

Until now, when we have studied the Optimal Transport problem we have been interested on the optimal plans and their properties. In this section the focus is going to shift to the optimal value of the problem, and how this allows to define a distance between measures.

As in the previous section, we will assume X to be a Polish space with metric d.

2.1 Definition and Wasserstein spaces

Definition 7 **Wasserstein distance**

For any $p \in [1, +\infty)$ and $\mu, \nu \in \mathcal{P}(X)$, consider the cost function $c(x, y) = d(x, y)^p$. Then the Wasserstein distance is

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \mathrm{Adm}(\mu,\nu)} \int d(x, y)^p d\gamma \right)^{\frac{1}{p}} = \left(\int d(x, y)^p d\tilde{\gamma} \right)^{\frac{1}{p}} \qquad \tilde{\gamma} \in \mathrm{Opt}(\mu, \nu)$$

A first easy example is that for any $x, y \in X$ then $W_p(\delta_x, \delta_y) = d(x, y)$. This is due to the fact that any plan $\gamma \in \mathrm{Adm}(\delta_x, \delta_y)$ must be concentrated on (x, y), so the only admisible plan is the optimal.

Although in this case the distance is the same for any value of p, in general it depends on p.

Proposition 2.1. *W_p is an extended distance on $\mathcal{P}(X)$, meaning that*

1. $W_p : \mathcal{P}(X) \times \mathcal{P}(X) \to [0, +\infty]$.
2. *It is symmetric.*
3. *If $W_p(\mu, \nu) = 0$ then $\mu = \nu$.*
4. *It satisfies the triangular inequality: $W_p(\mu, \nu) \leq W_p(\mu, \lambda) + W_p(\lambda, \nu)$.*

Proof. The first two points are clearly true. For the third, if $W_p(\mu,\nu) = 0$ then exists $\gamma \in \mathrm{Adm}(\mu,\nu)$ with

$$\int d(x,y)^p d\gamma = 0.$$

This means that γ is concentrated on the diagonal $\{x = y\}$, being induced by the map $Id : X \to X$, which makes $\mu = \nu$. □

For the last part we need the following lemma.

Lemma 2.2 (Gluing or composition lemma). *Let X, Y, Z be three polish spaces and $\gamma_1 \in \mathcal{P}(X \times Y)$, $\gamma_2 \in \mathcal{P}(Y \times Z)$ with $\pi^Y_{\#\gamma_1} = \pi^Y_{\#\gamma_2}$. Then exists $\gamma \in \mathcal{P}(X \times Y \times Z)$ with such that $\pi^{X,Y}_{\#\gamma} = \gamma_1$ and $\pi^{Y,Z}_{\#\gamma} = \gamma_2$.*

Proof. Let's call $\mu := \pi^Y_{\#\gamma_1} = \pi^Y_{\#\gamma_2}$. We can use the disintegration lemma to write $d\gamma_1(x,y) = d\mu(y)d\gamma_{1,y}(x)$ with the meaning that

$$\int_{X\times Y} f(x,y)\ d\gamma_1(x,y) = \int_Y \left(\int_X f(x,y)d\gamma_{1,y}(x)\right)\ d\mu(y).$$

Note that this is just the conditional probability for discrete and absolutely continuous, while for singular measures the disintegration theorem[1]gives the theoretical support.

We do the same for $d\gamma_2(y,z) = \mu(y)d\gamma_{2,y}(z)$. Then

$$d\gamma(x,y,z) = d\mu(y)d\gamma_{1,y}(x)d\gamma_{2,y}(z)$$

satisfies the wanted properties. □

Continuation of 2.1 proof: If we consider $\gamma_1 \in \mathrm{Opt}(\mu,\lambda)$ and $\gamma_2 \in \mathrm{Opt}(\lambda,\nu)$ then we can apply the Gluing lemma to get $\gamma \in \mathcal{P}(X^3)$ with marginals μ, λ, ν. As $\pi^{1,3}_{\#\gamma} \in \mathrm{Adm}(\mu,\nu)$,

$$\begin{aligned}W_p(\mu,\nu) &\le \left(\int d^p\ d\pi^{1,3}_{\#\gamma}\right)^{\frac{1}{p}} = \left(\int d(x_1,x_3)^p d\gamma(x_1,x_2,x_3)\right)^{\frac{1}{p}} \\ &\le \left(\int d(x_1,x_2)^p d\gamma(x_1,x_2,x_3)\right)^{\frac{1}{p}} + \left(\int d(x_2,x_3)^p d\gamma(x_1,x_2,x_3)\right)^{\frac{1}{p}} = W_p(\mu,\lambda) + W_p(\lambda,\nu).\end{aligned}$$

□

[1]The proof can be checked on [5], [3] or [9].

Corollary 2.3 **Wasserstein's spaces**

W_p is a distance on the set of Borel probabilities

$$\mathcal{P}_p(X) = \left\{ \mu \in \mathcal{P}(X) \ : \ \int d(x_0, x)^p d\mu(x) < \infty \text{ for some } x_0 \in X \right\}.$$

It is a common convention to abbreviate a metric space (X, d) to just X, as the metric can usually be inferred by the context. In this case, as the novelty is not on the space $\mathcal{P}_p$ but on the distance, it is frequent to just write $W_p(X)$. Naturally, the base space X may be omitted if it does not cause confusion and use instead W_p or $\mathcal{P}_p$.

Proof. First note that if $\int d(x_0, x) d\mu(x)$ is finite for some x_0 then it is finite for any $\tilde{x}_0 \in X$:

$$\int d(\tilde{x}_0, x)^p d\mu(x) \leq 2^{p-1} \int d(\tilde{x}_0, x_0)^p + d(x_0, x)^p d\mu(x) < \infty.$$

So for any $\mu, \nu \in \mathcal{P}_p$ and $\gamma \in \mathrm{Adm}(\mu, \nu)$

$$W_p(\mu, \nu)^p \leq \int d(x, y)^p d\gamma(x, y) \leq 2^{p-1} \left(\int d(x, x_0)^p d\mu(x) + \int d(x_0, y)^p d\nu(y) \right) < \infty$$

$\square$

Proposition 2.4. *If* $1 \leq p \leq q$ *then* $W_p \leq W_q$.

Proof. This is a consequence that when λ is a finite measure $\|\cdot\|_{L^1(\lambda)} \leq \|\cdot\|_{L^\alpha(\lambda)}$ for any $\alpha \geq 1$. Taking an optimal plan $W_q(\mu, \nu) = \left(\int d^q \, d\gamma\right)^{\frac{1}{q}}$ then

$$W_p(\mu, \nu)^p \leq \int d^p d\gamma \leq \left(\int d^q d\gamma \right)^{\frac{p}{q}} = W_q(\mu, \nu)^p \qquad \Rightarrow \qquad W_p \leq W_q.$$

$\square$

This tells us that the distance W_1 is dominated by any other. This case $p = 1$ is specially frequent, and is also known as the Kantorovich-Rubinstein distance.

Proposition 2.5 **Kantorovich-Rubinstein duality**

$W_1(\mu, \nu) = \sup \left\{ \int \varphi d\mu - \int \varphi d\nu \ : \ \|\varphi\|_{Lip} \leq 1 \right\}$ where $\|\varphi\|_{Lip}$ is the Lipschitz seminorm.

The proof is a direct consequence of duality combined with the following lemma. Indeed,

$$W_1(\mu,\nu) = \sup\left\{\int \varphi d\mu + \int \varphi^{c+} d\nu \ : \ \varphi \ c\text{-concave}\right\} = \sup\left\{\int \varphi d\mu - \int \varphi d\nu \ : \ \|\varphi\|_{Lip} \le 1\right\}.$$

Lemma 2.6. *If the cost function is a distance $c(x,y) = d(x,y)$ (not necessarily the one of the metric space X), then $\varphi : X \to \mathbb{R}$ is c-concave if and only if it is 1-Lipschitz (for the distance d). In this case $\varphi^{c+} = -\varphi$.*

Proof. If φ is c-concave, then

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

This implies $\varphi(x) - \varphi(z) \ge \inf_y d(x,y) - d(z,y)$, so

$$\varphi(z) - \varphi(x) \le \sup_y d(z,y) - d(x,y) \le d(z,x).$$

As the role of x and z is symmetric, this proves that φ is 1-Lipschitz.

If instead we assume that φ is 1-Lipschitz, this is,

$$|\varphi(x) - \varphi(y)| \le d(x,y) \qquad \forall x,y \in X.$$

Then $-\varphi(y) \le d(x,y) - \varphi(x)$ so $\varphi^{c+}(y) = \inf_x d(x,y) - \varphi(x) = -\varphi(y)$ and, as $-\varphi$ is still 1-Lipschitz, $\varphi^{c+c+} = (-\varphi)^{c+} = \varphi$ and it is c-concave. □

Definition 8 **Weak topology on $\mathcal{P}_p$**

Given (μ_k) a sequence measures in $\mathcal{P}_p$ and $\mu \in \mathcal{P}_p$, we say that $\mu_k \rightharpoonup \mu$ weakly on $\mathcal{P}_p$ if $\mu_k \to \mu$ narrowly and for some $x_0 \in X$

$$\int d(x_0,x)^p d\mu_k(x) \to \int d(x_0,x)^p d\mu(x).$$

Remark 2.7. *This condition is equivalent to the any of the following ones:*

1. $\mu_k \to \mu$ *narrow and* $\limsup_k \int d(x_0,x)^p d\mu_k(x) \to \int d(x_0,x)^p d\mu(x)$.
2. $\mu_k \to \mu$ *narrow and* $\lim_{R\to+\infty} \limsup_k \int_{d(x_0,x)\ge R} d(x_0,x)^p d\mu_k(x) = 0$.
3. *For all continuous functions* $f : X \to \mathbb{R}$ *with* $|f(x)| \le C\,(1 + d(x_0,x)^p)$ *we have that*

$$\int f \, d\mu_k \to \int f d\mu$$

Proof. Let's denote by (0) the condition of the definition. Then, there are some trivial implications: (3)⇒(0)⇒(1). So we have to check two more implications.

(1)⇒(2): For any $R \geq 0$, we can split the limit in

$$\begin{aligned}\int d(x_0,x)^p d\mu &= \limsup_k \int d(x_0,x)^p d\mu_k \\ &= \limsup_k \int d(x_0,x)^p \chi_{\{d(x_0,x)\leq R\}} d\mu_k + \limsup_k \int d(x_0,x)^p \chi_{\{d(x_0,x)\geq R\}} d\mu_k\end{aligned}$$

The function of the first limit is bounded, and as $\mu_k \to \mu$ narrowly we get

$$\limsup_k \int d(x_0,x)^p \chi_{\{d(x_0,x)\leq R\}} d\mu_k = \int d(x_0,x)^p \chi_{\{d(x_0,x)\leq R\}} d\mu \xrightarrow{R\to\infty} \int d(x_0,x)^p d\mu$$

where the last limit is a consequence of $\mu \in \mathcal{P}_p$. This proves the hypothesis of *(2)*.

(2)⇒(3): Given $f : X \to \mathbb{R}$ continuous with $|f(x)| \leq C(1 + d(x,x_0)^p)$, for any $R > 0$ we can consider the bounded continuous function $\tilde{f}(x) = \max\{-R, \min\{R, f(x)\}\}$. Then

$$\lim_k \int f\, d\mu_k = \lim_k \left(\int \tilde{f}\, d\mu_k + \int f - \tilde{f}\, d\mu_k \right) = \int \tilde{f}\, d\mu + \lim_k \int f - \tilde{f}\, d\mu_k.$$

It is clear that $\lim_{R\to\infty} \int \tilde{f}\, d\mu = \int f\, d\mu$, so we want to check that the limit as $R \to \infty$ of the second part is 0. For this we can reason as follows:

$$\left| \int f - \tilde{f}\, d\mu_k \right| \leq \int \left| f - \tilde{f} \right|\, d\mu_k = \int_{|f(x)|\geq R} |f(x)| - R\, d\mu_k \leq \int_{|f(x)|\geq R} |f(x)|\, d\mu_k.$$

But if $R \leq |f(x)| \leq C(1 + d(x,x_0)^p)$ we get $\frac{R}{C} - 1 \leq d(x,x_0)^p$ and we can take an R big enough so that $d(x,x_0) \geq R'$ with $R' \to \infty$ when $R \to \infty$, and when $R' \geq 1$ we get $|f(x)| \leq 2Cd(x,x_0)^p$. Using the hypothesis we finish the proof:

$$\lim_{R\to\infty} \limsup_k \left| \int f - \tilde{f}\, d\mu_k \right| \leq \lim_{R\to\infty} \limsup_k \int_{d(x,x_0)\geq R'} 2Cd(x,x_0)^p\, d\mu_k = 0.$$

□

This topology is not only metrizable, but is generated by the Wasserstein distance, as the following theorem shows.

Theorem 2.8

$\mu_k \rightharpoonup \mu$ in $\mathcal{P}_p$ if and only if they converge in the W_p, this is, $W_p(\mu_k, \mu) \to 0$.

First, we are going to require the following lemma, that has interest on its own.

Lemma 2.9. *If (μ_k) is a Cauchy sequence in W_p then $\{\mu_k\}$ is tight.*

Proof. Recall that to prove a sequence is tight we need to find for every $\varepsilon > 0$ a compact set $K_\varepsilon = K$ with $\mu_n(X\backslash K) \leq \varepsilon$ for all $n \in \mathbb{N}$.

That (μ_k) is Cauchy means that $W_p(\mu_k, \mu_l) \xrightarrow{k,l\to+\infty} 0$. Notably, this plus the triangular inequality imply that $W_p(\mu_1, \mu_k)$ is bounded. From this we get

$$W_p(\delta_{x_0}, \mu_k)^k = \int d(x_0, x)^p d\mu_k(x) \leq (W_p(\delta_{x_0}, \mu_1) + W_p(\mu_1, \mu_k))^p$$

is bounded too.

Because $W_p \geq W_1$, we get that $(\mu_k) \subseteq \mathcal{P}_p \subseteq \mathcal{P}_1$ are Cauchy for W_1 too. This means that for any given $\varepsilon > 0$ exists N with

$$k \geq N \Rightarrow W_1(\mu_N, \mu_k) < \varepsilon^2.$$

Note that for any $k \in \mathbb{N}$ exists $j = j(k) \in \{1, \ldots, N\}$ with $W_1(\mu_j, \mu_k) < \varepsilon^2$:

- If $k \geq N$, $j(k) = N$ works by the previous inequality.
- If $k \leq N$, we can take $j(k) = k$.

Recall that any Borel measure is tight, and as the finite union of tight sets is tight, the set $\{\mu_1, \ldots, \mu_N\}$ is tight:

Exists K compact with $\mu_j(X\backslash K) < \varepsilon$ for $1 \leq j \leq N$.

We can cover it with a finite number of balls like $K \subseteq \cup_{i=1}^m B(x_i, \varepsilon) =: U$. We are going to define

$$V = \{x \in X \ : \ d(x, U) < \varepsilon\} \qquad \varphi(x) = \left(1 - \frac{d(x,U)}{\varepsilon}\right)_+ = \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{if } x \notin V \\ \frac{1}{\varepsilon}(\varepsilon - d(x,U)) & \text{if } x \in V\backslash U \end{cases}$$

$U \subseteq V \subseteq \cup B(x_i, 2\varepsilon)$ and $\chi_U \leq \varphi \leq \chi_V$. It is also not difficult to check that φ is a $(\frac{1}{\varepsilon})$-Lipschitz function. This means that $\|\varepsilon\varphi\|_{Lip} = 1$, and using the Kantorovich-Rubinstein duality (2.5) we get

$$\int \varphi d\mu_k - d\mu_j = \frac{1}{\varepsilon} \int \varepsilon\varphi d\mu_k - d\mu_j \leq \frac{1}{\varepsilon} W_1(\mu_k, \mu_j) \leq \varepsilon$$

where $j = j(k)$ is described as before. This gives the estimate

$$\mu_k(V) = \int \chi_V d\mu_k \geq \int \varphi d\mu_k = \int \varphi d\mu_j + \left(\int \varphi d\mu_k - d\mu_j \right) \geq \mu_j(U) - \varepsilon \geq 1 - 2\varepsilon.$$

What we have proven is that for any $\varepsilon > 0$ exists a finite number of points $x_1, \ldots, x_m \in X$ with

$$W_\varepsilon = \cup_{i=1}^m \overline{B}(x_i, 2\varepsilon), \qquad \mu_k(X \backslash W_\varepsilon) \leq 2\varepsilon \quad \forall k \in \mathbb{N}.$$

We would like for W_ε to be compact, but in general it might not be. This has a classical solution: replacing ε by $\frac{\varepsilon}{2^{l+1}}$ and consider the union.

For each $l \in \mathbb{N}$ exists $m(l)$ and $x_1, \ldots, x_m \in X$ with $\mu \left(X \backslash \cup_{i=1}^{m(l)} \overline{B}(x_i, 2^{-l}\varepsilon) \right) \leq 2^{-l}\varepsilon$. We define

$$K = \cap_{l=1}^{\infty} \cup_{i=1}^{m(l)} \overline{B}(x_i, 2^{-l}\varepsilon).$$

Then $\mu_k(X \backslash K) \leq \varepsilon$, it is a closed set, and because X is complete then K is too. Also, it is *totally bounded*, this meaning that for any arbitrarily small $\varepsilon > 0$ it can be covered by finitely many balls of radius ε. But a complete totally bounded set is compact, so K is compact. □

Theorem 2.8 proof: Let's assume that $(\mu_k), \mu \in \mathcal{P}_p$ and $W_p(\mu_k, \mu) \to 0$. We want to show that $\mu_k \rightharpoonup \mu$ in $\mathcal{P}_p$. By the previous lemma, $\{\mu_k\}$ is a tight set. This means that it is relatively compact, so it has a subsequence $\mu_{k_n} \to \tilde{\mu}$ narrowly.

We want to check that $\mu = \tilde{\mu}$. For this recall that $\gamma \mapsto \int d(x,y)^p d\gamma$ is l.s.c. If we pick $\gamma_n \in \operatorname{Opt}(\mu, \mu_{k_n})$ then $\{\gamma_n\}$ is a tight set, and then it has another subsequence γ_m with $\gamma_m \xrightarrow{\text{nrw}} \gamma$. As $\gamma \in \operatorname{Adm}(\mu, \tilde{\mu})$ we can use the l.s.c. to get

$$W_p(\mu, \tilde{\mu})^p \leq \int d(x,y)^p d\gamma \leq \liminf_{m \to \infty} W_p(\mu, \mu_m)^p = 0.$$

So indeed $\mu = \tilde{\mu}$. As this works for any converging subsequence of μ_k, we get $\mu_k \xrightarrow{\text{nrw}} \mu$.

Recall the following real analysis fact: for $\varepsilon > 0$ exists C_ε such that $(a+b)^p \leq (1+\varepsilon)a^p + C_\varepsilon b^p$ for any $a, b \in \mathbb{R}$. So for any three points $d(x_0, x)^p \leq (1+\varepsilon) d(x_0, y)^p + C_\varepsilon d(x,y)^p$. If we consider $\pi_k \in \operatorname{Opt}(\mu_k, \mu)$:

$$\begin{aligned} \int d(x_0, x)^p d\mu_k(x) &= \int d(x_0, x)^p d\pi_k(x,y) \leq (1+\varepsilon) \int d(x_0, y)^p d\pi_k(x,y) + C_\varepsilon \int d(x,y)^p d\pi_k(x,y) \\ &= (1+\varepsilon) \int d(x_0, y) d\mu(y) + C_\varepsilon W_p(\mu_k, \mu)^p. \end{aligned}$$

So by taking $\limsup_k$ first and then $\varepsilon \to 0$ we get that

$$\limsup_k \int d(x_0,x)^p d\mu_k(x) \leq \limsup_k \int d(x_0,x)^p d\mu(x)$$

and $\mu_k \rightharpoonup \mu$ in $\mathcal{P}_p$.

For the other implication, if $\mu_k \rightharpoonup \mu$ in $\mathcal{P}_p$ we want to check that $W_p(\mu_k,\mu) \to 0$.

Note that $\mu_k \xrightarrow{\text{nrw}} \mu$ too, and this implies that $\{\mu_k\}$ is relatively closed, and by Prokhorov theorem it is tight. Pick $\pi_k \in \mathrm{Opt}(\mu,\mu_k)$, then $\{\pi_k\}$ is a tight set too. This means that there exists a subsequence $\pi_{k'} \xrightarrow{\text{nrw}} \pi$. As

$$\int d(x,y)^p d\pi_k \leq 2^{p-1}\int d(x,x_0)^p d\pi_k + 2^{p-1}\int d(x_0,y)^p d\pi_k = 2^{p-1}\int d(x,x_0)^p d\mu_k + 2^{p-1}\int d(x_0,y)^p d\mu,$$

when we take $\liminf_k$ it is bounded, and by Theorem 1.12 $\pi_{k'} \xrightarrow{\text{nrw}} \pi \in \mathrm{Opt}(\mu,\mu)$. But then $\pi = (Id,Id)_{\#\mu}$. As this works for any converging subsequence, $\pi_k \xrightarrow{\text{nrw}} \pi$.

Fix $x_0 \in X$. For any $R > 0$, if $d(x,y) \geq R$ the chain of inequalities

$$2\max\{d(x,x_0), d(x_0,y)\} \geq d(x,x_0) + d(x_0,y) \geq d(x,y) \geq R$$

gives $\max\{d(x,x_0), d(x_0,y)\} \geq d(x,y)$ and $\max\{d(x,x_0), d(x_0,y)\} \geq R$. More formally, this shows that

$$\chi_{\{d(x,y)>R\}} \leq \chi_{\{d(x,x_0)\geq\frac{R}{2},\ d(x,x_0)\geq\frac{d(x,y)}{2}\}} + \chi_{\{d(x_0,y)\geq\frac{R}{2},\ d(x_0,y)\geq\frac{d(x,y)}{2}\}}.$$

$$\begin{aligned} d(x,y)^p\chi_{\{d(x,y)\geq R\}} &\leq d(x,y)^p\chi_{\{d(x,x_0)\geq\frac{R}{2},\ d(x,x_0)\geq\frac{d(x,y)}{2}\}} + d(x,y)^p\chi_{\{d(x_0,y)\geq\frac{R}{2},\ d(x_0,y)\geq\frac{d(x,y)}{2}\}} \\ &\leq 2^p d(x,x_0)\chi_{\{d(x_0,x)\geq\frac{R}{2}\}} + 2^p d(x_0,y)\chi_{\{d(x_0,y)\geq\frac{R}{2}\}}. \end{aligned}$$

Consider

$$W_p(\mu_k,\mu)^p = \int d(x,y)^p d\pi_k(x,y) = \int \min\{d(x,y)^p, R^p\} d\pi_k + \int \Big(d(x,y)^p - R^p\Big)_+ d\pi_k(x,y).$$

The first integral goes to 0 when $k \to \infty$, as it is a continuous bounded function and $\pi_k \xrightarrow{\text{nrw}} \pi$ with π concentrated on $\{x = y\}$. For the second integral we can use the previous bounds:

$$\begin{aligned} \int \Big(d(x,y)^p - R^p\Big)_+ d\pi_k(x,y) &\leq \int d(x,y)^p\chi_{\{d(x,y)\geq R\}} d\pi_k(x,y) \\ &\leq 2^p\int_{d(x_0,x)\geq\frac{R}{2}} d(x,x_0)^p d\pi_k(x,y) + 2^p\int_{d(x_0,y)\geq\frac{R}{2}} d(y,x_0)^p d\pi_k(x,y) \\ &\leq 2^p\int_{d(x_0,x)\geq\frac{R}{2}} d(x,x_0)^p d\mu_k(x) + 2^p\int_{d(x_0,y)\geq\frac{R}{2}} d(y,x_0)^p d\mu(y) \end{aligned}$$

Therefor $\limsup_k W_p(\mu_k,\mu)^p \leq \lim_{R\to+\infty}\limsup_k \int d(x,y)^p\chi_{\{d(x,y)\geq R\}}d\pi_k(x,y) \leq 0$ by the definition of weak convergence in $\mathcal{P}_p$. This proves that $W_p(\mu_k,\mu)\to 0$. □

Theorem 2.10

The space $(\mathcal{P}_p, W_p)$ is a Polish space, this is, as a metric space it is complete and separable.

Proof. As we already know that it is a metric space, we only need to check that it is complete and separable.

Separability: As X is a separable space, let's take D a dense subsequence of X and consider

$$\mathcal{F} = \left\{\sum_{j=1}^N a_j\delta_{x_j} \ : \ 1\leq N<\infty,\ a_j\in\mathbb{Q},\ a_j\geq 0,\ \sum a_j = 1,\ x_j\in D\right\}\subseteq \mathcal{P}_p(X)$$

This set is numerable, so we would like to check that it is dense. Given $\mu\in\mathcal{P}_p$, exists a compact set $K\subseteq X$ such that

$$\int_{X\setminus K} d(x_0,x)^p d\mu(x)\leq \varepsilon^p.$$

Note that $B(x_j,\varepsilon)$ for $x_j\in D\cap K$ is an open cover of K, so we can cover K with a finite number of balls $B(x_j,\varepsilon)$ with $x_j\in D$, $1\leq j\leq N$. Define

$$B_k' = B(x_k,\varepsilon)\setminus\cup_{j<k} B(x_j,\varepsilon)$$

so that the B_k' are disjoint. We consider then $f: X\to X$ given by

$$f(x) = \begin{cases} x_0 & \text{if } x\in X\setminus K \\ x_k & \text{if } x\in B_k'\cap K\end{cases}$$

This satisfies $d(x,f(x))\leq\varepsilon$ for any $x\in K$, so

$$\int_X d(x,f(x))^p d\mu = \int_K d(x,f(x))^p d\mu + \int_{X\setminus K} d(x,x_0)^p\, d\mu(x)\leq 2\varepsilon^p.$$

As $(Id,f)_{\#\mu}\in\mathrm{Adm}(\mu,f_{\#\mu})$, we have computed that $W_p^p(\mu,f_{\#\mu})\leq 2\varepsilon^p$. It is easy to check that $f_{\#\mu} = \sum_{j=1}^N a_j\delta_{x_j}$ for $0\leq a_j\in\mathbb{R}$ (and the $x_j\in D$). We want then to approximate this probability with another with rational coefficients.

We are going to consider $b_j\in\mathbb{Q}$ with these conditions:

- $0\leq b_j\leq a_j$ for $j\geq 2$ and $|a_j-b_j|\leq\varepsilon'$.
- As we want to have $\sum b_j = 1 = \sum a_j$, pick b_1 given by $b_1 = 1-\sum_{j\geq 2} b_j = a_1+\sum_{j\geq 2} a_j - b_j$. So $0\leq a_1\leq b_1$.

This makes $\sum_{j=1}^{N} b_j \delta_{x_j}$ a probability measure on $\mathcal{F}$.

We can define $\gamma = \sum_{i,j=1}^{N} c_{i,j}\delta_{(x_i,x_j)} \in \mathrm{Adm}(\sum a_j\delta_{x_j}, \sum b_j\delta_{x_j})$ as

$$c_{1,1} = a_1, \qquad c_{j,j} = b_j \; j \geq 2, \qquad c_{i,1} = a_i - b_i \; i \geq 2, \qquad c_{i,j} = 0 \text{ otherwise.}$$

And by taking ε' small enough we get

$$W_p^p(\sum a_j\delta_{x_j}, \sum b_j\delta_{x_j}) \leq \sum_{i=2}^{N}(a_i - b_i)d(x_1,x_i)^p \leq \varepsilon' \sum_{i=2}^{N} d(x_1,x_i)^p \leq \varepsilon^p.$$

Completness: Given a Cauchy sequence $(\mu_k) \subseteq \mathcal{P}_p$, by Lemma 2.9 it is a tight set. It admits then a subsequence $\mu_n \xrightarrow{\text{nrw}} \mu$. Using the l.s.c.

$$\int d(x_0,x)^p d\mu(x) \leq \liminf_n \int d(x_0,x)^p d\mu_n(x) \leq \liminf_n \left[W_p(\delta_{x_0},\mu_1) + W_p(\mu_1,\mu_n)\right]^p < +\infty.$$

This is, $\mu \in \mathcal{P}_p$. Using again the l.s.c. and that the sequence is Cauchy we get that

$$\limsup_m W_p(\mu,\mu_m) \leq \limsup_m \liminf_n W_p(\mu_n,\mu_m) \leq \limsup_{m,n} W_p(\mu_n,\mu_m) = 0.$$

This proves that $\mu_n \to \mu$ in the W_p topology. A classical argument shows that when a Cauchy sequence has a convergent subsequence the whole sequence is convergent to the same point, so the proof is finished. $\square$

2.2 W_∞ distance

Note that we can write $W_\infty(\mu,\nu) = \inf\left\{\|d(x,y)\|_{L^p(\gamma)} : \gamma \in \mathrm{Adm}(\mu,\nu)\right\}$. This justifies the following definition.

Definition 9

For $\mu,\nu \in \mathcal{P}(X)$ two probabilities $W_\infty(\mu,\nu) = \inf\left\{\|d(x,y)\|_{L^\infty(\gamma)} : \gamma \in \mathrm{Adm}(\mu,\nu)\right\}$.

Note that this is not a Kantorovich's optimal transport problem. This, combined with the fact that it will not share some of the key properties of W_p spaces, motivates to just briefly study this concept.

We also define

$$\mathcal{P}_\infty(X) = \{\mu \in \mathcal{P}(X) \ : \ \mathrm{supp}\,\mu \text{ is bounded}\}\,.$$

Proposition 2.11. *For* $\mu,\nu \in \mathcal{P}_\infty(X)$

$$W_\infty(\mu,\nu) = \lim_{p\to\infty} W_p(\mu,\nu).$$

Proof. Consider $\gamma_k \in \mathrm{Adm}(\mu,\nu)$ such that $\|d(x,y)\|_{L^k(\gamma)} = W_k(\mu,\nu)$. As $\{\gamma_k\} \subseteq \mathrm{Adm}(\mu,\nu)$ the set $\{\gamma_k\}$ is tight, and then in has a convergent subsequence: $\gamma_{k'} \xrightarrow{\text{nrw}} \gamma \in \mathrm{Adm}(\mu,\nu)$.

Because μ,ν are bounded supported, γ_k is too, and the function $d(x,y)$ is bounded for $x,y \in \mathrm{supp}\,\gamma_k$. Then, the narrow convergence implies

$$\int d(x,y)^p d\gamma_k \to \int d(x,y)^p d\gamma$$

This means that

$$\|d(x,y)\|_{L^p(\gamma_k)} \xrightarrow{k\to\infty} \|d(x,y)\|_{L^p(\gamma)} \xrightarrow{p\to\infty} \|d(x,y)\|_{L^\infty(\gamma)}$$

And therefore $\lim_p W_p(\mu,\nu) \geq W_\infty(\mu,\nu)$. The other inequality is because for any $\gamma \in \mathrm{Adm}(\mu,\nu)$

$$\|d(x,y)\|_{L^\infty(\gamma)} \geq \|d(x,y)\|_{L^p(\gamma)} \geq W_p(\mu,\nu).$$

$\square$

Corollary 2.12. *(P_∞, W_∞) is a metric space.*

This is a consequence of the previous proposition, as the only non-trivial property to check is the triangular inequality, and now it can be proved by taking limits in W_p.

Remark 2.13. *$\mathcal{P}_\infty \subsetneq \cap_{p\geq 1}\mathcal{P}_p$. Indeed, it is easy to give an example of a probability in $\mathcal{P}_p$ for any p but not in $\mathcal{P}_\infty$. Consider $X = \mathbb{R}$ and*

$$\mu = \sum_{n=1}^{\infty} \frac{1}{2^n}\delta_n, \qquad \operatorname{supp}\mu = \{1, 2, 3, \dots\}.$$

Remark 2.14. *If $\mu \in \mathcal{P}_p$ for all $p \geq 1$ and $W_\infty(\mu, \nu) < \infty$ for all $\nu \in \mathcal{P}_\infty$ then $\mu \in \mathcal{P}_\infty$.*

We know that $W_\infty(\mu, \delta_{x_0}) < \infty$, but as $\operatorname{Adm}(\mu, \delta_{x_0}) = \{\gamma\}$ with γ induced by the map $T : X \to X$, $T(x) \equiv x_0$, then

$$W_\infty(\mu, \delta_{x_0}) = \|d(x, y)\|_{L^\infty(\gamma)} = \|d(x, x_0)\|_{L^\infty(\mu)}$$

and $\operatorname{supp}\mu$ is bounded.

Proposition 2.15. *W_∞ does not metrize the weak topology on $\mathcal{P}_\infty$.*

In $X = \mathbb{R}$ consider $\mu_n = \frac{n-1}{n}\delta_0 + \frac{1}{n}\delta_1$ and $\nu = \delta_0$. Then $\mu_n \rightharpoonup \nu$ but $W_\infty(\mu_n, \nu) = 1$.

Chapter 3
Classical geometrical inequalities

3.1 Isoperimetric Inequality

Theorem 3.1 **Isoperimetric inequality**

For any open set $E \subseteq \mathbb{R}^d$, let's denote by $P(E)$ its perimeter and $B \subseteq \mathbb{R}^d$ the unit ball. Then we have that

$$\mathcal{L}^d(E)^{1-\frac{1}{d}} \leq \frac{P(E)}{d\mathcal{L}^d(B)^{\frac{1}{d}}}.$$

Proof. We can restrict to ourselves to the case when E has finite measure, as when it is infinite both sides are $+\infty$. We can define the following probabilities then:

$$\mu = \frac{1}{\mathcal{L}^d(E)}\mathcal{L}^d_{|E} \in \mathcal{P}(\mathbb{R}^d), \qquad \nu = \frac{1}{\mathcal{L}^d(B)}\mathcal{L}^d_{|B} \in \mathcal{P}(\mathbb{R}^d).$$

Both measures are absolutely continuous, so we can apply Brenier's theorem 1.22. This means that, for the cost $c(x,y) = |x-y|^2$, there is only one admisible plan, that it is also induced by an optimal plan $T : \overline{E} \to \overline{B}$. Moreover, T is the gradient of a convex function and a bijection, after possibly ignoring a set of zero measure. Now, using the change of variable $T(x)$ on $T_{\#\mu} = \nu$, we get that for any Borel function $F : B \to \mathbb{R}$,

$$\frac{1}{\mathcal{L}^d(B)}\int_B F = \frac{1}{\mathcal{L}^d(E)}\int_E (F \circ T) = \int_B F\frac{1}{\mathcal{L}^d(E)\det(\nabla T(T^{-1}y))}.$$

Taking now a sequence of functions that approximate a Dirac delta we get the following pointwise inequality:

$$\frac{1}{\mathcal{L}^d(E)} = \det(\nabla T(x))\frac{1}{\mathcal{L}^d(B)}, \qquad \forall x \in E.$$

On the other hand, as $\nabla T(x)$ is the Hessian of a convex function, it is a symmetric matrix with eigenvalues $\lambda_i(x) = \lambda_i \geq 0$, so we have the inequality

$$(\det \nabla T(x))^{\frac{1}{d}} = (\lambda_1 \dots \lambda_d)^{\frac{1}{d}} \leq \frac{\lambda_1 + \cdots + \lambda_d}{d} = \frac{Tr(\nabla T(x))}{d} = \frac{\nabla \cdot T(x)}{d}. \tag{3.1}$$

Combaining both facts, $\dfrac{1}{\mathcal{L}^d(E)^{\frac{1}{d}}} \leq \dfrac{\nabla \cdot T(x)}{d\mathcal{L}^d(B)}$. Now, integrating over E and using the divergence theorem

$$\mathcal{L}^d(E)^{1-\frac{1}{d}} \leq \int_E \frac{\nabla \cdot T(x)}{d\mathcal{L}^d(B)^{\frac{1}{d}}} dx = \frac{1}{d\mathcal{L}^d(B)^{\frac{1}{d}}} \int_{\partial E} \langle T(x), \overrightarrow{n}(x) \rangle dS(x).$$

Because $T(x) \in B$, $\|T(x)\| \leq 1$ and therefore $|\langle T(x), \overrightarrow{n}(x) \rangle| \leq 1$, so we get the wanted result. □

Remark 3.2. *In (3.1) we have used the inequality between the arithmetic and geometric means. We know that for this to be an equality it would have to be that $\lambda_1 = \lambda_2 = \cdots = \lambda_d$. But if all the eigenvalues are the same, then the map $T : E \to B$ is just a homothety and E is a ball (of radius possibly not 1). Also, in this case the only other inequality transforms into a equality too.*

This shows both that the inequality is sharp, and that is only attained when E is a ball.

3.2 Sobolev Inequality

Theorem 3.3 **Sobolev Inequality**

For all functions $f \in W^{1,p}(\mathbb{R}^d)$ exists a constant $C = C(p, d)$ such that for $1 \leq p < d$

$$\left(\int |f|^{p*} \right)^{\frac{1}{p*}} \leq C \left(\int |\nabla f|^p \right)^{\frac{1}{p}}$$

where $p^* = \dfrac{dp}{d-p}$ (or equivalently $\frac{1}{p^*} + \frac{1}{d} = \frac{1}{p}$).

Proof. By normalization we can assume $\int |f|^{p*} = 1$. Also, we can assume $f \geq 0$ without loss of generality. Because the functions with compact support are dense we can assume that f has it. With this hypothesis, what we want to prove is that

$$\left(\int |\nabla f|^p \right)^{\frac{1}{p}} \geq C.$$

Let's fix a smooth function $g : \mathbb{R}^d \to \mathbb{R}$ with $g \geq 0$ and $\int g = 1$. With this we can consider the following probabilities on $\mathbb{R}^d$:

$$\mu = f^{p*}\mathcal{L}^d, \qquad\qquad \nu = g\mathcal{L}^d.$$

We are again in a situation where we can apply Brenier's Theorem to get the optimal transport map $T : \mathbb{R}^d \to \mathbb{R}^d$. Using the change of variable $T(x)$ on $T_{\#\mu} = \nu$ and taking aproximations we get the equality

$$g(T(x)) = \frac{f^{p*}(x)}{\det \nabla T(x)} \qquad\qquad \forall x \in \mathbb{R}^d.$$

Note that $T_{\#\mu} = \nu$ means that for every $F : \mathbb{R}^d \to \mathbb{R}$ Borel $\int Fg = \int (F \circ T) f^{p*}$. From this we can deduce that

$$\int g^{1-\frac{1}{d}} = \int g^{-\frac{1}{d}} g = \int (g \circ T)^{-\frac{1}{d}} f^{p*} = \int \det(\nabla T)^{\frac{1}{d}} (f^{p*})^{1-\frac{1}{d}} \leq \frac{1}{d} \int \nabla \cdot T f^{p*(1-\frac{1}{d})}.$$

where we have used that ∇T satisfies again (3.1). Using now a change of variables and that f has compact support, we get

$$\int g^{1-\frac{1}{d}} \leq -\frac{p*}{d}\left(1 - \frac{1}{d}\right) \int f^{\frac{p*}{q}} T \cdot \nabla f, \qquad\qquad \frac{1}{p} + \frac{1}{q} = 1.$$

At last, using the Hölder inequality we get the wanted bound:

$$\int g^{1-\frac{1}{d}} \leq \frac{p*}{d}\left(1 - \frac{1}{d}\right)\left(\int f^{p*}|T|^q\right)^{\frac{1}{q}}\left(\int |\nabla f|^p\right)^{\frac{1}{p}} = \frac{p*}{d}\left(1 - \frac{1}{d}\right)\left(\int g(x)|x|^q dx\right)^{\frac{1}{q}}\left(\int |\nabla f|^p\right)^{\frac{1}{p}}$$

□

Chapter 4
Brunn-Minkowski Inequality

Before presenting a proof of the Brunn-Minkowski inequality we need some preliminary results. Mainly, we are going to see first that $W_2(\mathbb{R}^d)$ is a geodesic space and then define a convex functional on it.

4.1 Geodesic space

Remark 4.1. *It will useful to remember the following concepts. In a metric space (X,d) we say that a curve or path is a continuous application $\omega : [0,1] \to X$. Its length is*

$$\mathrm{Len}(\omega) = \sup\left\{\sum_{k=0}^{n-1} d(\omega(t_k), \omega(t_{k+1})) \ : \ n \geq 1,\ 0 = t_0 < t_1 < t_2 < \cdots < t_n = 1\right\}.$$

We will consider only curves with finite length $\mathrm{Len}(\omega) < +\infty$.

We define the (metric) derivative of a curve as

$$|\omega'|(t) := \lim_{h\to 0} \frac{d(\omega(t+h), \omega(h))}{|h|}, \qquad\qquad t \in [0,1].$$

A curve is absolutely continuous $\omega \in \mathrm{AC}(X)$ *when exists $g \in L^1([0,1])$ such that*

$$d(\omega(t), \omega(s)) \leq \int_t^s g$$

for all $0 \leq t < s \leq 1$.

We can reparametrize an absolutely continuous so that it is Lipschitz. Also, for $\omega \in \mathrm{AC}(X)$ $\mathrm{Len}(\omega) = \int_0^1 |\omega'|(t)\, dt$.

A curve is a geodesic between x_0 and x_1 if it minimizes the length among all absolutely continuous curves with $\omega(0) = x_0$ and $\omega(1) = x_1$.

We say that X is a geodesic space if for every pair of points

$$d(x,y) = \min\left\{\mathrm{Len}(\omega) \;:\; \omega \in \mathrm{AC}(X),\ \omega(0)=x,\ \omega(1)=y\right\}.$$

A curve it is said to be a constant speed geodesic if

$$d(\omega(t),\omega(s)) = |t-s| d(\omega(0),\omega(1)).$$

It is easy to check that this curve is indeed a geodesic.

Theorem 4.2. *For $X \subseteq \mathbb{R}^d$ convex and the cost function $c(x,y) = |x-y|^p$, $p > 1$, given $\mu,\nu \in \mathcal{P}_p(X)$ and $\gamma \in \mathrm{Opt}(\mu,\nu)$ we define*

$$\begin{aligned} \pi_t : X \times X &\longrightarrow X, \\ (x,y) &\longmapsto (1-t)x + ty \end{aligned} \qquad \mu_t := (\pi_t)_{\#\gamma}.$$

Then for $t \in [0,1]$ the curve μ_t is a constant speed geodesic between μ and ν on $W_p(X)$.

If $\gamma = (Id,T)_{\#\mu}$ then $\mu_t = ((1-t)Id + tT)_{\#\mu}$.

Proof. We make the claim $W_p(\mu_t,\mu_s) \le |t-s| \cdot W_p(\mu,\nu)$. Applying it several times proves indeed the result: for $t < s$

$$W_p(\mu,\nu) \le W_p(\mu,\mu_t) + W_p(\mu_t,\mu_s) + W_p(\mu_s,\nu) \le W_p(\mu,\nu)\left(t + (s-t) + (1-s)\right) = W_p(\mu,\nu)$$

where we have used the claim with $\mu_0 = \mu$ and $\mu_1 = \nu$. But this forces all the inequalities to be equalities, so $W_p(\mu_t,\mu_s) = (s-t)W_p(\mu,\nu)$.

To prove the claim note that the following plan is admisible

$$\gamma_{s,t} = (\pi_t,\pi_s)_{\#\gamma} \in \mathrm{Adm}(\mu_t,\mu_s).$$

This is because $\pi^1_{\#\gamma_{s,t}} = (\pi_t)_{\#\gamma} = \mu_t$. We have then the inequalities:

$$\begin{aligned} W_p(\mu_t,\mu_s) &\le \left(\int |x-y|^p \, d\gamma_{s,t}\right)^{\frac{1}{p}} = \left(\int |\pi_t(x,y) - \pi_s(x,y)|^p d\gamma\right)^{\frac{1}{p}} \\ &= \left(\int |(1-t)x + ty - (1-s)x - sy|^p \, d\gamma\right)^{\frac{1}{p}} = |t-s| \left(\int |x-y|^p d\gamma\right)^{\frac{1}{p}} = |t-s| W_p(\mu,\nu). \end{aligned}$$

Finally, when the plan is induced by a map $\gamma = (Id,T)_{\#\mu}$, for any measurable function $F : X \to \mathbb{R}$

$$\int F \, d\mu_t = \int F \, d(\pi_t)_{\#\gamma} = \int F((1-t)x + ty) \, d\gamma(x,y) = \int F((1-t)x + tT(x)) \, d\mu(x)$$

So $\mu_t = ((1-t)Id + tT)_{\#\mu}$. □

This theorem has as a direct consequence the following corollary.

Corollary 4.3

For $p > 1$ and $X \subseteq \mathbb{R}^d$ convex $W_p(X)$ is a geodesic space.

4.2 A Convex Functional over W_2

Definition 10 **Internal energy functional**

Given $u : [0, +\infty) \to \mathbb{R} \cup \{+\infty\}$ convex with $u(0) = 0$ we define the funcional

$$\mathcal{E}(\mu) = \int_{\mathbb{R}^d} u(\mu(x))dx$$

for absolutely continuous probabilities $\mu = \mu(x)dx$ in $\mathbb{R}^d$.

Remark 4.4. *Although the integral domain is $\mathbb{R}^d$, we can restrict it to the support of μ thanks to the condition $u(0) = 0$. Moreover, we can get rid off this condition by considering*

$$\mathcal{E}(\mu) = \int_{\mu(x)>0} u(\mu(x))\ dx$$

Remark 4.5. *We have not checked that $\mathcal{E}(\mu)$ is well defined, this is, that the integral makes sense, as it could happen that both the positive and negative parts of $u(\mu(x))$ were not integrable. Indeed, in general it will not be, so we usually need to assume some additional assumptions either on u or on μ.*

A quite general and useful condition that guarantees this is when u is bounded from below and satisfies that

$$\liminf_{x\to 0^+} \frac{u(x)}{x^\alpha} > -\infty \qquad \text{for some } \alpha > \frac{d}{d+2},$$

Then the negative part satisfies $u^-(x) \leq ax + bx^\alpha$. We can assume that $\alpha < 1$, and then we get that

$$\alpha > \frac{d}{d+2} \iff \alpha d + 2\alpha > d \iff 2\alpha > d(1-\alpha) \iff \frac{2\alpha}{1-\alpha} > d$$

and then the following integral is finite

$$\int \mu(x)^\alpha \, dx = \int_{\mathbb{R}^d} \mu(x)^\alpha (1+|x|)^{2\alpha}(1+|x|)^{-2\alpha} \, dx$$
$$\leq \left(\int \mu(x)(1+|x|)^2 dx\right)^\alpha \left(\int (1+|x|)^{\frac{-2\alpha}{1-\alpha}} dx\right)^{1-\alpha} < +\infty$$

This makes the integral of $\mathcal{E}(\mu)$ well defined, perhaps being $\mathcal{E}(\mu) = +\infty$.

Definition 11 **Interpolating curves**

Given $\mu, \nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ with μ a regular measure, consider T_0 and T_1 the optimal maps from μ to ν_0 and ν_1 for the cost $c(x,y) = |x-y|^2$. Then the interpolating curve from ν_0 to ν_1 with base μ is

$$\nu_t = ((1-t)T_0 + tT_1)_{\#\mu}, \qquad t \in [0,1].$$

A special case is when $\mu = \nu_0$, as we then the curve is just the geodesic from μ_0 to μ_1. This also justifies the name of generalized geodesic to the curve μ_t, although I will favor the previous one, as this is not (in general) a geodesic.

Theorem 4.6

Given u, consider its functional $\mathcal{E}$. Assume the map $x \mapsto x^d u(x^{-d})$ is convex and non-increasing on $(0,+\infty)$, then the functional $\mathcal{E}$ is convex, meaning that for any $\mu, \nu_0, \nu_1 \in \mathcal{P}(\mathbb{R}^d)$ absolutely continuous and any $t \in (0,1)$

$$\mathcal{E}(\nu_t) \leq (1-t)\mathcal{E}(\nu_0) + t\mathcal{E}(\nu_1).$$

Notably, this theorem affirms that ν_t is absolutely continuous when the original measures are. As a corollary, when $\mu = \nu_0$ we get that $\mathcal{E}$ is geodesically convex.

On the other hand, it *does not* affirm that $\mathcal{E}$ is well defined over the interpolating curve, and we will need some additional assumption like the one of Remark 4.5 for this.

Proof. Thanks to Brenier's theorem, we know that the optimal maps T_0 and T_1 are gradients of convex functions and that they are injective in a set $A \subseteq \mathbb{R}^d$ with $\mu(A) = 1$. They are a bijection with the images, which has also mass 1: $\nu_0(T_0(A)) = 1$ $\nu_1(T_1(A)) = 1$. Also, the maps are differentiable in A, and their gradient is the Hessian matrix of a convex function, therefore with positive determinant and positive semi-definite.

Let's call $T_t = (1-t)T_0 + tT_1$. Note that as the gradient is linear, $\nabla T_t = (1-t)\nabla T_0 + t\nabla T_1$ is the convex combination of ∇T_0 and ∇T_1. Also, these 3 matrices are the Hessians of convex functions, so they are symmetric positive semi-definite (a.e. in $x \in \mathbb{R}^d$). Because the functional $A \mapsto \det(A)^d$ is concave (Lemma 4.7) and $\det(\nabla T_0), \det(\nabla T_1) > 0$, we get that $\det(\nabla T_t) > 0$ a.e. in x.

For any measurable function F we can do a change of variables to get that

$$\int F \, d\nu_t = \int F(T_t(x))\mu(x) \, dx = \int F(y)\frac{\mu(T_t^{-1}(y))}{\det \nabla T_t(T_t^{-1}(y))}dy.$$

So the measure ν_t is absolutely continuous with density is given by

$$\nu_t(y) = \frac{\mu(T_t^{-1}(y))}{\det \nabla T_t(T_t^{-1}(y))}, \qquad \nu_t(T_t(x)) = \frac{\mu(x)}{\det \nabla T_t(x)}.$$

Then the functional $\mathcal{E}$ is well defined along the curve ν_t:

$$\mathcal{E}(\nu_t) = \int_{\mathbb{R}^d} u(\nu_t(y)) \, dy = \int_{\mathbb{R}^d} u\left(\frac{\mu(x)}{\det \nabla T_t(x)}\right) \det \nabla T_t(x) \, dx.$$

Then, it is clear that if we prove that $A \mapsto \left(\frac{\mu(x)}{\det(A)}\right) \det(A)$ is a convex function for $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ symmetric positive semi-definite then the functional $\mathcal{E}$ is convex.

Recall that if $g : D \subseteq \mathbb{R}^n \to \mathbb{R}$ is concave and $h : \mathbb{R} \to \mathbb{R}$ is convex and non-increasing, then $f = h \circ g$ is convex. In our case,

- $x \mapsto x^d u\left(\frac{\mu(x)}{x^d}\right)$ is convex and non-increasing (a.e. in x, as we have to impose $\mu(x) > 0$).
- $A \mapsto (\det A)^{\frac{1}{d}}$ is concave (see Lemma 4.7 for the details).

Then the map is convex a.e.

$$A \longmapsto (\det A)^{\frac{1}{d}} = x \longmapsto x^d u\left(\frac{\mu(x)}{x^d}\right) = \det(A) u\left(\frac{\mu(x)}{\det A}\right).$$

$\square$

Lemma 4.7 (Minkowski determinant inequality). *The map $A \longmapsto (\det A)^{\frac{1}{n}}$ for A $n \times n$ symmetric positive semidefinite matrices is concave.*

Proof. Given two matrices of the form A and $A + B$, we want to check that

$$\begin{aligned} F : [0,1] &\longrightarrow \mathbb{R} \\ t &\longmapsto F(t) := \det(A+tB)^{\frac{1}{n}} \end{aligned}$$

is concave. Recall that a positive semidefinite symmetric matrix has a "square root", this is, exists $A^{\frac{1}{2}}$ positive semidefinite with $A^{\frac{1}{2}}A^{\frac{1}{2}} = A$. Then, we can write

$$\det(A+tB)^{\frac{1}{n}} = \det\left(A^{\frac{1}{2}}\left(Id + tA^{\frac{1}{2}}BA^{-\frac{1}{2}}\right)\frac{A^1}{2}\right)^{\frac{1}{n}} = \det(A)^{\frac{1}{n}}\det\left(Id + tA^{\frac{1}{2}}BA^{-\frac{1}{2}}\right)^{\frac{1}{n}}.$$

So we can assume without loss of generality that $A = Id$. To compute the derivatives of $F(t) = \det(Id+tB)^{\frac{1}{n}}$ we will use Jacobi's formula:

$$\frac{d}{dt}X(t) = (\det X(t))\cdot \operatorname{tr}\left(X(t)^{-1}\frac{dX(t)}{dt}\right)$$

With this $\dfrac{d}{dt}\det(Id+tB) = \det(Id+tB)\operatorname{tr}\left((Id+tB)^{-1}B\right)$ and

$$\begin{aligned} F'(t) &= \frac{1}{n}\det(Id+tB)^{\frac{1}{n}}\operatorname{tr}\left((Id+tB)^{-1}C\right) \\ F''(t) &= \frac{1}{n}\det(Id+tB)^{\frac{1}{n}}\left[\frac{1}{n}(\operatorname{tr}C)^2 - \operatorname{tr}(C^2)\right], \qquad\qquad C = (Id+tB)^{-1}B \end{aligned}$$

Finally, we have that $F''(t) \le 0$ because $(\operatorname{tr}C)^2 \le n\operatorname{tr}(C^2)$. Consider the eigenvalues of C: $\lambda_1, \lambda_2, \dots, \lambda_n$. Then $\operatorname{tr}C = \sum \lambda_i$ and $\operatorname{tr}(C^2) = \sum \lambda_i^2$ so using the Cauchy-Schwartz inequality

$$(\operatorname{tr}C)^2 = \left(\sum_{i=1}^n \lambda_i\right)^2 \le n\sum_{i=1}^n \lambda_i^2 = n\operatorname{tr}(C^2).$$

□

As we are imposing quite a few conditions on u for the functional $\mathcal{E}$ to be convex, a natural question is whether we can find functions satisfying all of them. The following proposition gives some families of these types of functions.

Proposition 4.8. *We can apply the previous theorem for the functions*

1. $u(x) = x\log(x)$.
2. $u(x) = \dfrac{x^\alpha - x}{\alpha - 1}$ *for* $\alpha \neq 1$, $\alpha \ge 1 - \frac{1}{d}$ *and dimension* $d > 2$.
3. $u(x) = x^p$ *for* $p > 1$.
4. $u(x) = -x^\alpha$ *for* $1 - \frac{1}{d} \le \alpha < 1$.

Note that although in the theory we have not asked for any smoothness in the function u, in practice this functions are at least C^2, so that we can check the convexity and non-increasing conditions easily.

Proof. As the previous comment says, many of the work is just computing the second derivative of u and $x \mapsto x^d u(x^{-d})$. Then the only condition worth commenting is that $\mathcal{E}$ is well defined. For the first two cases, we can easily fullfil the condition described in Remark 4.5. For the third case, $u^-(x) = 0$, while for the last $u^+(x) = 0$. $\square$

4.3 Brunn-Minkowski inequality proof

Finally, we are prepared to prove this classical result. The main idea of the proof is that we if we choose an adequate function u the functional $\mathcal{E}$ over uniform probabilities on a set is going to be an expression related to its measure, and using the convexity of the functional we can get the result.

Theorem 4.9 **Brunn-Minkowski inequality**

For any compact sets $A, B \subseteq \mathbb{R}^d$

$$\mathcal{L}\left(\frac{A+B}{2}\right)^{\frac{1}{d}} \geq \frac{\mathcal{L}(A)^{\frac{1}{d}} + \mathcal{L}(B)^{\frac{1}{d}}}{2}$$

Proof. If $\mathcal{L}(A) = 0$, then $\mathcal{L}(\frac{A+B}{2}) \geq \mathcal{L}(\frac{B}{2}) = \frac{1}{2^d}\mathcal{L}(B)$, so we have to focus on the case with $\mathcal{L}(A), \mathcal{L}(B) > 0$.

Consider the probabilities $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ given by the densities

$$\mu_0(x) = \frac{1}{\mathcal{L}(A)}\chi_A(x), \qquad \mu_1(x) = \frac{1}{\mathcal{L}(B)}\chi_B(x).$$

Consider μ_t the geodesic (in $W_2(\mathbb{R}^d)$) from μ_0 to μ_1. Then the functional

$$\mathcal{E}(\rho) = \int_{\mathbb{R}^d} u(\rho(x))\, dx, \qquad u(x) = -x^{1-\frac{1}{d}}\text{[1]}$$

is convex and therefore: $\mathcal{E}(\mu_{\frac{1}{2}}) \leq \dfrac{\mathcal{E}(\mu_0) + \mathcal{E}(\mu_1)}{2}$. We can easily compute the right side of the inequality:

$$\mathcal{E}(\mu_0) = \int -\mu_0(x)^{1-\frac{1}{d}}\, dx = -\mathcal{L}(A)^{\frac{1}{d}-1} \int \chi_A(x)^{1-\frac{1}{d}} dx \;= -\mathcal{L}(A)^{\frac{1}{d}}.$$

[1] For dimension $d = 1$ perhaps would be more precise to say $u(0) = 0$ and $u(x) = -1$ for $x > 0$.

By the same reasoning we compute $\mathcal{E}(\mu_1)$ and get that

$$\mathcal{E}(\mu_{\frac{1}{2}}) \leq -\left(\mathcal{L}(A)^{\frac{1}{d}} + \mathcal{L}(B)^{\frac{1}{d}}\right)$$

As u is a convex function, we can apply Jensen's inequality to de measure $\frac{dx}{\mathcal{L}^d\left(\frac{A+B}{2}\right)}$ to get

$$\int u(\mu_{\frac{1}{2}}(x)) \, \frac{dx}{\mathcal{L}^d\left(\frac{A+B}{2}\right)} \geq u\left(\int \mu_{\frac{1}{2}}(x) \frac{dx}{\mathcal{L}^d\left(\frac{A+B}{2}\right)}\right) = u\left(\mathcal{L}^d\left(\frac{A+B}{2}\right)^{-1}\right)$$

With this we can bound

$$\mathcal{E}(\mu_{\frac{1}{2}}) \geq \mathcal{L}^d\left(\frac{A+B}{2}\right) u\left(\mathcal{L}^d\left(\frac{A+B}{2}\right)^{-1}\right) = -\mathcal{L}^d\left(\frac{A+B}{2}\right)^{\frac{1}{d}}$$

that implies the wanted result. $\square$

Chapter 5
Application to Kronecker sequences

5.1 Bounds using Fourier coefficients

For $f(x)dx$ an absolutely continuous probability on $[0,1]$ it is clear that we can define its Fourier series as just taking $\widehat{f}(k)$. This generalizes nicely to general probabilities as

$$\widehat{\mu}(k) = \int e^{2\pi i k x} d\mu \qquad\qquad \mu \in \mathcal{P}([0,1]).$$

Proposition 5.1. *Given a probability $\mu \in \mathcal{P}([0,1])$ we have that for each $n \in \mathbb{N}$*

$$W_1(\mu,\ dx) \lesssim \frac{1}{n}\sum_{k=1}^{n} \frac{|\widehat{\mu}(k)|}{k}.$$

Proof. Note that $\mu \in W_1([0,1])$, so it can be approximated by $\nu_N = \frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}$, $x_k \in [0,1]$, when $N \to \infty$, as this kind of measures are dense. Applying this to the Wasserstein distance:

$$W_1(\mu, dx) \le W_1(\mu,\nu_N) + W_1(\nu_N, dx) \qquad \Rightarrow \qquad W_1(\mu, dx) \le \limsup_{N\to\infty} W_1(\nu_N, dx).$$

Using the duality, $W_1(\nu_N, dx) = \sup\left\{\sum_{k=1}^{N} f(x_k) - \int_0^1 f(x)dx \ : \ \|f\|_{Lip} \le 1\right\}$.

Using the Koksma-Hlawka[1] inequality we get that

$$\left|\frac{1}{N}\sum_{k=1}^{N} f(x_k) - \int_0^1 f(x)dx\right| \le var(f)\cdot \sup_{J\subseteq[0,1]} \left|\nu_N(J) - \mathcal{L}(J)\right|.$$

Where the supremum is taken over all intervals $J \subseteq [0,1]$. But the total variation of a 1-Lipschitz

[1]Check Appendix A for the details of the Koksma-Hlawka and the Erdős-Turan inequalities.

function is bounded by 1:

$$\begin{aligned}var(f) &= \sup\left\{\sum_{i=1}^{N-1} |f(x_{i+1}) - f(x_i)| \ : \ 0 \leq x_1 < x_2 < \cdots < x_N \leq 1\right\} \\ &\leq \sup\left\{\sum_{i=1}^{N-1} |x_{i+1} - x_i| \ : \ 0 \leq x_1 < x_2 < \cdots < x_N \leq 1\right\} = 1.\end{aligned}$$

The final step of the proof is applying the Erdős-Turan inequality:

$$\sup_{J\subseteq[0,1]} |\nu_N(J) - \mathcal{L}(J)| \lesssim \frac{1}{n} + \sum_{k=1}^{n} \frac{|\widehat{\nu}_N(k)|}{k}.$$

Taking $N \to \infty$, this is, making $\nu_N \to \mu$, we get the inequality of the statement. □

We now consider the Sobolev space of functions with zero average $g : \Omega \subseteq \mathbb{R}^d \to \mathbb{R}$ equipped with the norm

$$\|g\|_{\dot{H}^1} = \left(\int_\Omega |\nabla g|^2 \, dx\right)^{\frac{1}{2}}.$$

This defines the dual space $\dot{H}^{-1}$:

$$\|f\|_{\dot{H}^{-1}} = \sup\left\{\int_\Omega f \cdot g \, dx \ : \ \|g\|_{\dot{H}^1} \leq 1\right\}.$$

With a little more generality, for 2 measures on Ω we have that

$$\|\mu - \nu\|_{\dot{H}^{-1}} = \sup\left\{\int_\Omega g \, d(\mu - \nu) \ : \ \|g\|_{\dot{H}^1} \leq 1\right\}.$$

Note that because $\mu - \nu$ has average 0 we can ignore the condition for g to have 0 average, this is, replacing $\|g\|_{\dot{H}^1} \leq 1$ by $\|\nabla g\|_{L^2} \leq 1$.

Consider now the Neumann Problem

$$\begin{cases} -\Delta u = \mu - \nu & \text{in } \Omega \\ \frac{\partial u}{\partial n} = 0 & \text{on } \partial\Omega \end{cases}$$

where the first condition is understood in the weak sense: for all $f \in C^\infty(\Omega)$

$$\int f d(\mu - \nu) = \int \nabla f \cdot \nabla u \, dx.$$

If u is a solution of this problem, we claim that $\|\mu - \nu\|_{\dot{H}^{-1}} = \|\nabla u\|_{L^2}$. Indeed, it is enough to prove it in a dense set, and using the first condition of the problem

$$\int f\, d(\mu-\nu) = \int \nabla f \cdot \nabla u\, dx \leq \|\nabla f\|_{L^2} \cdot \|\nabla u\|_{L^2} \leq \|\nabla u\|_{L^2}.$$

Taking $f = \frac{u}{\|\nabla u\|_{L^2}}$ gives the equality.

Proposition 5.2. *Given an absolutely continuous probability $f(x)dx$ on $[0,1]$*

$$W_2(f(x)dx, dx) = \|1-f\|_{\dot{H}^{-1}} = \frac{1}{\pi}\left(\sum_{k=1}^{\infty} \frac{|\widehat{f}(k)|^2}{k^2}\right)^{\frac{1}{2}}.$$

Proof. Because $g(x)dx$ and dx are absolutely continuous probabilities we know thanks to Brenier's theorem that the W_2 distance is given by an optimal transport map (the cost $|x-y|^2$ plays the same role than $\frac{|x-y|^2}{2}$).

Then exists $\phi : [0,1] \to [0,1]$ convex function with $\phi' : [0,1] \to [0,1]$ satisfying

- $W_2(f(x)dx,\ dx) = \left(\int_0^1 |x-\phi'(x)|^2\, dx\right)^{\frac{1}{2}} = \|x-\phi'(x)\|_{L^2}$.
- $(\phi')_{\#dx} = f(x)dx$, meaning that for any measurable function $F : [0,1] \to \mathbb{R}$

$$\int_0^1 F(x)f(x)\, dx = \int_0^1 F(\phi'(x))\, dx.$$

- Up to a set of measure 0, ϕ' is injective.

The norm $\|1-f\|_{\dot{H}^{-1}}$ is determined by a Neumann problem, that as we are in dimension 1 can be written as

$$\left\{\begin{array}{ll} -u'' = 1-f & \text{in } (0,1) \\ u'(x) = 0 & \text{for } x = 0,1 \end{array}\right\} \Rightarrow \|1-f\|_{\dot{H}^{-1}} = \|u'\|_{L^2}.$$

For any $F \in C^\infty([0,1])$ we want to have $\int F(x)(1-g(x))\ dx = \int F'(x)u'(x)\ dx$. But using integration by parts

$$\int_0^1 F(x)(1-f(x))\, dx = -\int_0^1 F'(x)v(x), \qquad v(x) = \int_0^x (1-f(t))\, dt, \quad v(0)=v(1)=0.$$

So we can take $u'(x) = -v(x)\ dt$ as a solution. Now, we can use that ϕ' is monotone nondecreasing and (up to a null set) injective to get that a.e. in x

$$u'(x) = \int_0^x f(t)dt - x = \int_0^1 \chi_{[0,x]}(\phi'(t))dt - x = \phi'(x) - x$$

So $\|1-f\|_{\dot{H}^{-1}} = \|u'\|_{L^2} = \|\phi'(x)-x\|_{L^2} = W_2(f(x)dx,\ dx)$.

For the second part, because f is a probability on $[0,1]$ $1-f$ is a function with 0 average. Moreover its Fourier coefficients are $\widehat{(1-f)}(0)=0$ and $\widehat{(1-f)}(k)=-\widehat{f}(k)$ otherwise. By using Plancherel identity in the norms we get:

$$\|x\|_{\dot{H}^1(dx)}=\left(\sum_{k\in\mathbb{Z}}|\widehat{\nabla x}(k)|^2\right)^{\frac{1}{2}}.$$

As we are working on $[0,1]$, $\widehat{\nabla x}(k)=\widehat{x'}(k)=2\pi ik\widehat{x}(k)$. We can then bound $\|1-g\|_{\dot{H}^{-1}}$ by:

$$\int_0^1(1-f)x=\sum_{k\in\mathbb{Z}}\widehat{(1-f)}(k)\widehat{x}(k)=\frac{1}{2\pi}\sum_{k\neq 0}\frac{-\widehat{f}(k)}{k}\widehat{\nabla x}(k)\le\frac{1}{2\pi}\left(\sum_{k\neq 0}\frac{|\widehat{f}(k)|^2}{k^2}\right)^{\frac{1}{2}}\left(\sum_{k\neq 0}|\widehat{\nabla x}(k)|^2\right)^{\frac{1}{2}}$$
$$\le\frac{1}{\pi}\left(\sum_{k=1}^{\infty}\frac{|\widehat{f}(k)|^2}{k^2}\right)^{\frac{1}{2}}$$

When taking the supremum subject to $\|x\|_{\dot{H}^{-1}}\le 1$ we actually get the equality, as the only inequality used has been Cauchy-Schwarz. □

5.2 Kronecker sequences

A *badly approximable* number is a number $\alpha\in\mathbb{R}$ such that $\exists c>0$ with

$$\left|\alpha-\frac{p}{q}\right|\ge\frac{c}{q^2}\qquad\qquad\forall\frac{p}{q}\in\mathbb{Q}.$$

Note that the inequality is equivalent to $|q\alpha-p|\ge\frac{c}{q}$. Some examples of badly approximable numbers are $\sqrt{2},\sqrt{3},e,\dots$

We can get the following result

Theorem 5.3

For α badly approximable number we can define the probability $\mu_N=\frac{1}{N}\sum_{n=1}^{N}\delta_{\{n\alpha\}}$. Then,

$$W_2(\mu_N,dx)\lesssim_\alpha\frac{\sqrt{\log N}}{N}.$$

Proof. We start by defining the distance to nearest integer as

$$\|x\| = \min\{x - \lfloor x \rfloor, \lceil x \rceil - x\} = \min\{\{x\}, 1 - \{x\}\}.$$

Note that it satisfies these two properties:

- $\|k\alpha\| \gtrsim_\alpha \dfrac{1}{|k|}$ for α badly approximable.
- $\left|e^{2\pi i x} - 1\right| \sim \|x\|$.

With this we can easily bound the Fourier Coefficients of μ_N:

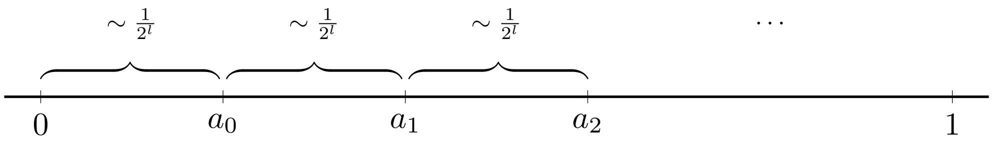
$$|\widehat{\mu}_N| = \left|\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i k n\alpha}\right| = \frac{1}{N}\left|\frac{e^{2\pi i k N\alpha} - 1}{e^{2\pi i k\alpha} - 1} e^{2\pi i k\alpha}\right| \leq \frac{1}{N}\frac{2}{|e^{2\pi i k\alpha} - 1|} \sim \frac{1}{N\,\|k\alpha\|}.$$

Let's consider the sets $A_l = \left\{\|k\alpha\| \;:\; 2^l \leq k \leq 2^{l+1}\right\}$. They satisfy the following

- $\|k\alpha\| \gtrsim \frac{1}{|k|} \geq \frac{1}{2^{l+1}} \sim \frac{1}{2^l}$.
- $|A_l| = 2^l$.
- The set is $\sim 2^{-l}$ separated, this is because for $a \neq b$ in the range $[2^l, 2^{l+1}]$ the difference is of the form

$$|\|a\alpha\| - \|b\alpha\|| = |\pm a\alpha \pm b\alpha + M| = |\alpha(\pm a \pm b) + M| \gtrsim \frac{1}{|\pm a \pm b|} \geq \frac{1}{2^{l+2}} \sim \frac{1}{2^l}.$$

From this we can deduce that $\displaystyle\sum_{x\in A_l} \frac{1}{x^2} \lesssim \sum_{m=1}^{2^l} \frac{1}{m^2/2^{2l}}$. Indeed, ordering the elements of A_l like $a_0 < a_1 < a_2 < \cdots < a_{2^l - 1}$ their distances are distributed as below:



Using Proposition 5.2 and that $|\widehat{\mu}_N(k)| \leq \|\mu\|_{L_1} = 1$ we get

$$W_2(\mu_N, dx) \sim \left(\sum_{k=1}^{\infty} \frac{|\widehat{\mu}_N(k)|^2}{k^2}\right)^{\frac{1}{2}} \leq \left(\sum_{k=1}^{N^2} \frac{|\widehat{\mu}_N(k)|^2}{k^2} + \sum_{k=N^2+1}^{\infty} \frac{1}{k^2}\right)^{\frac{1}{2}}$$

The tail of the series can be easily bounded by $\lesssim \dfrac{1}{N^2}$. For the first terms,

$$\begin{aligned}\sum_{k=1}^{N^2} \frac{|\widehat{\mu}_N(k)|^2}{k^2} &\lesssim \frac{1}{N^2}\sum_{k=1}^{N^2}\frac{1}{k^2\|k\alpha\|^2} \le \frac{1}{N^2}\sum_{l=0}^{\lfloor \log_2 N^2\rfloor}\sum_{k=2^l}^{2^{l+1}}\frac{1}{k^2\|k\alpha\|^2} \le \frac{1}{N^2}\sum_{l=0}^{\lfloor \log_2 N^2\rfloor}\frac{1}{2^{2l}}\sum_{k=2^l}^{2^{l+1}}\frac{1}{\|k\alpha\|^2}\\ &\lesssim \frac{1}{N^2}\sum_{l=0}^{\lfloor \log_2 N^2\rfloor}\frac{1}{2^{2l}}\sum_{m=1}^{2^l}\frac{2^{2l}}{m^2} \le \frac{1}{N^2}\sum_{l=0}^{\lfloor \log_2 N^2\rfloor}\sum_{m=1}^{\infty}\frac{1}{m^2} \lesssim \frac{1}{N^2}\sum_{l=0}^{\lfloor \log_2 N^2\rfloor} 1\\ &= \frac{\lfloor \log_2 N^2\rfloor}{N^2} \lesssim \frac{\log N}{N^2}.\end{aligned}$$

So we get that

$$W_2(\mu_N, dx) \lesssim \left(\frac{\log N}{N^2} + \frac{1}{N^2}\right)^{\frac{1}{2}} \lesssim \frac{\sqrt{\log N}}{N}.$$

□

This upper bound is relatively sharp, as the following classical result gives a close lower bound.

Proposition 5.4. *If we have a sequence* $(x_n) \subseteq [0,1]$ *and define* $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$ *then*

$$W_2(\mu_N, dx) \ge W_1(\mu_N, dx) \ge \frac{1}{4N}.$$

Proof. By the Kantorovich-Rubinstein duality,

$$W_1(\mu_N, dx) = \sup\left\{\left|\int_0^1 f d\mu_N - dx\right| \;:\; \|f\|_{Lip} \le 1\right\}.$$

For a fix N, assume the points are ordered like $0 \le x_1 \le x_2 \le \cdots \le x_N \le 1$, and define $f(x) = d(x, \{x_1, \dots, x_N\}) = \min\{d(x, x_i), d(x, x_1+1), d(x, x_N - 1)\}$, this is, the natural distance when we see the interval $[0,1]$ as the torus. We have then

$$W_1(\mu_N, dx) \ge \left|\int_0^1 f \; d\mu_N - dx\right| = \int_0^1 f \; dx.$$

Finding the minimum of this integral can be see as a easy optimization problem if we consider the new variables:

$$\begin{aligned} i = 1, 2, \dots, N-1, \qquad & y_i = x_{i+1} - x_i, & \int_{x_i}^{x_{i+1}} f \; dx = \frac{1}{2}y_i\frac{y_i}{2} = \frac{1}{4}y_i^2\\ & y_N = x_1 + 1 - x_N, & \int_0^{x^1} f \; dx + \int_{x_N}^1 f \; dx = \frac{1}{4}y_N^2. \end{aligned}$$

where the integrals are computed as the area of a triangle with base y_i and width $y_i/2$.

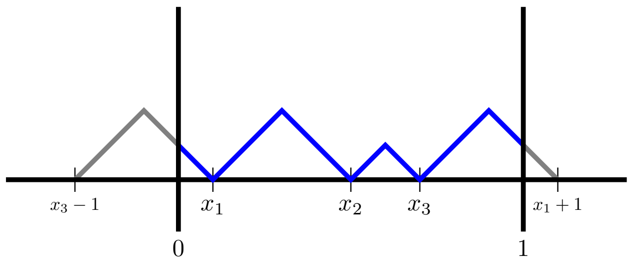


Figure 5.1: In blue, graph of $f(x)$ for 3 points.

So we are interested in the optimization problem

$$\text{Minimize: } F = \sum_{i=1}^{N} \frac{1}{4} y_i^2 \qquad \text{subject to: } \sum_{i=1}^{N} y_i = 1, \quad y_i \geq 0.$$

By using the multipliers of Lagrange, we get that the minimum is attained when exists λ such that $\frac{1}{2} y_i - \lambda = 0$, so all the y_i are equal and $y_i = \frac{1}{N}$. This gives us

$$W_1(\mu_N, dx) \geq F \geq \sum_{i=1}^{N} \frac{1}{4} \left(\frac{1}{N} \right)^2 = \frac{1}{4N}.$$

$\square$

Remark 5.5. *We can use the technics used to bound W_2 to try bound W_1, using now Proposition 5.1 to get the Fourier coefficients. In this case this is not very successful, as we would get*

$$W_1(\mu_N, dx) \lesssim_\alpha \frac{(\log N)^2}{N}$$

which is bigger than if we use the other bound: $W_1(\mu_N, dx) \leq W_2(\mu_N, dx) \lesssim \frac{\sqrt{\log N}}{N^2}$.

Chapter 6
Kantorovich's Formulation as a relaxation

The Kantorovich problem is to look for the minimum of the functional

$$\begin{aligned}\mathbf{K} : \mathrm{Adm}(\mu, \nu) &\longrightarrow \mathbb{R} \cup \{+\infty\} \\ \gamma &\longmapsto \int c \, d\gamma =: \mathbf{K}(\gamma).\end{aligned}$$

On the other hand, for the Monge's problem, we consider a map $T : X \to Y$ and consider the induced plan $\gamma_T = (Id, T)_{\#\mu} \in \mathrm{Adm}(\mu, \nu)$ and try to minimize $\mathbf{K}(\gamma_T)$. We can extend this to a functional defined over the whole set of admisible plans as

$$\mathbf{M}(\gamma) = \begin{cases} \mathbf{K}(\gamma) = \int c \, d\gamma & \text{if } \gamma = \gamma_T \\ +\infty & \text{otherwise} \end{cases}$$

It is clear that the infimum of Monge's formulation is the infimum of $\mathbf{M}$. Now that both formulations are expressed as minimizing a functional over the same set of measures, we can check that $\mathbf{K}$ is the relaxation of $\mathbf{M}$.

Definition 12 **Relaxation as an optimization concept**

Given Ω a metric space and $F : \Omega \to \cup\{+\infty\}$ a functional bounded from below we define its relaxation $\overline{F} : \Omega \to \mathbb{R} \cup \{+\infty\}$ as

$$\overline{F} = \sup \{G : \Omega \to \mathbb{R} \cup \{+\infty\} \ : \ G \leq F \text{ and it is l.s.c.}\}$$

Remark 6.1. *It is immediate to check that* $\inf F = \inf \overline{F}$*. Indeed,* $F \geq \overline{F}$ *implies* $\inf F \geq \inf \overline{F}$*. The other implication is because the constant function* $f \equiv \inf F$ *is l.s.c. and* $f \leq F$*, so* $\overline{F} \geq f$ *and* $\inf \overline{F} \geq \inf f = \inf F$*.*

So we can try to minimize F by minimizing $\overline{F}$. This is useful as the relaxation usually has nicer properties than the original one.

Remark 6.2. *Note that $\overline{F}$ is l.s.c. because it is the supremum of a family of l.s.c. functions. Moreover, we can write it as*

$$\overline{F}(x) = \inf\left\{\liminf_n F(x_n) \ : \ x_n \to x\right\}.$$

Proof. Let us denote $\overline{F}$ to the original definition and $\tilde{F}$ to the new expression.

- As $\tilde{F} \leq F$ and it is l.s.c. we have that $\tilde{F} \leq \overline{F}$.
- For any $x_n \to x$, $\overline{F}(x_n) \leq F(x_n)$ so $\overline{F}(x) = \liminf \overline{F}(x_n) \leq \liminf F(x_n) \leq \tilde{F}(x)$.

□

As the chapter's tittle suggests, the idea to prove that both formulation of the problem have the same infimum is to check that $\mathbf{K}$ is the relaxation of $\mathbf{M}$. We will prove it for measures $\mu \in \mathcal{P}(\mathbb{R}^N)$, $\nu \in \mathcal{P}(\mathbb{R}^M)$ with compact supports and a lower semicontinous cost function. We will also assume that μ is **atomless**, this is,

$$\mu(\{x\}) = 0 \qquad\qquad \forall x \in \operatorname{supp}(\mu).$$

This last assumption is reasonable, as we know that if μ is Dirac delta it may happen that there is no transport map, which we will see that does not happen in this case.

The idea of the proof is that, in this context, the plans induced by transport plans are going to be dense in $\mathrm{Adm}(\mu,\nu)$, and precisely for this kind of plans both functionals coincide. To get to this result we will require several lemmas.

Lemma 6.3. *If $\mu, \nu \in \mathcal{P}(\mathbb{R})$ have compact support and μ atomless then exists a transport map with $T_{\#\mu} = \nu$.*

Proof. In dimension 1, the set of no differentiability of a convex function is at most countable, so it has Lebesgue measure 0. This makes that an atomless measure is regular, and we can apply Brenier's Theorem to μ and get a transport map with $T_{\#\mu} = \nu$. □

Lemma 6.4. *There exists $\sigma_d : \mathbb{R}^d \to \mathbb{R}$ measurable Borel, injective, whose image is a Borel set and with its inverse σ_d^{-1} being Borel measurable too.*

Proof. If we prove it for the case $d = 2$ then we can define by induction

$$\sigma_d(x_1, x_2, \dots, x_d) =: \sigma_2\left(x_1, \sigma_{d-1}(x_2, \dots, x_d)\right).$$

For this case, as

$$\begin{aligned} \mathbb{R}^2 &\longrightarrow (0,1)^2 \\ (x,y) &\longmapsto \left(\frac{1}{2}+\frac{1}{\pi}\arctan x, \frac{1}{2}+\frac{1}{\pi}\arctan y\right) \end{aligned}$$

is a continuous bijection, it is enough to define $\sigma : (0,1)^2 \to \mathbb{R}$ satisfying all the wanted properties for σ_2.

Given $x, y \in (0,1)$ we can consider its decimal expansion, without numbers ending in 9 periodic,

$$x = 0.x_1x_2x_3\cdots = \sum_{i=1}^{\infty} x_i 10^i, \qquad y = 0.y_1y_2y_3\cdots = \sum_{i=1}^{\infty} y_i 10^i.$$

We define then $\sigma(x,y) := 0.x_1y_1x_2y_2\cdots$. Note that numbers like $0.393939\cdots$ are not in the image set $S := \sigma((0,1)^2) \subseteq (0,1)$.

To prove that σ is measurable it is enough to note that for $[a,b] \subseteq (0,1)$ with rational endpoints of the form

$$a = 0.a_1a_2\cdots a_{2n}, \qquad b = 0.b_1b_2\cdots b_{2n}$$

the preimage is a rectangle and therefore Borel measurable:

$$\sigma^{-1}([a,b]) = [0.a_1a_3\cdots a_{2n-1}, 0.b_1b_3\cdots b_{2n-1}] \times [0.a_2a_4\cdots a_{2n}, 0.b_2b_4\cdots b_{2n}].$$

The injectivity of σ is clear due to only allowing 1 periodic expansion for any number. For the measurability of the inverse function, a similar reasoning allows to prove that

$$(\sigma_2^{-1})^{-1}([a,b]\times[c,d]) = S \cap [A,B]$$

So it is only left to prove that S is measurable.

$$\begin{aligned} S &= \left\{0.x_1x_2\cdots \in (0,1) \;:\; \forall n\; \exists m,m' > n,\;\; m \text{ odd, } m' \text{ even } x_m \neq 9,\; x_{m'} \neq 0\right\} \\ (0,1)\backslash S &= \{0.x_1x_2\cdots \in (0,1)\; \exists N \text{ such that } \forall n \geq N \quad x_{2n} = 9 \text{ or } x_{2n+1} = 9\} \end{aligned}$$

Naturally, this last set we can split it in two sets depending wether the periodic 9 is in the even positions or in the odd ones: $(0,1)\backslash S = A_{odd} \cup A_{even}$. But then

$$A_{even} = \cup_{N=1}^{\infty} \{0.x_1x_2\cdots x_{2N-1}9x_{2N+1}9x_{2N+3}\cdots \in (0,1)\} = \cup_{N=1}^{\infty} B_N$$

As for any N, there are countably many point in B_N, this is a Borel set, which in turns implies that A_{even} is Borel. The same reasoning works for A_{odd}, and then we can get that S is Borel. $\square$

Corollary 6.5. *Given $\mu \in \mathcal{P}(\mathbb{R}^N)$, $\nu \in \mathcal{P}(\mathbb{R}^M)$ both with compact support and μ atomless there exists a Transport map $T_{\#\mu} = \nu$.*

Proof. As $(\sigma_N)_{\#\mu}, (\sigma_M)_{\#\nu} \in \mathcal{P}(\mathbb{R})$ have both compact support and the first one is atomless we can apply Lemma 6.3 and the exists $S_{\#(\sigma_N)_{\#\mu}} = (\sigma_M)_{\#\nu}$. Then $T = \sigma_M^{-1} \circ S \circ \sigma_N$ is the wanted map:

$$\begin{aligned} T_{\#\mu}(A) &= \mu(T^{-1}(A)) = \mu\left(\sigma_N^{-1} \circ S^{-1} \circ \sigma_M(A)\right) = (\sigma_N)_{\#\mu}\left(S^{-1} \circ \sigma_M(A)\right) \\ &= S_{\#(\sigma_N)_{\#\mu}}\left(\sigma_M(A)\right) = (\sigma_M)_{\#\nu}\left(\sigma_M(A)\right) = \nu(\sigma_M^{-1} \circ \sigma_M(A)) = \nu(A). \end{aligned}$$

□

Lemma 6.6. *Let X be a compact metric space, $\rho \in \mathcal{P}(X)$ and G_n a sequence of partitions: $C_{i,n}$ disjoint with $\cup_{i \in I_n} C_{i,n} = X$. Call $a_n = \sup_{i \in I_n} diam(C_{i,n})$ and assume that $a_n \to 0$.*

If we have a sequence of probability measures $\rho_n \in \mathcal{P}(X)$ with $\rho_n(C_{i,n}) = \rho(C_{i,n})$ for all $i \in I_n$, $n \geq 1$, then $\rho_n \to \rho$ narrowly.

Proof. Given $\varphi : X \to \mathbb{R}$ continuous, as X is compact, φ is then bounded and absolutely continuous. We have the bound $\left|\int \varphi d\rho_n - \int \varphi d\rho\right| \leq \sum_{i \in I_n} \left|\int_{C_{i,n}} \varphi d\rho_n - \int_{C_{i,n}} \varphi d\rho\right|$.

Using the absolute continuity we have that for every $\varepsilon > 0$ exists n such that if $x, y \in C_{i,n}$ then $|\varphi(x) - \varphi(y)| < \frac{\varepsilon}{2}$. Pick a point in each set $x_i \in C_{i,n}$, so we have $\varphi(x_i) - \frac{\varepsilon}{2} < \varphi(x) < \varphi(x_i) + \frac{\varepsilon}{2}$ for $x \in C_{i,n}$. As $\rho(C_{i,n}) = \rho_n(C_{i,n})$ we get the inequality

$$\left|\int_{C_{i,n}} \varphi \, d\rho_n - \int_{C_{i,n}} \varphi \, d\rho\right| \leq \varepsilon\rho(C_{i,n}) \qquad \Rightarrow \qquad \left|\int \varphi d\rho_n - \int \varphi d\rho\right| \leq \sum_{i \in I_n} \varepsilon\rho(C_{i,n}) = \varepsilon.$$

This proves that $\int \varphi d\rho_n \to \int \varphi \, d\rho$, so we have $\rho_n \xrightarrow{\text{nrw}} \rho$. □

Theorem 6.7

Given $X \subseteq \mathbb{R}^N$, $Y \subseteq \mathbb{R}^M$ compact subsets and probabilities $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ then the set of transport plans induced by a map is dense in $\mathrm{Adm}(\mu, \nu)$.

Proof. For any $n \in \mathbb{N}$ consider $(K_{i,n})_{i \in I_n}$ a partition by cubes of X with $diam(K_{i,n}) \leq \frac{1}{2n}$, and one $\tilde{K}_{j,n}$ of Y. Then $C_{i,j,n} = K_{i,n} \times K_{j,n}$ is a partition of $X \times Y$ with $diam(C_{i,j,n}) < \frac{1}{n}$.

Given $\gamma \in \mathrm{Adm}(\mu, \nu)$, consider $Col_{i,n} := K_{i,n} \times Y$ and $\gamma_{i,n}$ the restriction to $Col_{i,n}$ of γ. This measure has marginals $\mu_{i,n}$ and $\nu_{i,n}$ with compact support and $\mu_{i,n}$ being atomless, so it exists

$(T_{i,n})_{\#\mu_{i,n}} = \nu_{i,n}$. Because the $Col_{i,n}$ form a partition we can define

$$S_n(x) = T_{i,n} \text{ if } x \in Col_{i,n}.$$

Which is clear that sends μ to ν: $(S_n)_{\#\mu} = \nu$.

If we apply $\gamma_{S_n}(C_{i,j,n}) = \gamma(C_{i,j,n})$ then we can apply Lemma 6.6 to get that $\gamma_{S_n} \xrightarrow{\text{nrw}} \gamma$ and complete the proof. But this inequality is a result precisely of how we have defined $T_{i,n}$:

$$\gamma_{S_n}(C_{i,j,n}) = \gamma_{T_{i,n}}(K_{i,n} \times \tilde{K}_{j,n}) = \mu_{i,n}(T_{i,n}^{-1}(\tilde{K_{j,n}})) = \nu_{i,n}(\tilde{K}_{j,n}) = \gamma(K_{i,n} \times \tilde{K}_{j,n}).$$

□

Corollary 6.8

When $\mu \in \mathcal{P}(\mathbb{R}^N)$, $\nu \in \mathcal{P}(\mathbb{R}^M)$ have compact support and the cost function is l.s.c. then $\mathbf{K}$ is the relaxation of $\mathbf{M}$. Notably, $\inf \mathbf{M} = \min \mathbf{K}$.

Proof. Because of Theorem 1.5 we know that $\mathbf{K}$ is l.s.c. We also have that $\mathbf{K} \leq \mathbf{M}$ so $\mathbf{K} \leq \overline{\mathbf{M}}$. For the other implication, given any $\gamma \in \mathrm{Adm}(\mu, \nu)$ exists $\gamma_{T_n} \to \gamma$ narrowly so

$$\overline{\mathbf{J}}(\gamma) \leq \liminf_n \mathbf{M}(\gamma_{T_n}) \leq \liminf_n \mathbf{K}(\gamma_{T_n}) = K(\gamma).$$

□

Appendix A
Analysis inequalities

The Koksma-Hlawka inequality gives an upper bound to the difference between a Riemann sum and the actual integral. For this it defines the following concept, the discrepancy. This appendix will give a brief summary of this result, for a more complete text one can read [8, Chapter 2], sections 1 and 5. The paper [2] also gives a nice summary.

Given N points $0 \leq x_1 \leq x_2 \leq \cdots \leq x_N \leq 1$

$$D^*((x_k)) = \sup \left\{ \left| \frac{1}{N} \sum_{k=1}^{N} \chi_{[0,t]}(x_j) - t \right| \; : \; 0 \leq t \leq 1 \right\}.$$

Note that that if we call $\nu = \sum_{k=1}^{N} \delta_{x_k}$ we are just taking the supremum of $\nu([0,t]) - \mathcal{L}([0,t])$.

We also consider

$$D((x_k)) = \sup \left\{ \left| \frac{1}{N} \sum_{k=1}^{N} \chi_I(x_k) - \mathcal{L}(I) \right| \; : \; I \subseteq [0,1] \text{ interval} \right\}.$$

It is clear that $D^* \leq D$, but we also have that $D \leq 2D^*$ (check [8, Chapter 2, Theorem 1.3]).

Theorem A.1 **Koksma-Hlawka inequality**

If f has bounded variation then

$$\left| \frac{1}{N} \sum_{k=1}^{N} f(x_k) - \int_0^1 f(x)dx \right| \leq var(f) \cdot D^*((x_k)) \leq var(f) \cdot D((x_k)).$$

The proof can be consulted on [8, Chapter 2, Theorem 5.1].

The variation of a function is defined as

$$var(f) = \sup \left\{ \sum_{i=1}^{N-1} |f(x_{i+1}) - f(x_i)| \ : \ 0 \leq x_1 < x_2 < \cdots < x_N \leq 1 \right\}$$

The Erdős-Turan inequality is another classical result, that bounds the discrepancy.

Theorem A.2 **Erdős-Turan inequality**

Given N points $0 \leq x_1 \leq x_2 \leq \cdots \leq x_N \leq 1$ let $\nu = \sum_{k=1}^{N} \delta_{x_k}$. Then for any $n \in \mathbb{N}$

$$D((x_k)) \lesssim \frac{1}{n} + \sum_{k=1}^{n} \frac{|\widehat{\nu}(k)|}{k}$$

The proof can be consulted on [8, Chapter 2, Theorem 2.5 and comment at the end of page 114]

Bibliography

[1] Luigi Ambrosio and Nicola Gigli. *A User's Guide to Optimal Transport*. Springer Berlin Heidelberg, 2013.

[2] L. Brandolini, L. Colzani, G. Gigante, and G. Travaglini. On koksma-hlawka inequality, 2011.

[3] Laura Caravenna. *The Disintegration Theorem and Applications to Optimal Mass Transport*. PhD thesis, International School of Advanced Studies, 2009.

[4] Villani Cédric. *Optimal transport: Old and new*. Springer Berlin Heidelberg, 2009.

[5] J. T. Chang and D. Pollard. Conditioning as disintegration. *Statistica Neerlandica*, 51(3):287–317, 1997.

[6] Michael G. Crandall, Hitoshi Ishii, and P. L. Lions. *User's Guide to viscosity solutions of second order partial differential equations*. American Mathematical Society, 1992.

[7] Piotr Hajłasz. On an old theorem of erdös about ambiguous locus, 2021.

[8] Lauwerens Kuipers and Harald Niederreiter. *Uniform distribution of sequences*. Dover, 2006.

[9] Jan K. Pachl. Disintegration and compact measures. *MATHEMATICA SCANDINAVICA*, 43:157–168, 1978.

[10] Aldo Pratelli. On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation. *Annales de l'I.H.P. Probabilités et statistiques*, 43(1), 2007.

[11] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of variations, pdes, and modeling*. 2015.

[12] Stefan Steinerberger. Wasserstein distance, fourier series and applications. *Monatshefte für Mathematik*, 194(2):305–338, Jan 2021.

[13] Onno van Gaans. Probability measures on metric spaces. Avalible online.

[14] Luděk Zajíček. On the differentiation of convex functions in finite and infinite dimensional spaces. *Czechoslovak Mathematical Journal*, 29(3):340–348, 1979.

Alphabetical Index