Metrics on Spaces of Measures

Wouter Slegers

Bachelor thesis

Supervisors: dr. S.C. Hille,
M. Ziemlanska

Date: July 9, 2016

Mathematisch Instituut, Universiteit Leiden

Contents

Introduction

About Metrics on Spaces of Measures

We will investigate metrics that are defined on spaces of measures. Some of those metrics one can define only on probability measures, an example of this are the Wasserstein metrics. Other metrics one can define on the larger space of finite signed measures and are given by a norm on this vector space. Some of the metrics defined on probability measures are in essence a restriction of such metrics, derived from a norm on the space of finite signed measures, some of them are not. The main point of study in this thesis are the Wasserstein metrics. These metrics are constructed using only measure theoretic definitions. One of those metrics we will study in greater detail, namely the Wasserstein metric of order 1, otherwise referred to as the Kantorovich distance. We will take a look at the Kantorovich-Rubinstein Theorem, which tells us that the Kantorovich distance is equal to a metric that has the structure of a metric derived from a dual norm on the space of Lipschitz functions. Both of these metrics are defined only on probability measures, but we find an extension to the space of finite signed measures.
We wish to know whether, more generally, the Wasserstein metrics of order p, where $p > 1$, are in essence such a restriction too. We investigate if a proof similar to that of the Kantorovich-Rubinstein Theorem is possible for these more general metrics. This extension of the theorem seems new in literature, which suggests such an extension may not be possible. We carry the adapted proof as far as we can and pinpoint where it seems to stop working.

An overview of this thesis

The reader is assumed to have a general knowledge of measure theory and functional analysis.

In the first chapter we give all definitions and lemmas to form a sufficient basis for the rest of the thesis. These definitions concern Lipschitz continuous functions which play a crucial role in the Kantorovich-Rubinstein Theorem. Furthermore we will discuss finite signed measures and their connection with the Lipschitz continuous functions. In the last section of this chapter we give an extension of the Kantorovich norm. This chapter is largely based on [7], some aspects worked out in further detail than found in [7], complemented with some additional results.

In the second chapter we introduce the transportation problem, we state the Kantorovich Duality Theorem and our main focus in this chapter is giving a rigorous proof of the Kantorovich-Rubinstein Theorem. This chapter is mainly based on the first chapter of [9], and it is the proof of the Kantorovich-Rubinstein Theorem given there that we work out in detail.

The focus in the third chapter lies in applying the proof of the Kantorovich-Rubinstein Theorem to Wasserstein metrics of order p where $p > 1$, giving some solutions for problems we encounter and identifying in which step the proof stops to work.

In the Appendix we provide a note on lower semi-continuity. The first part of this chapter is based on [9, Remark 1.1.7.4] and the second part is based on [8, Theorem 3].

About this thesis

In this thesis we give an extension of the Kantorovich norm, defined on the space of probability measures, to the space of finite signed measures. Various extensions exist, such as the one given by Bogachev in [3, p. 234] and the one given by Hanin in [6]. Both extensions seem slightly artificial, the extension given in this thesis somewhat less so. This extension then also gives an extension for the metric derived from the Kantorovich norm, which we define on probability measures, to the space of finite signed measures.
The proof of the Kantorovich-Rubinstein Theorem we base on the one given by Villani in [9], we filled various gaps in reasoning and repaired some mistakes. For one [9, Remark 1.12] brushes over a few statements that take some work to formulate rigorously. The part given as an exercise in this remark actually turns out to be incorrect, see Example 0.1. This remark we split into parts, correct and work out in Chapter 2 of this thesis.
For each statement that is made in Chapter 2 we put extra care into determining the requirements that are needed, for the statement to hold. This streamlines the process of applying the statements to the more general case, that of the Wasserstein metrics of order p, since then we can conveniently distinguish which statements still hold, which do not and why. One useful addition in that respect are the c_X and c_Y functions defined in Remark 2.6, that depend on the cost function c we use. In the original statement, given in [9], it was required that c is bounded. We need only that these functions c_X and c_Y are bounded, which is a more general case. We use this increased generality to simplify the proof of the Kantorovich-Rubinstein Theorem and it helps when investigating the possibility of an extension of the Kantorovich-Rubinstein Theorem in Chapter 3.

Example 0.1
Let X and Y be Polish spaces. Let $c\colon X \times Y \to \mathbb{R}^+$ a measurable, bounded and lower semi-continuous function. Then there exists a nondecreasing sequence of nonnegative and uniformly continuous functions c_l converging pointwise to c. Let $\varphi\colon X \to \mathbb{R}$ be bounded and define

$$\varphi^c(y) := \inf_{x\in X}[c(x,y) - \varphi(x)], \quad \text{and for any } l \in \mathbb{N} \text{ define} \quad \psi_l(y) := \inf_{x\in X}[c_l(x,y) - \varphi(x)].$$

In [9, Remark 1.12] it was given as an exercise to prove that ψ_l converges pointwise to φ^c.
The following counter-example was taken from `http://math.stackexchange.com/questions/1270630` with slight modification to increase generality.
Suppose there exists a point $e \in X$ such that e is not isolated. Take the bounded function given by

$$\varphi(x) := \begin{cases} 0, & \text{if } x = e \\ 1, & \text{otherwise} \end{cases}$$

and for all $(x,y) \in X \times Y$ take $c(x,y) := \varphi(x)$. Clearly $\varphi^c(y) = \inf_{x\in X}[c(x,y) - \varphi(x)] = 0$ for all $y \in Y$. Suppose that c_l is a continuous function such that $c_l \leq c$. Take a sequence $(x_n)_{n\in\mathbb{N}} \subset X$ such that $x_n \neq e$ for all $n \in \mathbb{N}$ and that $x_n \to e$. Then for any $y \in Y$ we have

$$\psi_l(y) = \inf_{x\in X}[c_l(x,y) - \varphi(x)] \leq \lim_{k\to\infty}[c_l(x_n,y) - \varphi(x_n)] = c_l(e,y) - 1 \leq c(e,y) - 1 = -1.$$

Therefore $\psi_l(y) \leq -1$ for all $y \in Y$, so the sequence ψ_l will not converge to φ^c.

Chapter 1

Lipschitz and measure spaces

The first section of this chapter is mainly focused on introducing the required definitions and notations and it contains a few basic lemmas. Section two connects some of the definitions introduced here. The last section gives an extension of the Kantorovich norm, also introduced in the first section.

1.1 A few functions, spaces and norms

We start by defining the space of Lipschitz functions. This will play a crucial role in this thesis.

Definition 1.1
Let (S, d) be a metric space then define

$$\mathrm{Lip}(S) := \{f\colon S \to \mathbb{R} \mid \exists L \in \mathbb{R} \text{ such that } \forall x, y \in S\colon |f(x) - f(y)| \leq L d(x, y)\}\,.$$

The defining property of the space $\mathrm{Lip}(S)$ is called the Lipschitz property. Intuitively, a Lipschitz function is a function whose slope will never be bigger than a certain value. In Definition 1.1 this value is L and we have

$$\text{for all } x, y \in S \text{ such that } x \neq y : \frac{|f(x) - f(y)|}{d(x, y)} \leq L.$$

Remark 1.2: A Lipschitz function is often called Lipschitz continuous. This makes sense since Lipschitz continuity is actually a stronger requirement than uniform continuity which in turn is stronger than continuity. Namely let $f \in \mathrm{Lip}(S)$ and take L from Definition 1.1. For any $\epsilon > 0$ choose $\delta := \frac{\epsilon}{L}$, this gives for any $x, y \in S$ satisfying $d(x, y) \leq \delta$ that $|f(x) - f(y)| \leq L d(x, y) < \epsilon$ proving uniform continuity.

For each Lipschitz function we call its maximal slope the Lipschitz constant of f. This is the minimal constant L such that $f \in \mathrm{Lip}(S)$ satisfies the Lipschitz property. It is defined as follows.

Definition 1.3
For $f \in \mathrm{Lip}(S)$ we define

$$|f|_L := \sup_{\substack{x,y \in S \\ x \neq y}} \frac{|f(x) - f(y)|}{d(x, y)}.$$

The Lipschitz constant gives a seminorm on $\mathrm{Lip}(S)$. It is not a norm because $|f|_L = 0$ if and only if f is constant.

Lemma 1.4
A finite sum of Lipschitz functions is again Lipschitz. Actually let $f_1, \ldots, f_N \in \operatorname{Lip}(S)$ *then*

$$f := \sum_{n=1}^{N} f_n \text{ is Lipschitz with } |f|_L \leq \sum_{n=1}^{N} |f_n|_L .$$

Proof. We find for every $x, y \in S$, by the Lipschitz property, that

$$|f(x) - f(y)| \quad = \quad \left| \sum_{n=1}^{N} f_n(x) - \sum_{n=1}^{N} f_n(y) \right| \leq \sum_{n=1}^{N} |f_n(x) - f_n(y)| \leq \sum_{n=1}^{N} |f_n|_L \, d(x, y).$$

Since $L := \sum_{n=1}^{N} |f_n|_L < \infty$ we have $L \in \mathbb{R}$ and that means that f is Lipschitz with $|f|_L \leq L$. □

Lemma 1.5 ([5, Lemma 4])
For $f_1, \ldots, f_n \in \operatorname{Lip}(S)$ *we take*

$$g(x) := \min_{1 \leq i \leq n} f_i(x) \quad \text{and} \quad h(x) := \max_{1 \leq i \leq n} f_i(x).$$

Then $g, h \in \operatorname{Lip}(S)$ *and*

$$\max(|g|_L, |h|_L) \leq \max_{1 \leq i \leq n} |f_i|_L .$$

Proof. We prove the case where $n = 2$. We then get the general case by induction.
Let $f_1, f_2 \in \operatorname{Lip}(S)$, take $g := \min(f_1, f_2)$ and take $M = \max(|f_1|_L, |f_2|_L)$. Let $x, y \in S$. We get

$$\max(|f_1(x) - f_1(y)|, |f_2(x) - f_2(y)|) \leq M d(x, y). \tag{1.1}$$

This proves

$$|g(x) - g(y)| \leq M d(x, y),$$

if $g(x) = f_1(x)$ and $g(y) = f_1(y)$ and if $g(x) = f_2(x)$ and $g(y) = f_2(y)$.
If $g(x) = f_1(x)$ and $g(y) = f_2(y)$, then $f_1(x) \leq f_2(x)$ and $f_2(y) \leq f_1(y)$. Hence we get

$$f_1(x) - f_1(y) \leq f_1(x) - f_2(y) \leq f_2(x) - f_2(y),$$

so by (1.1) we get $|f_1(x) - f_2(y)| \leq M d(x, y)$. By exchanging x and y in the preceding argument we also prove this for the remaining case.
By definition for a function $f \in \operatorname{Lip}(S)$ it follows that $-f \in \operatorname{Lip}(S)$ with $|-f|_L = |f|_L$. Hence for $h = \max(f_1, f_2) = -\min(-f_1, -f_2)$ we also get $h \in \operatorname{Lip}(S)$ with $|h|_L \leq M$. □

Lemma 1.6
Let $\emptyset \neq A \subset S$. *Then for any* $x, y \in S$ *the following holds*

$$|d(x, A) - d(y, A)| \leq d(x, y).$$

Proof. If $x \in A$ or $y \in A$ then this is clear. Assume $x, y \notin A$ then let $(x_n)_{n \in \mathbb{N}} \subset A$ such that $\lim_{n \to \infty} d(x, x_n) = d(x, A)$. For any $n \in \mathbb{N}$ we find $d(y, A) \leq d(y, x_n) \leq d(x, x_n) + d(x, y)$ hence $d(y, A) - d(x, x_n) \leq d(x, y)$. Now in the limit $n \to \infty$, we get $d(y, A) - d(x, A) \leq d(x, y)$. From symmetry we also get $d(x, A) - d(y, A) \leq d(x, y)$, proving the claim. □

Remark 1.7: Note that Lemma 1.6 proves that $x \mapsto d(x, A)$ is a Lipschitz function and that $|d(\cdot, A)|_L \leq 1$. Note also that this makes $x \mapsto d(x, y)$ Lipschitz continuous for any $y \in S$.

We call a function $f \in \mathrm{Lip}(S)$ with $|f|_L \leq 1$ a 1-Lipschitz function.
There are a few commonly used norms on $\mathrm{Lip}(S)$ related to the seminorm $|\cdot|_L$. One norm on $\mathrm{Lip}(S)$ was introduced in [7]. It is given in the following definition.

Definition 1.8
Let $e \in S$. We define the norm $\|\cdot\|_e$ on $\mathrm{Lip}(S)$ by

$$\|f\|_e := |f(e)| + |f|_L.$$

Let $e \in S$. For any $f \in \mathrm{Lip}(S)$ we have, by the Lipschitz property, for all $x \in S$ that

$$|f(x)| \leq |f(x) - f(e)| + |f(e)| \leq |f|_L \, d(x,e) + |f(e)| . \tag{1.2}$$

For $e, e' \in S$ we find by (1.2)

$$\begin{aligned} \|f\|_e &= |f(e)| + |f|_L \leq |f|_L d(e,e') + |f(e')| + |f|_L = |f|_L(1 + d(e,e')) + |f(e')| \\ &\leq (|f|_L + |f(e')|)\,(1 + d(e,e')) = \|f\|_{e'}\,(1 + d(e,e')). \end{aligned}$$

Consequently $\|\cdot\|_e$ and $\|\cdot\|_{e'}$ are equivalent. From now on we take $e \in S$ fixed and denote the normed space $\mathrm{Lip}(S)$ with norm $\|\cdot\|_e$ as $\mathrm{Lip}_e(S)$.
Let $\mathrm{BL}(S)$ denote the space of bounded Lipschitz maps.

Definition 1.9
We define the norm $\|\cdot\|_{BL}$ on $\mathrm{BL}(S)$ by

$$\|f\|_{\mathrm{BL}} := \|f\|_\infty + |f|_L.$$

We write $\|\cdot\|_e^*$ for the dual norm of $\|\cdot\|_e$ on $\mathrm{Lip}_e(S)^*$ i.e. for $\phi \in \mathrm{Lip}_e(S)^*$ we have

$$\|\phi\|_e^* := \sup\left\{|\phi(f)| : f \in \mathrm{Lip}_e(S), \|f\|_e \leq 1\right\}.$$

We will write $\|\cdot\|_{\mathrm{BL}}^*$ for the dual norm of $\|\cdot\|_{\mathrm{BL}}$ on $\mathrm{BL}(S)^*$.

Let $\mathcal{P}(S)$ denote the space of Borel probability measures on S.
Let $\mathcal{M}(S)$ denote the *finite signed Borel measures* on S, i.e. all $\mu\colon \mathcal{B} \to \mathbb{R}$ satisfying only the null empty set and countable additivity requirements, not the nonnegativity requirement that would make μ an ordinary measure. Note that $\mathcal{B}$ is the set of Borel sets and that we used $\mathbb{R}$ not $\overline{\mathbb{R}}$ because we want μ to be finite: $-\infty < \mu(S) < \infty$.
Let $\mathcal{M}^+(S)$ denote the subset of $\mathcal{M}(S)$ that consist of *finite positive measures.*
For every measure $\mu \in \mathcal{M}(S)$ there exists a unique decomposition $\mu = \mu^+ - \mu^-$ where $\mu^+, \mu^- \in \mathcal{M}^+(S)$. This is called the Jordan decomposition.
We let $\mathcal{M}_1(S)$ be the set of finite signed measures with *finite first moment*, i.e.

$$\mathcal{M}_1(S) := \left\{\mu \in \mathcal{M}(S) : \int_S d(x,e) d|\mu|(x) < \infty\right\}.$$

Similarly, let $\mathcal{P}_1(S) \subset \mathcal{M}_1(S)$ be the set of probability measures with finite first moment.

We write $\mathcal{M}_1^0(S)$ for the subspace of $\mathcal{M}_1(S)$ consisting of $\mu \in \mathcal{M}_1(S)$ such that $\mu(S) = 0$.
For any (positive or signed) measure μ we write $\mathcal{L}^1(d\mu)$ for the space of all measurable functions $f\colon S \to \mathbb{R}$ that are μ-integrable.
We will write $\|\cdot\|_{\mathrm{TV}}$ for the total variation norm on $\mathcal{M}(S)$, i.e. for $\mu \in \mathcal{M}(S)$ we have

$$\|\mu\|_{\mathrm{TV}} := |\mu|(S) = \mu^+(S) + \mu^-(S).$$

We write $\|\cdot\|_1$ for the norm on $\mathcal{M}_1(S)$ given by

$$\|\mu\|_1 := \int_S \max(1, d(x,e)) d|\mu| .$$

Lemma 1.10
Let $f \in \mathrm{Lip}_e(S)$ and $\mu \in \mathcal{M}_1(S)$. Then f is μ-integrable.

Proof. Since f is Lipschitz continuous it is continuous hence (Borel-)measurable. By (1.2) it follows that

$$\int_S |f(x)|\, d|\mu| \leq \int_S \big[|f(e)| + |f|_L\, |d(x,e)|\big] d|\mu| = |f(e)|\, \|\mu\|_{\mathrm{TV}} + |f|_L \int_S d(x,e) d|\mu| < \infty$$

holds since μ has finite first moment. □

For a function f we will write $[f(x)]^+$ for $\max(f(x), 0)$. We will now introduce a function that will find various uses throughout this thesis.

Lemma 1.11
For $\emptyset \neq A \subset S$ and $n \in \mathbb{N}$ we define the function $f_{n,A} \colon S \to [0,1]$ by $f_{n,A}(x) := [1 - nd(x,A)]^+$. Then $f_{n,A} \in \mathrm{Lip}(S)$ with $|f_n|_L \leq n$. If A^c contains a point x such that $0 < d(x,A) \leq \frac{1}{n}$, then we have $|f_{n,A}|_L = n$. Moreover, $f_{n,A}$ converges to $\mathbb{1}_A$ pointwise if and only if A is closed.

Proof. From Remark 1.7 we get that $x \mapsto d(x,A)$ is Lipschitz continuous with $|d(\cdot,A)|_L \leq 1$, hence from Lemma 1.4 and Lemma 1.5 we get that $f_{n,A}$ is Lipschitz with $|f_{n,A}|_L \leq n$.
If there exists an $x \in S$ such that $0 < d(x,A) \leq \frac{1}{n}$, let $(x_k)_{k\in\mathbb{N}} \subset A$ such that $\lim_{k\to\infty} d(x,x_k) = d(x,A)$, then

$$\frac{|f_{n,A}(x) - f_{n,A}(x_k)|}{d(x,x_k)} = \frac{|[1 - nd(x,A)] - 1|}{d(x,x_k)} = \frac{nd(x,A)}{d(x,x_k)} \to n$$

as $k \to \infty$. So we find that $|f_{n,A}|_L = n$.
For $x \in \overline{A}$ we have $d(x,A) = 0$ hence for all $n \in \mathbb{N}$ we have $f_{n,A}(x) = 1$. For $x \notin \overline{A}$ we have $d(x,A) > 0$ so we get $f_{n,A}(x) = [1 - nd(x,A)]^+ = 0$ for n sufficiently large. That means that $f_{n,A}$ converges pointwise to $\mathbb{1}_A$ if and only if $A = \overline{A}$, i.e. if and only if A is closed. □

Definition 1.12 (Kantorovich norm)
For $\sigma \in \mathcal{M}_1^0(S)$ we define

$$\|\sigma\|_{KR} = \sup\left\{ \int_S f d\sigma : f \in \mathrm{Lip}(S), |f|_L \leq 1 \right\}.$$

Lemma 1.13
The function $\|\cdot\|_{KR}$ defines a norm on $\mathcal{M}_1^0(S)$.

Proof. Let $\sigma \in \mathcal{M}_1^0(S)$ such that $\|\sigma\|_{KR} = 0$. We will prove that $\sigma = 0$. Let $C \subset S$ be closed. For any $n \in \mathbb{N}$ we define $f_n(x) := [1 - nd(x,C)]^+$. From Lemma 1.11 we obtain that f_n is Lipschitz and that the pointwise limit of f_n is $\mathbb{1}_C$ since C is closed. There does not exist an $n \in \mathbb{N}$ such that $\int_S f_n d\sigma \neq 0$. Otherwise we have

$$\int_S \frac{f_n}{n} d\sigma \neq 0, \text{ whilst } \left|\frac{f_n}{n}\right|_L \leq 1,$$

as $|f_n|_L \leq n$, contradicting that $\|\sigma\|_{KR} = 0$. Hence we get that

$$\sigma(C) = \int_S \mathbb{1}_C d\sigma = \int_S \lim_{n\to\infty} f_n d\sigma = \lim_{n\to\infty} \int_S f_n d\sigma = 0$$

by Lebesgue's Dominated Convergence Theorem, using that $|f_n| \leq 1$ for all $n \in \mathbb{N}$ and that the constant function 1 is σ-integrable since σ is finite. Since the closed sets generate the Borel sets, this proves that $\sigma = 0$. It is clear that if $\sigma = 0$ we get $\|\sigma\|_{KR} = 0$. The remaining properties follow easily. □

Remark 1.14: The norm $\|\cdot\|_{KR}$ can only be defined on $\mathcal{M}_1^0(S)$ and not on $\mathcal{M}_1(S)$, see Remark 1.15. Extensions of this norm to the space $\mathcal{M}_1(S)$ exist, one is given by Hanin in [6] and another is given by Bogachev in [3, p. 234]. We will work with yet another extension of this norm, which we will provide in Section 1.3.

Remark 1.15: The reason we cannot define $\|\cdot\|_{KR}$ on all of $\mathcal{M}_1(S)$ is the following. Let $\sigma \in \mathcal{M}_1(S)$ such that $\sigma(S) \neq 0$. We have that for any $n \in \mathbb{Z}$ the function $f_n \colon S \to \mathbb{R}$ given by $f_n \equiv n$ is in $\mathrm{Lip}(S)$ and furthermore $|f_n|_L = 0 \leq 1$ since it is constant. We then find

$$\int_S f_n d\sigma = n\sigma(S) \to \infty$$

by letting $n \to \infty$ if $\sigma(S) > 0$ and $n \to -\infty$ if $\sigma(S) < 0$. This means that the supremum, from Definition 1.12, does not exist. Note that this seems forgotten in [9, p.35].

Definition 1.16 (The d_{KR} metric)
On $\mathcal{P}_1(S)$ we define the metric d_{KR} by

$$d_{KR}(\mu, \nu) := \|\mu - \nu\|_{KR}.$$

Remark 1.17: For Definition 1.16 we can use the norm $\|\cdot\|_{KR}$ to define d_{KR}, since for two probability measures $\mu, \nu \in \mathcal{P}_1(S)$ we have $\mu - \nu \in \mathcal{M}_1^0(S)$.
At this point, the metric d_{KR} on the convex set $\mathcal{P}_1(S)$ does not (yet) derive from a norm on a vector space enveloping $\mathcal{P}_1(S)$, i.e. $d_{KR} = \|\mu - \nu\|$ for some norm $\|\cdot\|$ on $\mathcal{M}_1(S)$. We know $\|\cdot\|_{KR}$ is not a norm on $\mathcal{M}_1(S)$ by Remark 1.15. In Section 1.3, Theorem 1.26 in particular, we shall show that d_{KR} derives from a norm on $\mathcal{M}_1(S)$ indeed. Despite the norm that defines d_{KR} being only defined on $\mathcal{M}_1^0(S)$, not on $\mathcal{P}_1(S)$, we do get that d_{KR} is a metric. The proof is very similar to how one would prove that from any norm one can derive a metric when both are defined on the same space.

Let the subscript KR (which stands for Kantorovich-Rubinstein), and the fact that d_{KR} is derived from the Kantorovich norm, not let you confuse this metric with the Kantorovich distance which has a quite different definition that we will give later on.

Remark 1.18: Any Borel measure μ on a metric space is inner regular [3, Theorem 7.1.7], which means that for any Borel set $A \subset S$ we have

$$\mu(A) = \sup\{\mu(K) : K \text{ closed}, K \subset A\}.$$

Lemma 1.19
Let $\mu \in \mathcal{M}(S)$. The set $\mathrm{BL}(S)$ lies dense in $\mathcal{L}^1(d\mu)$ with respect to the seminorm $\|\cdot\|_{\mathcal{L}^1}$, i.e. for every $f \in \mathcal{L}^1(d\mu)$ there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathrm{BL}(S)$ such that

$$\lim_{n\to\infty} \|f - f_n\|_{\mathcal{L}^1} = \lim_{n\to\infty} \int_S |f - f_n|\, d|\mu| = 0. \tag{1.3}$$

Proof. Lipschitz continuous functions are continuous hence (Borel-)measurable. Any bounded measurable function is μ-integrable since μ is finite. Hence we have $\mathrm{BL}(S) \subset \mathcal{L}^1(d\mu)$.
Since for any $\mu \in \mathcal{M}(S)$ a function f is μ-integrable if and only if f is $|\mu|$-integrable and since we have $|\mu| \in \mathcal{M}^+(S)$, we will, for proving (1.3), without loss of generality, take μ to be a finite positive measure. Let $g\colon S \to [0,\infty]$ be μ-integrable. Then there exists a nondecreasing sequence φ_n of nonnegative step-functions converging pointwise to g.
Let $n \in \mathbb{N}$. We can write the step-function as $\varphi_n = \sum_{i=1}^{N_n} \alpha_i \mathbb{1}_{A_i}$ where A_i are pairwise disjoint Borel sets, $\alpha_i \in \mathbb{R}^+$ and $N_n \in \mathbb{N}$. Take $1 \leq i \leq N_n$, we will consider the function $\alpha_i \mathbb{1}_{A_i}$.
By Remark 1.18 we can take a sequence $(C_{k,i})_{k\in\mathbb{N}}$ such that for each $k \in \mathbb{N}$ we have that $C_{k,i} \subset A_i$ is closed and that $\lim_{k\to\infty} \mu(C_{k,i}) = \mu(A_i)$. For each $k \in \mathbb{N}$ we have that $\mathbb{1}_{C_{k,i}} \leq \mathbb{1}_{A_i}$ so this gives

$$\lim_{k\to\infty} \int_S \left|\mathbb{1}_{A_i} - \mathbb{1}_{C_{k,i}}\right| d\mu = \lim_{k\to\infty} \int_S \left[\mathbb{1}_{A_i} - \mathbb{1}_{C_{k,i}}\right] d\mu = \lim_{k\to\infty} [\mu(A_i) - \mu(C_{k,i})] = 0. \tag{1.4}$$

Let $k \in \mathbb{N}$. Define the function $\sigma_{k,m,i} := [1 - md(x, C_{k,i})]^+$ for any $m \in \mathbb{N}$. Clearly $\sigma_{k,m,i}$ is bounded and by Lemma 1.11 we find that $\sigma_{k,m,i}$ is Lipschitz. The lemma also tells us that for any $x \in S$ we get

$$\lim_{m\to\infty} \sigma_{k,m,i}(x) = \mathbb{1}_{C_{k,i}}(x)$$

since $C_{k,i}$ closed. Furthermore we have by definition that $\sigma_{k,m,i}(x) = 1$ for $x \in C_{k,i}$, so we get for any $m \in \mathbb{N}$ that $\mathbb{1}_{C_{k,i}} \leq \sigma_{k,m,i}$. Hence we get for any $k \in \mathbb{N}$ that

$$\lim_{m\to\infty} \int_S \left|\mathbb{1}_{C_{k,i}} - \sigma_{k,m,i}\right| d\mu = \lim_{m\to\infty} \int_S \left[\sigma_{k,m,i} - \mathbb{1}_{C_{k,i}}\right] d\mu = \int_S \mathbb{1}_{C_{k,i}} d\mu - \int_S \mathbb{1}_{C_{k,i}} d\mu = 0. \tag{1.5}$$

Let $l \in \mathbb{N}$. By (1.4) we can take $k_l \in \mathbb{N}$ such that

$$\int_S \left|\mathbb{1}_{A_i} - \mathbb{1}_{C_{k_l,i}}\right| d\mu < \frac{1}{2l}.$$

For this k_l we can by (1.5) take $m_l \in \mathbb{N}$ such that

$$\int_S \left|\mathbb{1}_{C_{k_l,i}} - \sigma_{k_l,m_l,i}\right| d\mu < \frac{1}{2l}.$$

Thus we construct a sequence $((k_l, m_l))_{l\in\mathbb{N}} \subset \mathbb{N} \times \mathbb{N}$. We find for any $l \in \mathbb{N}$ that

$$\begin{aligned}
\int_S |\mathbb{1}_{A_i} - \sigma_{k_l,m_l,i}|\, d\mu &= \int_S \left|\mathbb{1}_{A_i} - \mathbb{1}_{C_{k_l,i}} + \mathbb{1}_{C_{k_l,i}} - \sigma_{k_l,m_l,i}\right| d\mu \\
&\leq \int_S \left|\mathbb{1}_{A_i} - \mathbb{1}_{C_{k_l,i}}\right| + \int_S \left|\mathbb{1}_{C_{k_l,i}} - \sigma_{k_l,m_l,i}\right| d\mu \\
&< \frac{1}{2l} + \frac{1}{2l} = \frac{1}{l},
\end{aligned}$$

proving that

$$\lim_{l\to\infty} \int_S |\mathbb{1}_{A_i} - \sigma_{k_l,m_l,i}|\, d\mu = 0.$$

Now, for any $l, n \in \mathbb{N}$ we define, as described above,

$$\psi_{l,n} := \sum_{i=1}^{N_n} \alpha_i \sigma_{k_l,m_l,i}.$$

By Lemma 1.4 we get that $\psi_{l,n}$ is Lipschitz. Clearly $\psi_{l,n}$ is bounded and we get for any $n \in \mathbb{N}$ that

$$\begin{aligned}\lim_{l\to\infty}\int_S |\varphi_n - \psi_{l,n}|\,d\mu &= \lim_{l\to\infty}\int_S \left|\sum_{i=1}^{N_n}\alpha_i \mathbb{1}_{A_i} - \sum_{i=1}^{N_n}\alpha_i\sigma_{k_l,m_l,i}\right| d\mu \\ &\leq \lim_{l\to\infty}\int_S \sum_{i=1}^{N_n}\alpha_i\,|\mathbb{1}_{A_i} - \sigma_{k_l,m_l,i}|\,d\mu \\ &= \sum_{i=1}^{N_n}\alpha_i \lim_{l\to\infty}\int_S |\mathbb{1}_{A_i} - \sigma_{k_l,m_l,i}|\,d\mu = 0. \end{aligned} \tag{1.6}$$

Let $t \in \mathbb{N}$. Since φ_n is a nondecreasing sequence converging pointwise to g, we can, in a similar way to (1.5), prove that we may take $n_t \in \mathbb{N}$ such that

$$\int_S |g - \varphi_{n_t}|\,d\mu < \frac{1}{2t}.$$

By (1.6) we can, for this n_t, take $l_t \in \mathbb{N}$ such that

$$\int_S |\varphi_{n_t} - \psi_{l_t,n_t}|\,d\mu < \frac{1}{2t}.$$

Therefore we get for any $t \in \mathbb{N}$

$$\begin{aligned}\int_S |g - \psi_{l_t,n_t}|\,d\mu &= \int_S |g - \varphi_{n_t} + \varphi_{n_t} - \psi_{l_t,n_t}|\,d\mu \\ &\leq \int_S |g - \varphi_{n_t}|\,\mu + \int_S |\varphi_{n_t} - \psi_{l_t,n_t}|\,d\mu \\ &< \frac{1}{2t} + \frac{1}{2t} < \frac{1}{t}.\end{aligned}$$

Hence

$$\lim_{t\to\infty}\int_S |g - \psi_{l_t,n_t}|\,d\mu = 0,$$

which means that we have constructed a sequence $(\psi_{l_t,n_t})_{t\in\mathbb{N}} \subset \mathrm{BL}(S)$ such that

$$\lim_{t\to\infty}\|g - \psi_{l_t,n_t}\|_{\mathcal{L}^1} = 0.$$

For any $f \in \mathcal{L}^1(d\mu)$ we have that f^+ and f^- are nonnegative integrable functions. Hence, by what was just proven, we can construct sequences $(f_n^+)_{n\in\mathbb{N}} \subset \mathrm{BL}(S)$ and $(f_n^-)_{n\in\mathbb{N}} \subset \mathrm{BL}(S)$ such that

$$\lim_{n\to\infty}\left\|f^+ - f_n^+\right\|_{\mathcal{L}^1} = \lim_{n\to\infty}\int_S \left\|f^- - f_n^-\right\|_{\mathcal{L}^1} = 0.$$

Now taking $f_n := f_n^+ - f_n^-$ gives

$$\lim_{n\to\infty}\|f - f_n\|_{\mathcal{L}^1} \leq \lim_{n\to\infty}\int_S \left\|f^+ - f_n^+\right\|_{\mathcal{L}^1} d\mu + \lim_{n\to\infty}\int_S \left\|f^- - f_n^-\right\|_{\mathcal{L}^1} d\mu = 0$$

proving the result. □

We will end this section by giving a lemma that will be quite useful later on. Note that the following lemma does not require the set, that is the domain of the functions, to be a metric space.

Lemma 1.20
Let $X \neq \emptyset$ be any set. We find for $f, g, h\colon X \to \mathbb{R}$ that

$$\left| \inf_{x\in X}[f(x)+h(x)] - \inf_{x\in X}[g(x)+h(x)] \right| \leq \sup_{x\in X} |f(x)-g(x)|$$

if f, g and h are such that said supremum and infima are finite.

Proof. Without loss of generality we assume that $\inf_{x\in X}[f(x)+h(x)] \geq \inf_{x\in X}[g(x)+h(x)]$. Let $\epsilon > 0$ and choose $x' \in X$ such that

$$g(x') + h(x') < \inf_{x\in X}[g(x)+h(x)] + \epsilon, \text{ so } -[g(x')+h(x')-\epsilon] > -\inf_{x\in X}[g(x)+h(x)].$$

We get

$$\begin{aligned}
\left| \inf_{x\in X}[f(x)+h(x)] - \inf_{x\in X}[g(x)+h(x)] \right| &= \inf_{x\in X}[f(x)+h(x)] - \inf_{x\in X}[g(x)+h(x)] \\
&\leq f(x') + h(x') - \inf_{x\in X}[g(x)+h(x)] \\
&< f(x') + h(x') - [g(x')+h(x')-\epsilon] \\
&= f(x') - g(x') + \epsilon \\
&\leq |f(x') - g(x')| + \epsilon \\
&\leq \sup_{x\in X} |f(x)-g(x)| + \epsilon.
\end{aligned}$$

Since the last inequality holds for any $\epsilon > 0$, the result follows. □

1.2 Embedding of $\mathcal{M}(S)$ into $\mathrm{BL}(S)^*$

Each $\mu \in \mathcal{M}(S)$ defines $I_\mu \in \mathrm{BL}(S)^*$ given by

$$I_\mu(f) = \int_S f d\mu. \tag{1.7}$$

Note that $\int_S |f|\, d\mu < \infty$ holds since f is bounded and μ is finite. For any $\mu \in \mathcal{M}(S)$ we have that I_μ is a continuous functional. Namely for linear maps between normed linear spaces it is enough to find a $k \in \mathbb{R}$ such that $|I_\mu(f)| \leq k$ for all $f \in \mathrm{BL}(S)$ with $\|f\|_{\mathrm{BL}} \leq 1$. We find such a k in the following lemma.

Lemma 1.21
For $\mu \in \mathcal{M}(S)$ we have $\|I_\mu\|^*_{\mathrm{BL}} \leq \|\mu\|_{\mathrm{TV}}$.

Proof. We find for $\mu \in \mathcal{M}(S)$ that

$$\begin{aligned} \|I_\mu\|_{\mathrm{BL}}^* &= \sup\left\{\left|\int_S f d\mu\right| : \|f\|_{\mathrm{BL}} \leq 1\right\} \\ &\leq \sup\left\{\int_S |f|\, d|\mu| : \|f\|_{\mathrm{BL}} \leq 1\right\} \\ &\leq \sup\left\{\int_S \|f\|_\infty\, d|\mu| : \|f\|_{\mathrm{BL}} \leq 1\right\} \\ &\leq \int_S 1 d|\mu| = |\mu|\,(S) = \|\mu\|_{\mathrm{TV}}\,. \end{aligned}$$

□

Lemma 1.22
For $\mu \in \mathcal{M}^+(S)$ we have $\|I_\mu\|_{\mathrm{BL}}^* = \|\mu\|_{\mathrm{TV}}$.

Proof. From Lemma 1.21 we obtain $\|I_\mu\|_{\mathrm{BL}}^* \leq \|\mu\|_{\mathrm{TV}}$. Since the constant function 1 is in $\mathrm{BL}(S)$ with $\|1\|_{\mathrm{BL}}^* = 1$, we find

$$\|\mu\|_{\mathrm{TV}} = |\mu|(S) = \mu(S) = \int_S 1 d\mu \leq \|I_\mu\|_{\mathrm{BL}}^*$$

because μ is a positive measure. This proves the lemma. □

We define the map $\chi\colon \mathcal{M}(S) \to \mathrm{BL}(S)^*$ by $\mu \mapsto I_\mu$, where $I_\mu \in \mathrm{BL}(S)^*$ is defined by (1.7).

Lemma 1.23
The map χ is injective.

Proof. Let $\mu, \nu \in \mathcal{M}(S)$ such that $\mu \neq \nu$. Take $A \subset S$ a Borel set such that $\mu(A) \neq \nu(A)$, then $A \neq \emptyset$. We assume that $I_\mu = I_\nu$ and work towards a contradiction.
Let $B \subset S$ be closed. We define $f_n(x) := [1 - nd(x,B)]^+$ for all $n \in \mathbb{N}$. By Lemma 1.11 we find that f_n is Lipschitz continuous with $|f_n|_L \leq n$ and that the pointwise limit of f_n is $\mathbb{1}_B$. Now using the assumption that $I_\mu = I_\nu$ we find that

$$\begin{aligned} \mu(B) &= \int_S \mathbb{1}_B d\mu = \int_S \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_S f_n d\mu \\ &= \lim_{n\to\infty} I_\mu(f_n) = \lim_{n\to\infty} I_\nu(f_n) = \lim_{n\to\infty} \int_S f_n d\nu \\ &= \int_S \lim_{n\to\infty} f_n d\nu = \int_S \mathbb{1}_B d\nu = \nu(B), \end{aligned}$$

by Lebesgue's Dominated Convergence Theorem, using that $|f_n| \leq 1$ for all $n \in \mathbb{N}$ and that the constant function 1 is μ and ν-integrable, since μ and ν are finite. By Remark 1.18 we know μ and ν to be inner regular, so we find that

$$\mu(A) = \sup\{\mu(B) : B \subset A, B \text{ closed}\} = \sup\{\nu(B) : B \subset A, B \text{ closed}\} = \nu(A),$$

giving a contradiction. We conclude that $I_\mu \neq I_\nu$, hence χ is injective. □

Similar to χ we define $\xi\colon \mathcal{M}_1(S) \to \mathrm{Lip}_e(S)^*$ given by $\mu \to I_\mu$, where $I_\mu \in \mathrm{Lip}_e(S)^*$ is given by 1.7. In the next lemma we prove that ξ is an embedding of $\mathcal{M}_1(S)$ into $\mathrm{Lip}_e(S)^*$. Note that we can only find such an embedding for $\mathcal{M}_1(S)$, not $\mathcal{M}(S)$, since when f is Lipschitz we want either f to be bounded (like in the $\mathcal{M}(S) \subset \mathrm{BL}(S)^*$ case) or μ to have finite first moment to ensure that $I_\mu(f) = \int_S f d\mu$ is finite.

Lemma 1.24
The map ξ is injective and for $\mu \in \mathcal{M}_1(S)$ we find that

$$\|I_\mu\|_e^* \leq \|\mu\|_1 .$$

Proof. If for $\mu \in \mathcal{M}_1(S)$ we have $I_\mu(f) = 0$ for all $f \in \mathrm{Lip}_e(S)$ then we also have $I_\mu(f) = 0$ for all $f \in \mathrm{BL}(S)$, since $\mathrm{BL}(S) \subset \mathrm{Lip}_e(S)$, hence $\mu = 0$ by Lemma 1.23, proving injectivity for ξ. The second statement follows since for any $f \in \mathrm{Lip}_e(S)$ we find for all $x \in S$ that

$$\begin{aligned} |f(x)| &\leq |f(x) - f(e)| + |f(e)| \leq |f|_L \, d(x,e) + |f(e)| \\ &\leq (|f|_L + |f(e)|)(\max(1, d(x,e)) \leq \|f\|_e \max(1, d(x,e)) \end{aligned}$$

hence we get

$$\left| \int_S f d\mu \right| \leq \int_S |f| \, d|\mu| \leq \|f\|_e \int_S \max(1, d(x,e)) d|\mu| = \|f\|_e \, \|\mu\|_1$$

proving the inequality. □

Remark 1.25: In this thesis we are mostly interested in norms and metrics on measures. The embedding ξ of $\mathcal{M}_1(S)$ into $\mathrm{Lip}_e^*(S)$ is a valuable result in this respect. Namely, consider the norm $\|\cdot\|_e^*$ we defined on $\mathrm{Lip}_e^*(S)$. This now gives a norm on $\mathcal{M}_1(S)$. For $\mu \in \mathcal{M}_1(S)$ take

$$\|\mu\|_e^* := \|I_\mu\|_e^* .$$

For this to define a norm on $\mathcal{M}_1(S)$ we need ξ to be an embedding since then $\|\mu\|_e^* = 0$ if and only if $\mu = 0$. The other properties for norms follow easily.

1.3 An extension of the Kantorovich Norm

We gave the definition of the $\|\cdot\|_{KR}$ norm on $\mathcal{M}_1^0(S)$ in Definition 1.12. This norm can be extended to the space $\mathcal{M}_1(S)$. In the following theorem we prove that the norm $\|\cdot\|_e^*$ on $\mathcal{M}_1(S)$, from Remark 1.25, gives us such an extension.

Theorem 1.26
Let $\mu \in \mathcal{M}_1^0(S)$. Then we find that

$$\|\mu\|_{KR} = \|\mu\|_e^* .$$

Proof. Recall that

$$\|\mu\|_{KR} = \sup \left\{ \int_S f d\mu : f \in \mathrm{Lip}(S), |f|_L \leq 1 \right\}$$

and

$$\|\mu\|_e^* = \sup \left\{ \int_S f d\mu : f \in \mathrm{Lip}_e(S), \|f\|_e \leq 1 \right\}$$

from which follows directly that $\|\mu\|_{KR} \geq \|\mu\|_e$ since $\|f\|_e = |f(e)| + |f|_L \geq |f|_L$.
To prove the other inequality we consider a sequence $(f_n)_{n \in \mathbb{N}} \subset \mathrm{Lip}(S)$ such that $|f_n|_L \leq 1$ and

$$\lim_{n \to \infty} \left[\int_S f_n d\mu \right] = \|\mu\|_{KR} .$$

We define for all $n \in \mathbb{N}$ functions $g_n \in \text{Lip}_e(S)$ given by $g_n(x) := f_n(x) - f_n(e)$ for all $x \in S$. Note that both $|f_n|_L = |g_n|_L$ and $|g_n(e)| = |f_n(e) - f_n(e)| = 0$ hold, hence $\|g_n\|_e \leq 1$.
Since $\mu \in \mathcal{M}_1^0(S)$, we have $\mu(S) = 0$, hence the following holds

$$\begin{aligned} \int_S g_n d\mu &= \int_S (f_n - f(e))\, d\mu \\ &= \int_S f_n d\mu - \int_S f(e) d\mu \\ &= \int_S f_n d\mu - f(e)\mu(S) \\ &= \int_S f_n d\mu. \end{aligned}$$

This implies that

$$\|\mu\|_e^* \geq \lim_{n\to\infty} \left[\int_S g_n d\mu\right] = \lim_{n\to\infty} \left[\int_S f_n d\mu\right] = \|\mu\|_{KR}.$$

We conclude that $\|\mu\|_{KR} = \|\mu\|_e^*$. □

Remark 1.27: Theorem 1.26 not only gives an extension of the norm $\|\cdot\|_{KR}$ to the space $\mathcal{M}_1(S)$, we also get an extension of d_{KR}, a metric on $\mathcal{P}_1(S)$, to the space $\mathcal{M}_1(S)$. Namely, for $\mu, \nu \in \mathcal{M}_1(S)$ we have that $\|\mu - \nu\|_e^*$ defines a metric. So we get for $\mu, \nu \in \mathcal{P}_1(S)$ by Theorem 1.26 that

$$d_{KR}(\mu, \nu) = \|\mu - \nu\|_{KR} = \|\mu - \nu\|_e^*$$

holds, giving an extension of d_{KR} to $\mathcal{M}_1(S)$.

Chapter 2

A rigorous proof of the Kantorovich-Rubinstein Theorem

In this chapter we will introduce the Kantorovich distance. This is a metric on probability measures, based on the Monge-Kantorovich mass transportation problem. Thereafter we will carefully study the proof of the Kantorovich-Rubinstein Theorem, which proves that this metric is equal to the d_{KR} metric defined earlier. We will first have a look at the transportation problem.

2.1 The Monge-Kantorovich mass transportation problem

A Polish space is a topological space that is metrizable such that it becomes a complete, separable metric space. Any metric that metrizes the space in this way is called admissible.
Let X and Y be Polish spaces. Let $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$. Let d_X and d_Y be admissible metrics on X and Y respectively. Furthermore, $X \times Y$ is equipped with the Borel σ-algebra corresponding to the product (metric) topology.

Definition 2.1
A cost function is a nonnegative, measurable function $c\colon X \times Y \to \mathbb{R}^+ \cup \{\infty\}$.

The Monge-Kantorovich problem can be described as follows. We view μ as representing a distribution of sand on space X where $\mu(A)$ denotes the amount of sand that is on the subset $A \subset X$. Similarly ν represents a hole on Y where sand can be placed, there is room for $\nu(B)$ sand on the subset $B \subset Y$. Transporting sand from $x \in X$ to $y \in Y$ costs $c(x, y)$. Minimizing the cost for transporting all sand from X to Y is known as the Monge-Kantorovich mass transportation problem, see Figure 2.1. To define the minimal cost of transportation we first need a transference plan. This is a measure π on $X \times Y$. Now $\pi(A \times B)$ tells us how much sand will be moved from $A \subset X$ to $B \subset Y$. We consider an example (Figure 2.2) where we have the sets $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2, y_3\}$. We denote the amount of sand moved from x_i to y_j by a_{ij}. We want all sand from x_i to be moved to somewhere on Y. We express this by $\mu(x_i) = a_{i1} + a_{i2} + a_{i3}$. When we translate this into the language of transference plans we write $\mu(x_i) = \pi(\{x_i\} \times Y)$. Likewise we want the hole at y_j to be completely filled up, for which we write $\nu(y_j) = \pi(X \times \{y_j\})$.

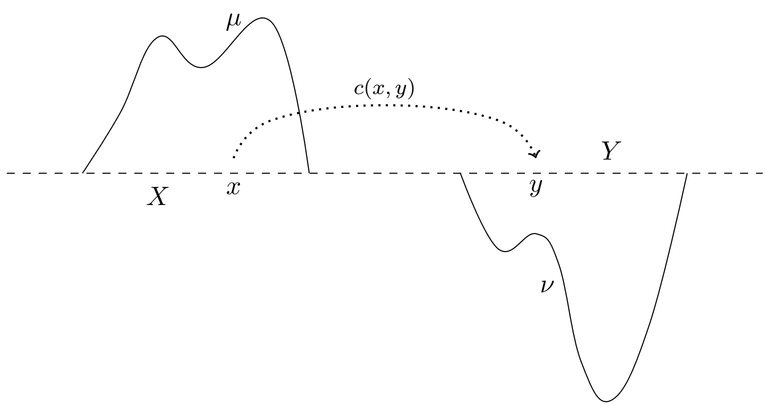

Figure 2.1: Monge-Kantorovich's mass transportation problem.

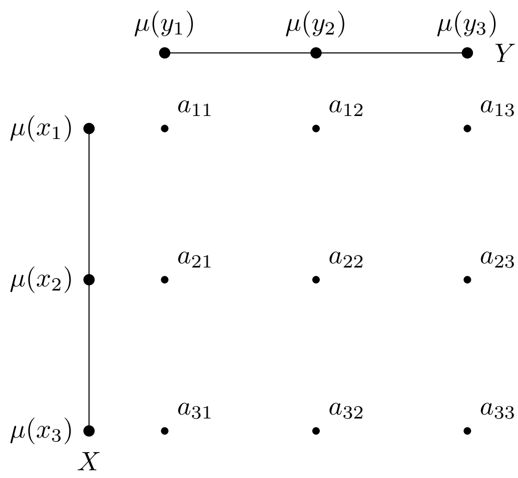

Figure 2.2: A simplified transportation plan.

The continuous case, for arbitrary Polish spaces X and Y, is quite similar. The requirement that the whole pile of sand is to be emptied and the complete hole filled up, comes down to π having to satisfy

$$\pi(A \times Y) = \mu(A) \quad \text{and} \quad \pi(X \times B) = \nu(B) \tag{2.1}$$

for all measurable subsets $A \subset X$ and $B \subset Y$.
In our example the cost of transportation would be given by

$$\sum_{i=1}^{3}\sum_{j=1}^{3} a_{ij} \cdot c(x_i, y_i), \text{ which we write as } \sum_{i=1}^{3}\sum_{j=1}^{3} \pi(\{x_i\}, \{y_j\}) \cdot c(x_i, y_i).$$

In the continuous case this cost of transportation by transportation plan π is $I(\pi)$, given by

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

The Monge-Kantorovich mass transportation problem concerns finding the minimal cost for transporting this mass. This optimal transportation cost $\mathcal{T}_c(\mu, \nu)$ is thus given by

$$\mathcal{T}_c(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} I(\pi),$$

where $\Pi(\mu, \nu)$ is the set of all admissible transference plans i.e. the set of all measures π on $X \times Y$ satisfying (2.1). Note that every transportation plan is also a probability measure since (2.1) implies $\pi(X \times Y) = \mu(X) = 1$.

2.2 Kantorovich Duality Theorem

This minimization problem allows a so called dual representation. The theorem that describes this is called the Kantorovich Duality Theorem. As a preparation for the Kantorovich Duality Theorem we define

$$J(f, g) = \int_X f d\mu + \int_Y g d\nu, \text{ for } (f, g) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$$

and the set

$$\Phi_c := \left\{(f, g) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu) : f(x) + g(y) \leq c(x, y) \text{ for } \mu\text{-a.e. } x \in X \text{ and } \nu\text{-a.e. } y \in Y\right\}.$$

Since $\mathcal{L}^1(d\mu)$ denotes the space of all measurable functions $f\colon X \to R$ that are μ-integrable, $f \in \mathcal{L}^1(d\mu)$ is every where defined, hence $f(x)$ has a meaning, as we do not consider equivalence classes of functions equal μ almost everywhere.

Theorem 2.2 (Kantorovich Duality Theorem)
Let X and Y be Polish spaces, let $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ and let $c\colon X \times Y \to \mathbb{R}^+ \cup \{\infty\}$ be a lower semi-continuous cost function. Then

$$\inf_{\pi \in \Pi(\mu,\nu)} I(\pi) = \sup_{(f,g) \in \Phi_c} J(f, g).$$

A proof of this theorem can be found in [9, p.25]. For the concept of lower semi-continuity we refer to Appendix A.

Villani presents a nice explanation of this Theorem, originally from Carafelli, in [9, p.20] that we closely reproduce here. You want to minimize the costs transporting a pile of sand μ on X to a hole ν on Y. You are using trucks to transport the sand. You do this by considering all transference plans, π, and find the minimal cost for transportation, $\inf_{\Pi(\mu,\nu)} I(\pi)$. Someone else comes along and says, you know what, you do not have to worry about how the sand gets from X to Y, I will take care of that. All I do is set a price for loading sand onto a truck at point $x \in X$, namely $f(x)$, and a price for unloading at $y \in Y$, namely $g(y)$. It will always be in your financial interest to let me take care of the transportation because $f(x)+g(y) \leq c(x,y)$! In order to achieve this I will even compensate for loading or unloading in certain places, by setting negative prices. Having set these prices determines the price for transporting even if we do not know what happens in between. The cost for loading will be $\int_X f d\mu$ and for unloading will be $\int_Y f d\nu$, making the total cost of transportation $J(f,g)$. We will always have $J(f,g) \leq I(\pi)$ by construction. You will of course accept the deal, and what the theorem tells us is that if the other is smart enough and he sets the prices in a clever enough way, then the cost will be (almost) as much as you were ready to spent on the other method anyway.
So for us the benefit of this theorem is that we do not have to care about the infimum over the set of transference plans, instead we have a supremum over the set Φ_c, which, as we will see in the next section, is nicely manageable.

2.3 The f^c and f^{cc} function

In this section we will give and prove lemmas that will be used in the proof of the Kantorovich-Rubinstein Theorem. Let us first define the functions f^c and f^{cc}, they play an important part in all this. In [9] these functions were simply introduced in the proof of the Kantorovich Duality Theorem, without detailed proof of their properties.

Definition 2.3
For c a cost function we define for any bounded $f \in \mathcal{L}^1(d\mu)$ the functions $f^c\colon Y \to \mathbb{R}$ and $f^{cc}\colon X \to \mathbb{R} \cup \{-\infty\}$ by

$$f^c(y) := \inf_{x\in X}[c(x,y) - f(x)] \quad \text{and} \quad f^{cc}(x) := \inf_{y\in Y}[c(x,y) - f^c(y)].$$

Remark 2.4: We assume for this definition that for each $x \in X$ there exists a $y \in Y$ such that $c(x,y) < \infty$. Otherwise take $A \subset X$ the set of $x \in X$ such that $c(x,y) = \infty$ for all $y \in Y$. We would have that $I(\pi) = \infty$ if $\mu(A) \neq 0$. In that case it is not interesting to consider this cost function and the results in this chapter.
If $\mu(A) = 0$ then $I(\pi)$ would not change if we took $c(x,y) = 0$ for all $x \in A$ and $y \in Y$.
That $I(\pi) = \infty$ if $\mu(A) \neq 0$ follows from the fact that $\pi(A \times Y) \neq 0$ for all $\pi \in \Pi(\mu,\nu)$ whilst $c(x,y) = \infty$ for all $x \in A$ and $y \in Y$, or in words, we have to move sand from A to somewhere on Y but it will always cost us infinitely much to do so. Another reason for making this assumption is that otherwise we would have $f^{cc}(x) = \infty$ for any $x \in A$.
Of course we also assume for all $y \in Y$ that there exists an $x \in X$ such that $c(x,y) < \infty$. Otherwise we would again have $I(\pi) = \infty$ and there would exist some $y \in Y$ such that $f^c(y) = \infty$.

Remark 2.5: The infimum in the definition of f^c always exists since c is nonnegative and f is bounded. Note that some authors define f^c and f^{cc} for functions $f \in \mathcal{L}^1(d\mu)$ that are not necessarily bounded. Then the range of f^c would be $\mathbb{R} \cup \{-\infty\}$. The range of f^{cc} is $\mathbb{R} \cup \{-\infty\}$ in any case since f^c does not have to be bounded from above, even if f is bounded. If the cost function c is bounded then both f^c and f^{cc} are bounded for bounded $f \in \mathcal{L}^1 d(\mu)$. See the discussion below in Remark 2.6.

We may consider Definition 2.3 in the following way. For a pair $(f,g) \in \Phi_c$ we know that $f(x) + g(y) \leq c(x,y)$ holds for μ-a.e $x \in X$ and ν-a.e. $y \in Y$. For proving the Kantorovich-Rubinstein Theorem we will use the Kantorovich Duality Theorem 2.18, which gives us the following expression

$$\sup_{(f,g)\in\Phi_c} J(f,g).$$

To make this expression easier to handle we will use f^c and f^{cc}. Namely f^c will act as a replacement for g as a function that still satisfies $f(x) + f^c(y) \leq c(x,y)$ and gives a possibly higher value $J(f,f^c) \geq J(f,g)$. Actually by definition it has, for each $y \in Y$, the largest value that a function satisfying $f(x) + f^c(y) \leq c(x,y)$ (for μ-a.e. $x \in X$) can have. Therefore it maximizes the value of $J(f,f^c)$. Then we do similar thing with f^{cc} replacing f. Note that the functions f^c and f^{cc} erase the need for a pair $(f,g) \in \Phi_c$. We can now take the supremum over bounded $f \in \mathcal{L}^1(d\mu)$ where each f generates a pair $(f^{cc}, f^c) \in \Phi_c$. That this does not change the value of the supremum will be proven in Lemma 2.14. One might then wonder if taking f^{ccc} would further increase the value J. It does not, since $(f^{cc})^c = f^c$ which we prove in Lemma 2.11. Before we may do any of this we will have to proof that f^c and f^{cc} are actually measurable and integrable functions which we do in Corollary 2.10 and Lemma 2.12 respectively.

Remark 2.6 (c_X and c_Y)**:** Some of the upcoming lemmas require f^c and f^{cc} to take only values in $\mathbb{R}$ or we might even want them to be bounded. Taking the cost function to be bounded would ensure that both f^c and f^{cc} are bounded as well, provided that f is bounded. We can be more general though. Let us define $c_X : X \to \mathbb{R}^+$

$$x \mapsto \inf_{y\in Y} c(x,y)$$

and $c_Y : Y \to \mathbb{R}^+$ by

$$y \mapsto \inf_{x\in X} c(x,y).$$

We know that since c is nonnegative and since f is bounded that f^c is bounded from below, hence takes only values in $\mathbb{R}$. For bounded f we find that f^c is bounded if and only if c_Y is bounded, because when c_Y is bounded then f^c is bounded from above. With f^c being bounded we prevent f^{cc} from ever taking value $-\infty$. If we also have that c_X is bounded we even get that f^{cc} is bounded. We prove this in Lemma 2.12.
To clarify, see Table 2.1 for an overview. Note that these results hold when f is bounded.

	any cost function	c_X bounded	c_Y bounded	c_X and c_Y bounded	c bounded
f^c	values in $\mathbb{R}$	values in $\mathbb{R}$	bounded	bounded	bounded
f^{cc}	values in $\mathbb{R} \cup \{-\infty\}$	values in $\mathbb{R} \cup \{-\infty\}$	values in $\mathbb{R}$	bounded	bounded

Table 2.1: The consequences of c_X and/or c_Y being bounded.

Clearly, if c is bounded, both c_X and c_Y are bounded. So requiring that only c_X and/or c_Y are bounded is more general than taking c bounded. What we gain from defining c_X and c_Y becomes clear in Example 2.7.
Note that whenever we only require c_Y to be bounded we might as well take only c_X to be bounded. In that case we would have to consider g^c and g^{cc} for bounded $g \in \mathcal{L}^1(d\nu)$, instead of f^c and f^{cc} for bounded $f \in \mathcal{L}^1(d\mu)$, in order to get similar results.

Example 2.7
Let $X = Y$ and c a cost function such that $c(x,x) = 0$ for all $x \in X$. Then we have

$0 \leq \inf_{y \in Y} c(x, y) \leq c(x, x) = 0$ *for all* $x \in X$ *and we have* $0 \leq \inf_{x \in X} c(x, y) \leq c(y, y) = 0$ *for all* $y \in X$ *. That means that* $c_X = c_Y = 0$ *holds, hence* c_X *and* c_Y *are bounded. So* c_X *and* c_Y *are bounded if we take* $c = d$ *to be a (not necessarily bounded) metric on* X*.*

We know that for bounded $f \in \mathcal{L}^1(d\mu)$, when the cost function c is bounded, both f^c and f^{cc} are bounded as well. There are other properties that the f^c and f^{cc} functions inherit from the cost function. We will see two such examples in Lemma 2.8 and Lemma 2.9.

Lemma 2.8
Let c *be a Lipschitz continuous cost function, such that* c_Y *is bounded. Then for any bounded* $f \in \mathcal{L}^1(d\mu)$ *we find that* f^c *and* f^{cc} *are Lipschitz continuous with* $|f^c|_L \leq |c|_L$ *and* $|f^{cc}|_L \leq |c|_L$*.*

Proof. We need c_Y to be bounded so that f^c and f^{cc} take only values in $\mathbb{R}$. Let $y, y' \in Y$. Then we obtain from Lemma 1.20 that

$$\begin{aligned} |f^c(y) - f^c(y')| &= \left| \inf_{x \in X} [c(x, y) - f(x)] - \inf_{x \in X} [c(x, y') - f(x)] \right| \\ &\leq \sup_{x \in X} |c(x, y) - c(x, y')| \\ &\leq |c|_L \, d_{X \times Y}((x, y), (x, y')) \\ &= |c|_L \, (d_X(x, x) + d_Y(y, y')) \\ &= |c|_L \, d_Y(y, y'). \end{aligned}$$

This proves the statement for f^c. In a similar way we can prove the result for f^{cc}. □

Lemma 2.9
Let c *be a lower semi-continuous cost function. Then for any bounded* $f \in \mathcal{L}^1(d\mu)$ *we find that* f^c *and* f^{cc} *are lower semi-continuous.*

Proof. Let $(x_0, y_0) \in X \times Y$. Let $\epsilon > 0$. Since c is lower semi-continuous we can take $\delta > 0$ such that

$$c(x', y') > c(x_0, y_0) - \frac{\epsilon}{2}$$

for all $(x', y') \in X \times Y$ that satisfy $d_{X \times Y}((x_0, y_0), (x', y')) = d_X(x_0, x') + d_Y(y_0, y') < \delta$.
We see that

$$c(x', y') > c(x_0, y_0) - \frac{\epsilon}{2}, \text{ gives us that } c(x_0, y_0) - c(x', y') < \frac{\epsilon}{2}.$$

Take $y \in Y$ such that $d_Y(y_0, y) < \delta$. We need to show that $f^c(y) > f^c(y_0) - \epsilon$ holds to prove lower semi-continuity for f^c.
We take $\hat{x} \in X$ such that $f^c(y) > c(\hat{x}, y) - f(\hat{x}) - \frac{\epsilon}{2}$. We can do this since $f^c(y) = \inf_{x \in X} [c(x, y) - f(x)]$. Note that we also have $f^c(y_0) \leq c(\hat{x}, y_0) - f(\hat{x})$. Now we find

$$\begin{aligned} f^c(y_0) - f^c(y) &\leq c(\hat{x}, y_0) - f(\hat{x}) - f^c(y) \\ &< c(\hat{x}, y_0) - f(\hat{x}) - [c(\hat{x}, y) - f(\hat{x}) - \epsilon/2] \\ &= c(\hat{x}, y_0) - c(\hat{x}, y) + \frac{\epsilon}{2} \\ &\leq \epsilon \end{aligned}$$

since $d_{X \times Y}((\hat{x}, y_0), (\hat{x}, y)) = d_X(\hat{x}, \hat{x}) + d_Y(y_0, y) < \delta$. What we have is

$$f^c(y_0) - f^c(y) < \epsilon, \text{ which implies } f^c(y) > f^c(y_0) - \epsilon,$$

proving lower semi-continuity for f^c. Similar reasoning applies to f^{cc}. □

Corollary 2.10
Let c be a lower semi-continuous cost function on $X \times Y$. Then for any bounded $f \in \mathcal{L}^1(d\mu)$ the functions f^c and f^{cc} are Borel measurable.

Proof. By definition, for a lower semi-continuous function $f\colon X \to \overline{R}$ we have for any $r \in \mathbb{R}$ that $f^{-1}((r,\infty])$ is open in X. This actually corresponds to one among various equivalent definitions of Borel measurability as proven in [4, Theorem 9.2]. Hence by Lemma 2.9 both f^c and f^{cc} are measurable. $\square$

Lemma 2.11
Let c be a lower semi-continuous cost function on $X \times Y$ such that c_Y is bounded. Then for bounded $f \in \mathcal{L}^1(d\mu)$ we have $f^{ccc} = f^c$.

Proof. By definition we have $f^{cc}(x) + f^c(y) \leq c(x,y)$ for all $(x,y) \in X \times Y$ so we find $f^c \leq f^{ccc}$ for the function f^{ccc} defined for every $y \in Y$ by

$$f^{ccc}(y) := \inf_{x \in X}[c(x,y) - f^{cc}(x)].$$

For any $y \in Y$ we find that

$$\begin{aligned}
f^{ccc}(y) &\overset{\text{def}}{=} \inf_{x\in X}\big[c(x,y) - f^{cc}(x)\big] \\
&\overset{\text{def}}{=} \inf_{x\in X}\Big[c(x,y) - \inf_{y'\in Y}\big[c(x,y') - f^c(y')\big]\Big] \\
&= \inf_{x\in X}\Big[c(x,y) + \sup_{y'\in Y}\big[-c(x,y') + f^c(y')\big]\Big] \\
&\overset{\text{def}}{=} \inf_{x\in X}\Big[c(x,y) + \sup_{y'\in Y}\Big[-c(x,y') + \inf_{x'\in X}\big[c(x',y') - f(x')\big]\Big]\Big] \\
&\leq \inf_{x\in X}\Big[c(x,y) + \sup_{y'\in Y}\Big[-c(x,y') + \big[c(x,y') - f(x)\big]\Big]\Big] \\
&= \inf_{x\in X}\Big[c(x,y) + \sup_{y'\in Y}\big[-f(x)\big]\Big] \\
&= \inf_{x\in X}[c(x,y) - f(x)] = f^c(y)
\end{aligned}$$

hence we also find $f^c \geq f^{ccc}$. $\square$

Lemma 2.12
Let c be a lower semi-continuous cost function on $X \times Y$ such that both c_X and c_Y are bounded. If $f \in \mathcal{L}^1(d\mu)$ is bounded then f^c and f^{cc} are integrable for any finite signed measure on Y respectively X.

Proof. From Corollay 2.10 we obtain that f^c and f^{cc} are measurable.
Let $M, m \in \mathbb{R}$ such that $c_Y(y) = \inf_{x\in X} c(x,y) \leq M$ for all $y \in Y$ and $|f(x)| \leq m$ for all $x \in X$. We will show that f^c and f^{cc} are bounded. For any $y \in Y$ we have that

$$f^c(y) = \inf_{x\in X}[c(x,y) - f(x)] \leq \inf_{x\in X}[c(x,y) + m] = \inf_{x\in X}[c(x,y)] + m \leq M + m$$

and we also find that

$$f^c(y) = \inf_{x\in X}[c(x,y) - f(x)] \geq \inf_{x\in X}[c(x,y) - m] \geq -m$$

since c is nonnegative. Since f^c is bounded we can now prove in a similar way that f^{cc} is bounded. Hence f^c and f^{cc} are integrable for finite signed measures on Y and X respectively. $\square$

Lemma 2.13
Let c be a cost function. Then

$$\sup_{(f,g)\in\Phi_c} J(f,g) = \sup_{\substack{(f,g)\in\Phi_c,\\ f,g \text{ bounded}}} J(f,g).$$

Proof. Let $(f,g)\in\Phi_c$ and take for all $n\in\mathbb{N}$ the functions f_n and g_n where

$$f_n := \min(n, \max(f,-n)) \quad \text{and} \quad g_n := \min(n, \max(g,-n)).$$

Let $n\in\mathbb{N}$. Clearly since $(f,g)\in\mathcal{L}^1(d\mu)\times\mathcal{L}^1(d\nu)$ we have that $(f_n,g_n)\in\mathcal{L}^1(d\mu)\times\mathcal{L}^1(d\nu)$. Now let $(x,y)\in X\times Y$ such that $f(x)+g(y)\le c(x,y)$.

(1) If $f(x)\ge -n$ and $g(y)\ge -n$ then $f_n(x)+g_n(y)\le f(x)+g(y)\le c(x,y)$.

(2) If $f(x)\le 0$ and $g(y)\le 0$ then $f_n(x)+g_n(y)\le 0\le c(x,y)$.

(3) If $f(x)\ge 0$ and $g(y)\le -n$ then $f_n(x)+g_n(y)=f_n(x)-n\le n-n=0\le c(x,y)$.

For any case that is not covered by (1) or (2) we have that either $f(x)<-n$ or $g(y)<-n$ (else it is covered by (1)) and either $f(x)>0$ or $g(y)>0$ (else it is covered by (2)). From this it follows that in a similar way to how we proof (3) we get all cases. Hence we find that $(f_n,g_n)\in\Phi_c$. Clearly both f_n and g_n are bounded and we have that $\lim_{n\to\infty} J(f_n,g_n)=J(f,g)$. □

Lemma 2.14
Let c be a lower semi-continuous cost function on $X\times Y$ such that c_X and c_Y are bounded. Then

$$\sup_{(f,g)\in\Phi_c} J(f,g) = \sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(f^{cc},f^c).$$

Proof. By Lemma 2.13 we find that

$$\sup_{(f,g)\in\Phi_c} J(f,g) = \sup_{\substack{(f,g)\in\Phi_c,\\ f,g \text{ bounded}}} J(f,g).$$

Let $(f,g)\in\Phi_c$ such that f and g are bounded. Since f is bounded we find by Lemma 2.12 that f^c and f^{cc} are integrable. By definition $(f,g)\in\Phi_c$ gives for μ-a.e. $x\in X$ and ν-a.e. $y\in Y$ that

$$f(x)+g(y)\le c(x,y), \quad \text{thus} \quad g(y)\le c(x,y)-f(x).$$

Now, by the definition $f^c(y)=\inf_{x\in X}[c(x,y)-f(x)]$, we see that f^c is the largest function satisfying $f^c(y)\le c(x,y)-f(x)$ for μ-a.e. $x\in X$ and ν-a.e. $y\in Y$. This gives that $g\le f^c$ ν-a.e. so we have $\int_Y g d\nu \le \int_Y f^c d\nu$ which implies $J(f,g)\le J(f,f^c)$. In a similar way we get $f\le f^{cc}$ μ-a.e. giving that $J(f,g)\le J(f,f^c)\le J(f^{cc},f^c)$. Hence

$$\sup_{(f,g)\in\Phi_c} J(f,g) = \sup_{\substack{(f,g)\in\Phi_c,\\ f,g \text{ bounded}}} J(f,g) \le \sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(f^{cc},f^c).$$

For any bounded $f\in\mathcal{L}^1(d\mu)$ we have, by definition, that $f^{cc}(x)+f^c(y)\le c(x,y)$ holds for all $(x,y)\in X\times Y$. Since f^c and f^{cc} are integrable we find $(f^{cc},f^c)\in\Phi_c$. This gives

$$\sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(f^{cc},f^c) \le \sup_{(f,g)\in\Phi_c} J(f,g).$$

Hence we get

$$\sup_{(f,g)\in\Phi_c} J(f,g) = \sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(f^{cc},f^c).$$ □

2.4 The Kantorovich distance

To derive a metric on probability measures from the transportation problem we will choose $X = Y$ and for the cost function a metric $c = d$ on X. Note that d need not be the admissible metric d_X on X defining the topology. We require that the metric d is a lower semi-continuous function on $X \times X$, because then d is measurable, as we have seen in the proof of Corollary 2.10. When we write $\mathcal{P}_1(X)$ this refers to the set of probability measures that have finite first moment with respect to d.

Definition 2.15 (Kantorovich distance)
Let $X = Y$ be a Polish space, $\mu, \nu \in \mathcal{P}_1(X)$ and let d be a metric on X. Let $\mathcal{T}_d$ be the cost of optimal transportation for the cost $d(x, y)$, so

$$\mathcal{T}_d(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d(x, y)\, d\pi(x, y).$$

This metric is known as the Kantorovich distance. It is a slightly more general definition than that of the Wasserstein metric of order 1, where one takes d to be equal to d_X, the metric that defines the topology of X. A metric is always continuous with respect to the topology derived from it, therefore the requirement of lower semi-continuity, i.e. measurability, is in that case automatically fulfilled. In Chapter 3 we will discuss Wasserstein metrics of order p where $p \geq 1$.

Now that we have chosen the cost function to be a metric we find some very nice properties. For example the c_X and c_Y functions from Remark 2.6 are now bounded (actually constant zero), so previous results from this chapter hold even if d is not bounded. Moreover we get some nice properties for the f^d and f^{dd} function.

Lemma 2.16
Let $X = Y$ be a Polish space and let d be a metric on X. For bounded $f \in \mathcal{L}^1(d\mu)$ we find that f^d and f^{dd} are 1-Lipschitz functions with respect to d, i.e. $f^d, f^{dd} \in \mathrm{Lip}(X)$ with $\left|f^d\right|_L \leq 1$ and $\left|f^{dd}\right|_L \leq 1$.

Proof. For $x, x' \in X$ we have by Lemma 1.20

$$\begin{aligned} \left|f^d(x) - f^d(x')\right| &= \left|\inf_{y \in X}[d(x, y) - f(y)] - \inf_{y' \in X}[d(x', y') - f(y')]\right| \\ &\leq \sup_{y \in X} |d(x, y) - d(x', y)| \\ &\leq d(x, x'). \end{aligned}$$

Hence we find that f^d is 1-Lipschitz. A similar result can be proven for f^{dd}. □

Lemma 2.17
Let $X = Y$ be a Polish space and let d be a metric on X. For bounded $f \in \mathcal{L}^1(d\mu)$ ones has

$$f^{dd} = -f^d.$$

Proof. By Lemma 2.16 we have that f^d and f^{dd} are 1-Lipschitz. This gives for all $x, y \in X$ that

$$-f^d(x) + f^d(y) \leq \left|f^d(x) - f^d(y)\right| \leq d(x, y).$$

Therefore we have for any $x \in X$ that

$$-f^d(x) \leq \inf_{y \in X}[d(x, y) - f^d(y)] = f^{dd}(x).$$

Furthermore we have

$$f^{dd}(x) = \inf_{y \in X}[d(x,y) - f^d(y)] \leq d(x,x) - f^d(x) = -f^d(x).$$

This gives us equality, hence for all $x \in X$ we get

$$-f^d(x) = f^{dd}(x).$$

□

2.5 The Kantorovich-Rubinstein Theorem

Most of the work for the proof of the Kantorovich-Rubinstein Theorem has already been done, the proof will mainly consist of connecting previous results. In the Kantorovich-Rubinstein Theorem we will write $\mathrm{Lip}(X,d) \cap \mathcal{L}^1$ for the set of functions $f\colon X \to \mathbb{R}$ that are Lipschitz continuous with respect to the metric d and are μ and ν-integrable, see Remark 2.19.

Theorem 2.18 (Kantorovich-Rubinstein Theorem)
Let $X = Y$ be a Polish space, let d be a lower semi-continuous metric on X and $\mu, \nu \in \mathcal{P}_1(X)$. We find that

$$\mathcal{T}_d(\mu,\nu) = d_{KR}(\mu,\nu)$$

where

$$d_{KR}(\mu,\nu) = \sup\left\{\int_X f d\mu - \int_X f d\nu : f \in \mathrm{Lip}(X,d) \cap \mathcal{L}^1, |f|_L \leq 1\right\}. \tag{2.2}$$

Proof. By the Kantorovich Duality Theorem 2.2 we get

$$\mathcal{T}_d(\mu,\nu) = \sup_{(f,g)\in\Phi_d} J(f,g). \tag{2.3}$$

So it remains for us to prove that

$$\sup_{(f,g)\in\Phi_d} J(f,g) = d_{KR}(\mu,\nu).$$

By Lemma 2.14 and Lemma 2.17 we get

$$\sup_{(f,g)\in\Phi_d} J(f,g) = \sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(f^{dd}, f^d) = \sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(-f^d, f^d). \tag{2.4}$$

Furthermore, we obtain from Lemma 2.12 and Lemma 2.16 the following inequality

$$\sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(-f^d, f^d) \leq \sup_{\substack{f\in\mathrm{Lip}(X,d)\cap\mathcal{L}^1,\\ |f|_L\leq 1}} J(f,-f), \tag{2.5}$$

since these lemmas give $\left\{f^d : f \in \mathcal{L}^1(d\mu), f \text{ bounded}\right\} \subset \left\{f \in \mathrm{Lip}(X,d) \cap \mathcal{L}^1 : |f|_L \leq 1\right\}$.
Now we find that

$$\sup_{\substack{f\in\mathrm{Lip}(X,d)\cap\mathcal{L}^1,\\ |f|_L\leq 1}} J(f,-f) \leq \sup_{(f,g)\in\Phi_d} J(f,g) \tag{2.6}$$

since $\{(f,-f) : |f|_L \leq 1\} \subset \Phi_d$. This holds as 1-Lipschitz functions are μ and ν-integrable and satisfy $f(x) - f(y) \leq d(x,y)$, giving $(f,-f) \in \Phi_d$. In conclusion we have

$$\sup_{(f,g)\in\Phi_d} J(f,g) \leq \sup_{\substack{f\in\mathrm{Lip}(X,d)\cap\mathcal{L}^1,\\ |f|_L\leq 1}} J(f,-f) \leq \sup_{(f,g)\in\Phi_d} J(f,g),$$

giving equality. Since we have

$$\sup_{\substack{f \in \mathrm{Lip}(X,d)\cap\mathcal{L}^1,\\ |f|_L \leq 1}} J(f,-f) = \sup_{\substack{f \in \mathrm{Lip}(X,d)\cap\mathcal{L}^1,\\ |f|_L \leq 1}} \left[\int_X f\, d\mu - \int_X f\, d\nu\right] = d_{KR}(\mu,\nu) \tag{2.7}$$

the result follows. □

Remark 2.19: Since the metric d need not be the metric that defines the topology of X, a (Lipschitz) continuous function with respect to d is not necessarily measurable. Therefore we have slightly adapted Definition 1.16, the result is (2.2). In the case where d does define the topology of X we get $\mathrm{Lip}(X) = \mathrm{Lip}(X,d) \cap \mathcal{L}^1$ by Lemma 1.10.

Chapter 3

The (im)possibility of extension of the Kantorovich-Rubinstein Theorem

The Kantorovich distance we have described in Chapter 2 is actually a specific case of a family of Wasserstein metrics. Let $p \geq 1$ and take X a Polish space, $e \in X$ fixed and d an admissible metric on X. Here we take d to be the metric that defines the topology on X, otherwise the results on embeddings of measures in $\mathrm{Lip}_e(X)^*$ may not hold. We take $\mathcal{P}_p(X)$ to be the set of probability measures with finite p-th moment, i.e.

$$\mathcal{P}_p(X) = \left\{ \mu \in \mathcal{P}(X) : \int_X d(x, e)^p d\mu(x) < \infty \right\}.$$

Now we define for $\mu, \nu \in \mathcal{P}_p(X)$ the p-th Wasserstein metric by

$$\mathcal{W}_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d(x, y)^p d\pi(x, y) \right)^{1/p}.$$

For $p = 1$ we get that $\mathcal{W}_1$ is equal to $\mathcal{T}_d$ from Definition 2.15. By using the Kantorovich-Rubinstein Theorem and Theorem 1.26 we found for this case that

$$\mathcal{W}_1(\mu, \nu) = d_{KR}(\mu, \nu) = \|\mu - \nu\|_e^*$$

where the latter expression gives an extension, from $\mathcal{P}_1(X)$, to the space $\mathcal{M}_1(X)$. One might wonder if a similar result holds for $\mathcal{W}_p$ where $p > 1$. We have worked out the Kantorovich-Rubinstein Theorem in great detail because we want to investigate if by similar steps, to the ones in this proof, we can acquire a similar result for $\mathcal{W}_p$, or otherwise find the location(s) in the argument that fail for $p > 1$.

3.1 The structure of the Kantorovich-Rubinstein Theorem

We can divide the proof of the Kantorovich-Rubinstein Theorem into four parts, given by equations (2.3),(2.4), (2.5) and (2.6). We give a short summary of what these various parts entail.

(1) The first part, (2.3), relies on the use of the Kantorovich Duality Theorem 2.18.

(2) The second part, (2.4), depends on the application of a few lemmas from Chapter 2. These lemmas mostly still hold in the general case, $p > 1$.

(3) The right-hand side of equation (2.5) is equal to $d_{KR}(\mu, \nu)$, by equation (2.7). This worked for $p = 1$, now we will have to find a substitute for this supremum.

(4) The substitute from part 3 is an upper bound for the left hand side of equation (2.5), and should be chosen carefully so as to satisfy an equation similar to (2.6), in order to get equality.

3.2 Analysing the separate parts

Let X be a Polish space, d_X an admissible metric on X. Let d be a lower semi-continuous metric on X, $p > 1$ and $\mu, \nu \in P_p(X)$.
Take $c(x, y) := d(x, y)^p$. Note that c is not a metric since the triangle inequality does not hold for $p > 1$. Note also that $c = d^p$ is still lower semi-continuous, which we need for the Kantorovich Duality Theorem 2.2. Furthermore we get that $c_X = c_Y = 0$ like in Example 2.7, as a result we can use most of the lemmas proved in Chapter 2. We will now consider each separate part of the proof, and see how it holds up in the more general case.

Part 1. The first part of the proof remains largely unchanged. Namely by the Kantorovich Duality Theorem 2.2 we find that

$$\mathcal{W}_p(\mu, \nu)^p = \inf_{\Pi(\mu,\nu)} \int_{X \times X} c(x, y)\, d\pi(x, y) = \sup_{(f,g)\in\Phi_c} J(f, g).$$

Part 2. By Lemma 2.14 we find that

$$\sup_{(f,g)\in\Phi_c} J(f, g) = \sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(f^{cc}, f^c).$$

We cannot apply Lemma 2.17, because c must be a metric in order to do that. Fortunately we can alter Lemma 2.14, the result is Lemma 3.1 below. This alteration uses that for any $y \in X$ we get

$$f^c(y) = \inf_{x\in X}[c(x, y) - f(x)] \leq c(y, y) - f(y) = d(y, y)^p - f(y) = -f(y) \tag{3.1}$$

which gives $f \leq -f^c$. Consequently, we can use $-f^c$ to replace f^{cc} in the proof of Lemma 2.14. The idea of using $-f^c$ instead of f^{cc} was taken from [1, p.3].

Lemma 3.1
Let d be a lower semi-continuous metric on X. For $c = d^p$ we have

$$\sup_{(f,g)\in\Phi_c} J(f, g) \leq \sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(-f^c, f^c).$$

Proof. The proof below closely resembles the proof of Lemma 2.14.
By Lemma 2.13 we find that

$$\sup_{(f,g)\in\Phi_c} J(f, g) = \sup_{\substack{(f,g)\in\Phi_c,\\ f,g \text{ bounded}}} J(f, g).$$

Let $(f,g) \in \Phi_c$ such that f and g are bounded. Since f is bounded we find by Lemma 2.12 that f^c is integrable, note that this holds for $\mu, \nu \in \mathcal{P}_p(X)$, since they are finite. By definition $(f,g) \in \Phi_c$ gives for μ-a.e. $x \in X$ and ν-a.e. $y \in Y$ that $f(x) + g(y) \leq c(x,y)$, thus $g(y) \leq c(x,y) - f(x)$. Now by the definition we see that $g \leq f^c$ ν-a.e. so we get $J(f,g) \leq J(f,f^c)$.
By (3.1) we get $f \leq -f^c$ which then gives $J(f,g) \leq J(f,f^c) \leq J(-f^c,f^c)$. Hence

$$\sup_{(f,g)\in\Phi_c} J(f,g) = \sup_{\substack{(f,g)\in\Phi_c,\\ f,g \text{ bounded}}} J(f,g) \leq \sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(-f^c,f^c). \qquad \square$$

We do not follow the proof of Lemma 2.14 to the end to get equality, since this does not hold. We do not have that $(-f^c,f^c) \in \Phi_c$, but then we do not need equality in the first place. Note that this alteration of Lemma 2.14 is possible because for any $x \in X$ we have $c(x,x) = 0$, which we use to obtain (3.1). This would not work for any cost function.

Part 3. The next step is to find a substitute for

$$\sup_{\substack{f\in\mathrm{Lip}(X),\\ |f|_L\leq 1}} J(f,-f)$$

in equation (2.5). We can do this by proving that f^c satisfies some nice property for any bounded $f \in \mathcal{L}^1(d\mu)$. In the $p=1$ case that was 1-Lipschitzianity, which we proved in Lemma 2.16. Unfortunately, for $p>1$ there does not exist an $M \in \mathbb{R}$ such that, for any bounded $f \in \mathcal{L}^1(d\mu)$, f^c is Lipschitz with $|f^c|_L \leq M$. This follows from Example 3.2.

Example 3.2
Let $X = \mathbb{R}$ with the usual metric d. We take the cost function to be $c = d^p$. Take $\mu \in \mathcal{M}_1(\mathbb{R})$. Suppose there exists an $M \in \mathbb{R}$ such that for any bounded $f \in \mathcal{L}^1(d\mu)$ we have $|f^c|_L \leq M$. Since $p > 1$ we have $1 - 1/p > 0$, therefore we can take $K \in \mathbb{R}$ such that $K^{1-1/p} \geq M$. Now define $f \in \mathcal{L}^1(d\mu)$ by

$$f(x) := \begin{cases} K, & \text{for } x = 0 \\ 0, & \text{for } x \neq 0 \end{cases},$$

which clearly is bounded. For any $x, y \in \mathbb{R}$ with $x \neq 0$ we have $d(x,y)^p - f(x) = d(x,y)^p \geq 0$. So if for $x = 0$ we get $d(0,y)^p - f(0) < 0$ then $\inf_{x\in\mathbb{R}}[d(x,y)^p - f(x)]$ attains its infimum for $x = 0$. Otherwise it attains its infimum for $x = y$, giving $f^c(y) = 0$. We have that

$$d^p(0,0) - f(0) = -K < 0, \quad \text{and} \quad d(0,K^{1/p})^p - f(0) = K - K = 0,$$

and as a result $f^c(0) = -K$ and $f^c(K^{1/p}) = 0$. It follows that

$$|f^c|_L = \sup_{\substack{x,y\in\mathbb{R}\\ x\neq y}} \frac{|f^c(x) - f^c(y)|}{d(x,y)} \geq \frac{\left|f^c(0) - f^c(K^{1/p})\right|}{d(0,K^{1/p})} = \frac{K}{K^{1/p}} = K^{1-1/p} > M$$

giving a contradiction.

This example tells us that we *cannot* have

$$\sup_{\substack{f\in\mathcal{L}^1(d\mu),\\ f \text{ bounded}}} J(-f^c,f^c) \leq \sup_{\substack{f\in\mathrm{Lip}(X),\\ |f|_L\leq M}} J(f,-f) \qquad \text{(false)}$$

for some $M \in \mathbb{R}$, like we had in the third step of proof of the Kantorovich-Rubinstein Theorem. Note that without a requirement like $|f|_L \leq M$, this gives no appropriate substitute because

$$\sup_{f \in \mathrm{Lip}(X)} J(f, -f)$$

does not exist by Lemma 1.19, to not even speak of the problems we would encounter in step 4. To prove 1-Lipschitzianity in the $p = 1$ case, we used Lemma 1.20 to get for any $y, y' \in X$ that

$$|f^c(y) - f^c(y')| \leq \sup_{x \in X} |c(x, y) - c(x, y')| \leq c(y, y')$$

where we could use the triangle inequality since we had $c = d$ a metric. As we discussed, this proof does not work for $c = d^p$ with $p > 1$. However, we can use a similar argument to prove that f^c is locally Lipschitz.

Definition 3.3
Let (X, d) be a metric space. The space of locally Lipschitz functions is given by

$$\mathrm{LocLip}(X) := \{f \colon X \to \mathbb{R} \mid f \colon B \to \mathbb{R} \textit{ is Lipschitz for any } B \subset X \textit{ bounded}\}.$$

For any $\theta > 0$ we define for all $f \in \mathrm{LocLip}(X)$ the constant

$$|f|_{L,\theta} := \sup \left\{ \frac{|f(x) - f(y)|}{d(x, y)} : x, y \in \overline{B_e(\theta)}, x \neq y \right\}$$

where we take $|f|_{L,\theta} = 0$ if there exist no $x, y \in \overline{B_e(\theta)}$ such that $x \neq y$.

Note that proving for a function $f \colon X \to \mathbb{R}$ that $f \colon B_e(\theta) \to \mathbb{R}$ is Lipschitz for any $\theta > 0$, suffices when proving local Lipschitzianity.

Lemma 3.4
Let $f \in \mathcal{L}^1(d\mu)$ be bounded. Take $m \in \mathbb{R}$ such that $|f(x)| \leq m$ for all $x \in X$.
Then $f^c \in \mathrm{LocLip}(X)$ with $|f^c|_{L,\theta} \leq p(2\theta + (2m)^{1/p})^{p-1}$ for any $\theta > 0$.

Proof. Let $\theta > 0$. Take $M := \theta + (2m)^{1/p}$.
Let $y \in \overline{B_e(\theta)}$. We get for all $x \in [B_e(M)]^c$ that

$$d(x, y) \geq (2m)^{1/p}, \text{ hence } c(x, y) \geq 2m.$$

This means that

$$c(x, y) - f(x) \geq 2m - m = m.$$

Since

$$c(y, y) - f(y) = -f(y) \leq m$$

we find $\inf_{x \in X}[c(x, y) - f(x)] \leq m$. We may conclude that for any $y \in \overline{B_e(\theta)}$ we have

$$f^c(y) = \inf_{x \in X}[c(x, y) - f(x)] = \inf_{x \in B_e(M)}[c(x, y) - f(x)].$$

For any $y, y' \in \overline{B_e(\theta)}$ we find by Lemma 1.20 that

$$\begin{aligned} |f^c(y) - f^c(y')| &= \left| \inf_{x \in B_e(M)}[c(x, y) - f(x)] - \inf_{x \in B_e(M)}[c(x, y') - f(x)] \right| \\ &\leq \sup_{x \in B_e(M)} |c(x, y) - c(x, y')| = \sup_{x \in B_e(M)} |d(x, y)^p - d(x, y')^p|. \end{aligned}$$

Let $x' \in B_e(M)$. Since the function on $\mathbb{R}$ given by $z \mapsto z^p$ is differentiable on $[0, \infty)$ we get by the Mean Value Theorem that

$$|d(x', y)^p - d(x', y')^p| = pc^{p-1} |d(x', y) - d(x', y')|$$

for some $c \in \mathbb{R}$ between $d(x', y)$ and $d(x', y')$. Note that we have $d(x', y) \leq M + \theta$ and $d(x', y') \leq M + \theta$ since $y, y' \in \overline{B_e(\theta)}$ and $x' \in B_e(M)$. Hence we find $c \leq M + \theta$, so we get

$$|d(x', y)^p - d(x', y')^p| \leq p(M + \theta)^{p-1} |d(x', y) - d(x', y')| \leq p(M + \theta)^{p-1} d(y, y').$$

Since this inequality holds for all $x' \in B_e(M)$, we get for $y, y' \in B_e(\theta)$ that

$$|f^c(y) - f^c(y')| \leq \sup_{x \in B_e(M)} |d(x, y)^p - d(x, y')^p| \leq p(M + \theta)^{p-1} d(y, y').$$

We conclude that f^c is locally Lipschitz with, for any $\theta > 0$,

$$|f^c|_{L,\theta} \leq p(M + \theta)^{p-1} = p(2\theta + (2m)^{1/p})^{p-1}. \qquad \square$$

Remark 3.5: Note that Lemma 3.4 actually also holds for $p = 1$, and then gives $|f^c|_{L,\theta} \leq 1$ for all $\theta > 0$, i.e. that f^c is 1-Lipschitz. Therefore we can view Lemma 3.4 as a generalization of Lemma 2.16.

The bound that Lemma 3.4 provides on $|f|_{L,\theta}$ depends on f, because the m is chosen such that we have $|f(x)| \leq m$ for all $x \in X$. We have not found an upper bound for $|f^c|_{L,\theta}$, that works for any bounded $f \in \mathcal{L}^1(d\mu)$. Therefore the lemma does not give

$$\sup_{\substack{f \in \mathcal{L}^1(d\mu), \\ f \text{ bounded}}} J(-f^c, f^c) \leq \sup_{\substack{f \in \mathrm{LocLip}(X), \\ |f|_{L,\theta} \leq M(\theta)}} J(f, -f) \qquad \text{(not proven)}$$

for some $M\colon (0, \infty) \to \mathbb{R}$. And even if this were to hold, there would be the problem that $f \in \mathrm{LocLip}(X)$ is not necessarily integrable. It would be integrable if M were bounded but then we would have $f \in \mathrm{Lip}_e(S)$ with $|f|_L \leq \sup_{\theta > 0} M(\theta)$, which is impossible by Example 3.2. Despite proving that the f^c is locally Lipschitz, we have not found a suitable substitute for the right-hand side of equation (2.5). Possibly, with more time further results may be obtained.

Appendix A

Lower semi-continuity

Let X be a topological space.

Definition A.1
A function $f\colon X \to \overline{\mathbb{R}}$ is called lower semi-continuous if for every $r \in \mathbb{R}$ the set $f^{-1}((r,\infty])$ is open in X.

Remark A.2: Since for any $r \in \mathbb{R}$ we have that $f^{-1}((r,\infty])$ is open if and only if $\left[f^{-1}((r,\infty])\right]^c$ is closed we find that f is lower semi-continuous if and only if for every $r \in \mathbb{R}$

$$\left[f^{-1}((r,\infty])\right]^c = f^{-1}([-\infty,r]) = \{x \in X : f(x) \leq r\}$$

is closed. This will be used in the proof of Lemma A.6.

The following proposition gives us an alternate definition of lower semi-continuity, in case X is metrizable with metric d.

Proposition A.3
A function $f\colon X \to \overline{\mathbb{R}}$ is lower semi-continuous if and only if for every $x \in X$ and $\epsilon > 0$ there exists a $\delta > 0$ such that for all $y \in X$ satisfying $d(x,y) < \delta$ we have $f(y) > f(x) - \epsilon$.

Proof. Assume f to be lower semi-continuous. Let $x \in X$ and $\epsilon > 0$. Now $f^{-1}((f(x),\infty])$ is open and contains x. Hence we can take $\delta > 0$ such that $B_\delta(x) \subset f^{-1}((f(x),\infty])$ meaning that for every $y \in B_\delta(x)$ we have $f(y) > f(x) > f(x) - \epsilon$.
Now assume that for any $x \in X$ and $\epsilon > 0$ we have a $\delta > 0$ such that $f(y) > f(x) - \epsilon$ for all $y \in X$ satisfying $d(x,y) < \delta$. Let $r \in \mathbb{R}$ and consider $f^{-1}((r,\infty])$. Let $x \in f^{-1}((r,\infty])$. We will now find a neighbourhood for x that is contained in $f^{-1}((r,\infty])$ to prove that this set is open. We define $\epsilon := \frac{f(x)-r}{2} > 0$. Take $\delta > 0$ such that for all $y \in B_\delta(x)$ we have $f(y) > f(x) - \epsilon$. Then we find

$$f(y) > f(x) - \epsilon = f(x) - \frac{f(x)-r}{2} = \frac{1}{2}f(x) + \frac{1}{2}r > \frac{1}{2}r + \frac{1}{2}r = r.$$

This means that $B_\delta(x) \subset f^{-1}((r,\infty])$, proving that this set is open for any $r \in \mathbb{R}$, thus proving lower semi-continuity for f. □

The alternative definition provided by Proposition A.3 says that by choosing an appropriate neighbourhood for $x \in X$ we can get the value of f to be either arbitrarily close to $f(x)$ (very much like continuity) or higher than $f(x)$ (making the statement less powerful).

In [9, p.26] it is stated but not proven that for any nonnegative lower semi-continuous function we have a nondecreasing sequence of nonnegative uniformly continuous functions that converges pointwise to the lower semi-continuous function. Using the same construction given in [9, p.26] we will prove a slightly stronger statement. Let (X, d) be a metric space.

Lemma A.4
Let $f\colon X \to [0, \infty]$ be lower semi-continuous. Then there exists a nondecreasing sequence $(f_n)_{n\in\mathbb{N}}$ of bounded nonnegative Lipschitz continuous functions such that $|f_n|_L \leq n$ and $f_n \to f$ pointwise.

Proof. We define the sequence $(f_n)_{n\in\mathbb{N}}$ by

$$f_n(x) := \inf_{y\in X}[f(y) + nd(x, y)].$$

The function f_n is well-defined as f and d are bounded below. We will prove that this sequence satisfies all requirements. Since $d \geq 0$ and $f \geq 0$ we see that each f_n is indeed nonnegative and that for every $x, y \in X$ we have $f(y) + md(x, y) \geq f(y) + nd(x, y)$ for $m \geq n$ so we find $f_m \geq f_n$. To prove Lipschitz continuity let $x, x' \in X$. Let $n \in \mathbb{N}$. Then by Lemma 1.20 it follows that

$$\begin{aligned} |f_n(x) - f_n(x')| &= \left|\inf_{y\in X}[f(y) + nd(x, y)] - \inf_{y'\in X}[f(y') + nd(x', y')]\right| \\ &\leq \sup_{y\in X}|nd(x, y) - nd(x', y)| \\ &\leq nd(x, x'), \end{aligned}$$

hence f_n is Lipschitz with $|f_n|_L \leq n$.
Note that for any $x \in X$ we have

$$f_n(x) = \inf_{y\in X}[f(y) + nd(x, y)] \leq f(x) + nd(x, x) = f(x)$$

so we have that $f_n \leq f$. Let $x \in X$ and $\epsilon > 0$. We will show that there exists an $N \in \mathbb{N}$ such that $f_n(x) \geq f(x) - \epsilon$ for all $n \geq N$. This will imply $f_n \to f$ pointwise since $(f_n)_{n\in\mathbb{N}}$ is a nondecreasing sequence and $f_n \leq f$. Since f is lower semi-continuous we get by Proposition A.3 that there exists a $\delta > 0$ such that $f(y) > f(x) - \epsilon$ for all $y \in B_\delta(x)$. Now choose $N \in \mathbb{N}$ such that $N\delta \geq f(x) - \epsilon$. Let $n \geq N$. For $y \in B_\delta(x)$ we find

$$f(y) + nd(x, y) \geq f(x) - \epsilon + nd(x, y) \geq f(x) - \epsilon$$

since $nd(x, y) \geq 0$. For $y \notin B_\delta(x)$ we have

$$f(y) + nd(x, y) \geq f(y) + n\delta \geq n\delta \geq N\delta \geq f(x) - \epsilon$$

since f is nonnegative.
So for all $y \in X$ we get $f(y) + nd(x, y) \geq f(x) - \epsilon$. Hence

$$f_n(x) = \inf_{y\in X}[f(y) + nd(x, y)] \geq f(x) - \epsilon$$

holds for $n \geq N$, proving pointwise convergence. Furthermore we can assume this sequence to be bounded by replacing f_n by $\min(f_n, n)$ which clearly still satisfies all other requirements. □

In Lemma A.6 we will closely follow the proof of Hing Tong [8]. We say that a space R is perfectly normal if any closed subset of R can be written as the intersection of countably many open subsets of R. In [8, Theorem 3] it is proven that R is perfectly normal if and only if every lower semi-continuous function has a nondecreasing sequence of nonnegative and continuous functions converging to it. Since every Polish space, or in fact any metric space, is perfectly normal we can use this result. We adapt the statement and the proof to suit our situation. We will need f to be bounded and we will also use the following lemma.

Lemma A.5
For a sequence $(f_n)_{n\in\mathbb{N}}$ of Lipschitz functions satisfying $0 \le f_n \le 1$ and $|f_n|_L \le n$ let

$$g := \sum_{n=1}^{\infty} \frac{1}{2^n} f_n.$$

Then g is Lipschitz continuous and $|g|_L \le 2$.

Proof. We get that g is well defined for any $x \in X$ since

$$0 \le \sum_{n=1}^{\infty} \frac{1}{2^n} f_n(x) \le \sum_{n=1}^{\infty} \frac{1}{2^n} = 1$$

holds. We find for any $x, y \in X$ that

$$\begin{aligned} |g(x) - g(y)| &= \left| \sum_{n=1}^{\infty} \frac{1}{2^n} f_n(x) - \sum_{n=1}^{\infty} \frac{1}{2^n} f_n(y) \right| = \left| \sum_{n=1}^{\infty} \frac{1}{2^n} [f_n(x) - f_n(y)] \right| \\ &\le \sum_{n=1}^{\infty} \frac{1}{2^n} |f_n(x) - f_n(y)| \le \sum_{n=1}^{\infty} \frac{1}{2^n} |f_n|_L \, d(x,y) \\ &\le \left[\sum_{n=1}^{\infty} \frac{n}{2^n} \right] d(x,y) = 2d(x,y). \end{aligned}$$

Hence g is Lipschitz with $|g|_L \le 2$. □

Lemma A.6
Let $f\colon X \to [0,\infty]$ be a bounded lower semi-continuous function. Then there exists a nondecreasing sequence of nonnegative, Lipschitz continuous functions f_n such that $f_n \to f$ pointwise.

Proof. We will go about this step by step.
Step 1. First suppose f takes only two different values α and β. Without loss of generality we can assume $\alpha = 0$ and $\beta = 1$. The set $A := \{x \in X : f(x) = 0\}$ is closed by Remark A.2.
For any $n \in \mathbb{N}$ let

$$B_n := \left\{ x \in X : d(x, A) < \frac{1}{n} \right\}.$$

We have that the B_n are open and $B_{n+1} \subset B_n$ holds for any $n \in \mathbb{N}$. Moreover, since A is closed we have that $A = \bigcap_{n\in\mathbb{N}} B_n$. Define functions $g_n := [1 - nd(x, B_n^c)]^+$ for all $n \in \mathbb{N}$. By Lemma 1.11 we get that g_n is Lipschitz continuous with $|g_n|_L \le n$.
By definition g_n satisfies $0 \le g_n \le 1$ and $g_n(x) = 1$ for $x \in B_n^c$. Also, it satisfies $g_n(x) = 0$ for $x \in A$. Namely for $x \in A$ we have

$$d(x, B_n^c) = \inf_{y \in B_n^c} d(x,y) \ge \inf_{y \in B_n^c} \inf_{x \in A} d(x,y) = \inf_{y \in B_n^c} d(y, A) \ge \frac{1}{n}$$

since for every $y \in B_n^c$ we have $d(y, A) \geq \frac{1}{n}$. Hence, by definition of g_n, we find that $g_n(x) = 0$. Now we define

$$g := \sum_{n=1}^{\infty} \frac{1}{2^n} g_n$$

which by Lemma A.5 is Lipschitz continuous. Also we have that $g(x) = 0$ for $x \in A$. Since A is closed we have for every $x \notin A$ that $d(x, A) > 0$ hence there exists an $n \in \mathbb{N}$ such that $x \notin B_n$. This gives us that $g_n(x) = 1$, thus we get $g(x) \neq 0$ for $x \notin A$. Now $f_n(x) := \min(1, ng(x))$ gives a sequence of nonnegative Lipschitz continuous functions that is nondecreasing and it's pointwise limit is f.

Step 2. Now suppose f takes finitely many different values $\alpha_1 > \alpha_2 > ... > \alpha_m$. In order to use step 1 we want functions that take only two values. Define for any $1 \leq i \leq m-1$ the function

$$\varphi_i(x) := \begin{cases} \alpha_{i+1} & \text{for } x \in X \text{ such that } f(x) \leq \alpha_{i+1} \\ \alpha_1 & \text{for } x \in X \text{ such that } f(x) > \alpha_{i+1}. \end{cases}$$

Then $\varphi_i \geq f$ and since f only takes finitely many values we get by construction that

$$f = \min_{1 \leq i \leq m-1} \varphi_i. \tag{A.1}$$

Since f is lower semi-continuous we get by Remark A.2 for any $1 \leq i \leq m-1$ that the set $\{x \in X : f(x) \leq \alpha_{i+1}\}$ is closed. By construction we have that

$$\{x \in X : f(x) \leq \alpha_{i+1}\} = \{x \in X : \varphi_i(x) \leq \alpha_{i+1}\}.$$

This implies that $\{x \in X : \varphi_i \leq \alpha_{i+1}\}$ is closed for every $1 \leq i \leq m-1$. Since each φ_i takes only two values we get by Remark A.2 that each φ_i is lower semi-continuous. So by Step 1 we get that for each $1 \leq i \leq m-1$ there exists a sequence $(\varphi_{ij})_{j \in \mathbb{N}}$ of nonnegative Lipschitz continuous functions that is nondecreasing and converges to φ_i pointwise. Let

$$f_n(x) := \min_{1 \leq i \leq m-1} \varphi_{in}(x).$$

Then $(f_n)_{n \in \mathbb{N}}$ is a nondecreasing sequence of nonnegative and Lipschitz continuous functions by Lemma 1.5. For any $x \in X$ we have by equation (A.1) that

$$\lim_{n \to \infty} f_n(x) = \min_{1 \leq i \leq m-1} \varphi_i(x) = f(x).$$

Step 3. Now we let $f\colon X \to [0, \infty]$ be any bounded nonnegative lower semi-continuous function. Since f is bounded we may restrict ourselves to the situation where $0 \leq f(x) \leq 1$. For any $m, p \in \mathbb{Z}_{\geq 0}$ such that $1 \leq p \leq m-1$ we take the set

$$A_{mp} := \left\{ x \in X : \frac{p+1}{m} \geq f(x) > \frac{p}{m} \right\}$$

and for $p = 0$ we take

$$A_{m0} := \left\{ x \in X : \frac{1}{m} \geq f(x) \geq 0 \right\}.$$

For now we fix $m \in \mathbb{N}$. Note that

$$X = \bigcup_{0 \leq p \leq m-1} A_{mp}.$$

What we have done is, we sliced the range $[0,1]$ into m pieces and for each piece we took the subset of X that f sends to that piece. This will help us define a function that takes only finitely many values, in order to use Step 2. Let $g_m(x) = p/m$ for $x \in A_{mp}$. By construction this functions takes m values, satisfies $0 \leq g_m \leq f$ and is lower semi-continuous since $\{x \in X : f(x) \leq (p+1)/m\} = \{x \in X : g_m(x) \leq p/m\}$ is closed for any $0 \leq p \leq m-1$ by Remark A.2. Hence, by Step 2, there exists a nondecreasing sequence $(g_{mj})_{j\in\mathbb{N}}$ of nonnegative Lipschitz continuous functions converging pointwise to g_m.
Since g_m converges pointwise to f, we have that g_{mm} converges pointwise to f. We need an increasing sequence, so for all $n \in \mathbb{N}$ we take

$$f_n := \max_{1\leq i\leq n} g_{ii}.$$

We find that f_n is Lipschitz continuous by Lemma 1.5. Furthermore the f_n are nonnegative and the sequence is now nondecreasing since from n to $n+1$ we just add another function to take the maximum over. Also $(f_n)_{n\in\mathbb{N}}$ still converges pointwise to f since $g_{ii} \leq f$ for all $i \in \mathbb{N}$. Hence this sequence satisfies all requirements. □

Bibliography

[1] Basso, Giuliano. (2015) , *A Hitchhiker's guide to Wasserstein distances.*

[2] Billingsley, Patrick. (1999), *Convergence of probability measures*, second ed. John Wiley & Sons Inc., New York, A Wiley-Interscience Publication.

[3] Bogachev, Vladimir I. (2007), *Measure theory*, Vol. 2, Springer Berlin Heidelberg.

[4] Burckel, Robert B. & Bauer, Heinz. (2011), *Measure and Integration Theory*, Berlin, Boston: De Gruyter.

[5] Dudley, Richard M. (1966), *Convergence of Baire measures*, Stud. Math. 27, 251–268.

[6] Hanin, Leonid G. (1999), *An Extension of the Kantorovich Norm*, Contemporary Math. Vol. 226.

[7] Hille, Sander C. and Worm, Daniël T.H. (2009), *Embedding of semigroups of Lipschitz maps into positive linear semigroups on Banach spaces generated by measures*, Integr. Equ. Oper. Theory 63, 351–371.

[8] Tong, Hing. (1952), *Some characterizations of normal and perfectly normal spaces*, Duke Math. J. 19, no. 2, 289–292.

[9] Villani, Cédric. (2003), *Topics in Optimal Transportation*, American Math. Society.

[10] Worm, Daniël T.H. (2010), *Semigroups on Spaces of Measures* (Doctoral thesis, Leiden University).