

## CONTENTS

1. Optimal Transport	1
1.1. Existence of Optimal Transport	2
1.2. Kantorovich as a Relaxation of Monge	4
1.3. Kantorovich Duality and the Kantorovich-Rubinstein Theorem	4
2. Basic Properties of the Wasserstein Space	5
2.1. The Wasserstein Distance	5
2.2. Triangular Inequality: The case $p = 1$	6
2.3. Triangular Inequality: The case $p \in (1, +\infty)$	6
2.4. Topology of the Wasserstein Space	8
3. Curves and Geodesics in the Metric Space	8
4. AC curves in the Wasserstein Space and the Benamou-Brenier Formula	8

## 1. OPTIMAL TRANSPORT

Consider the problem of optimally transporting some mass. The spatial distributions of the mass can be described (after normalizing the total amount of mass to unity) by a probability measure on some space. Let  $\mu \in \mathcal{P}(X)$  and  $\nu \in \mathcal{P}(Y)$  describe the initial and target mass distribution, respectively. Here,  $\mathcal{P}$  is used to denote the space of Borel probability measures on a given space.

Focusing solely on the initial distribution of each piece of mass, we can describe each **transport plan** using a probability measure on the product space,  $\gamma \in \mathcal{P}(X \times Y)$ , where  $\gamma(A \times B)$  gives the amount of mass initially located in  $A$  and subsequently transferred to  $B$ . The requirement that the initial and target mass distributions are respected then amounts to

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

for all measurable sets  $A \subseteq X$  and  $B \subseteq Y$ . That is, each transport plan  $\gamma$  has marginals  $\mu$  and  $\nu$ . Each measure with such properties will be called a **coupling** of  $\mu$  and  $\nu$ . The set of all couplings of  $\mu$  and  $\nu$  will be denoted by  $\Pi(\mu, \nu)$ .

Now, let the measurable **cost function**  $c : X \times Y \rightarrow [0, +\infty]$  describe the cost of transporting a unit mass from point  $x \in X$  to point  $y \in Y$  by  $c(x, y)$ . For each transport plan  $\gamma \in \Pi(\mu, \nu)$ , the total transport cost is then given by

$$\mathcal{K}(\gamma) := \int_{X \times Y} c(x, y) \, d\gamma(x, y).$$

The first version of the optimal transport problem, the **Kantorovich problem**, is to minimize this cost:

*Problem* (Kantorovich Problem). Minimize

$$\mathcal{K}(\gamma) := \int_{X \times Y} c(x, y) \, d\gamma(x, y).$$

over all  $\gamma \in \Pi(\mu, \nu)$ . The infimum cost for cost  $c$  will be denoted by  $\mathcal{C}_c(\mu, \nu)$ .

**linearity**

The Kantorovich problem can be considered as a relaxation of a related, and historically earlier, version of the optimal transport problem, the **Monge problem**. In the Monge problem, we impose the additional constraint that mass from each point  $x \in X$  cannot be divided. With this constraint, a transport plan can be described by a **transport map**, a measurable function  $T : X \rightarrow Y$  such that each piece of mass initially located at  $x$  is transported to  $T(x)$ . The requirement that the initial and target mass distributions are respected then amounts to

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

for all measurable sets  $B \subseteq Y$ . Equivalently, the transport map  $T$  must satisfy  $T_{\#}\mu = \nu$ , where  $T_{\#}\mu$  is the **pushforward** of  $\mu$  by  $T$ . For a transport map  $T$  with this property, we have the associated transport plan  $\gamma_T = (\text{Id}, T)_{\#}\mu$  and cost

$$\mathcal{M}(T) := \mathcal{K}(\gamma_T) = \int_{X \times Y} c(x, y) \, d\gamma_T(x, y) = \int_X c(x, T(x)) \, d\mu(x).$$

The Monge problem can be summarized as follows:

*Problem* (Monge). Minimize

$$\mathcal{M}(T) := \int_X c(x, T(x)) \, d\mu(x).$$

for all measurable maps  $T : X \rightarrow Y$  such that  $T_{\#}\mu = \nu$ .

Since for each transport map  $T$ , there is an associated transport plan  $\gamma_T$ , we may view the Kantorovich problem as a relaxation of the Monge problem where we minimize over a larger set of transport plans. From this we have

$$\inf_{\gamma \in \Pi(\mu, \nu)} \mathcal{K}(\gamma) \leq \inf_{T: T_{\#}\mu = \nu} \mathcal{M}(T).$$

The next two subsections will discuss the existence of optimal transport for both problems and when equality can be achieved.

**1.1. Existence of Optimal Transport.** We start with the existence of optimal transport for the Kantorovich problem. A standard compactness and lower semi-continuity argument will be used: we choose a sequence  $\gamma_n \in \Pi(\mu, \nu)$  such that  $\mathcal{K}(\gamma_n) \rightarrow \inf_{\gamma \in \Pi(\mu, \nu)} \mathcal{K}(\gamma)$ , extract by compactness a subsequence that converges to some  $\gamma \in \Pi(\mu, \nu)$ , and show that  $\gamma$  achieves the infimum.

The compactness of  $\Pi(\mu, \nu)$  will be given by Prokhorov's theorem, which we now recall:

**Theorem 1.1** (Prokhorov). *Let  $(X, d)$  be a complete separable metric space. A set  $\mathcal{M} \subset \mathcal{P}(X)$  is relatively compact in the weak topology if and only if it is **tight**, that is, for every  $\varepsilon > 0$ , there exists a compact set  $K \subset X$  such that  $\sup_{\mu \in \mathcal{M}} \mu(K^c) < \varepsilon$ .*

**Proof.** See [Zitkovic2013weak]. □

can we improve this to Polish?

Find reference that proves compactness instead of just sequential compactness.

When  $X, Y \subset \mathbb{R}^d$ , by considering the increasing sequence  $([-n, n]^d)_{n \in \mathbb{N}} \nearrow \mathbb{R}^d$ , we can find compact sets  $K_X \subset X$  and  $K_Y \subset Y$  such that  $\mu(K_X^c), \nu(K_Y^c) < \varepsilon/2$ . More generally, when  $X, Y$  are Polish spaces,  $K_X$  and  $K_Y$  can be found using **Ulam's lemma**. Now, for any  $\gamma \in \Pi(\mu, \nu)$ , we have

$$\gamma((K_X \times K_Y)^c) \leq \gamma(K_X^c \times Y) + \gamma(X \times K_Y^c) \leq \mu(K_X^c) + \nu(K_Y^c) \leq \varepsilon.$$

Thus  $\Pi(\mu, \nu)$  is tight and, by Theorem 1.1, relatively weakly compact. This is sufficient for proving existence for the Kantorovich problem, but we state the following obvious generalization, which will be useful later on:

**Proposition 1.2.** *Let  $\mathcal{M}$  and  $\mathcal{N}$  be two tight sets of measures. Then the set of all couplings*

$$\{\gamma \in \Pi(\mu, \nu) : \mu \in \mathcal{M}, \nu \in \mathcal{N}\}$$

*is relatively compact in the weak topology. In particular, if  $X$  and  $Y$  are Polish and  $\mu \in \mathcal{P}(X)$ ,  $\nu \in \mathcal{P}(Y)$ , then  $\Pi(\mu, \nu)$  is relatively compact in the weak topology.*

The other ingredient of the existence proof, that  $\mathcal{K}$  is lower semi-continuous, is an easy consequence of the following approximation theorem:

**Lemma 1.3.** *Let  $f$  be a nonnegative lower semi-continuous function on a metric space  $X$ . Then, there exists a sequence  $(f_n)_{n \in \mathbb{N}}$  of bounded nonnegative uniformly continuous functions converging pointwise monotonically to  $f$ .*

**Proof.** For each  $n$ , define

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

It is clear that each  $f_n$  is nonnegative and  $f_n(x)$  is nondecreasing in  $n$  for each  $x$ . As the upper envelope of  $n$ -Lipschitz functions, each  $f_n$  is also  $n$ -Lipschitz and thus uniformly continuous.

Finally, we show that  $f_n$  converges to  $f$  pointwise. Fix any  $x \in X$ . By  $f$  being lower semi-continuous, for each  $\varepsilon > 0$ , there exists some  $\delta > 0$  such that  $f(y) \geq f(x) - \varepsilon$  for all  $y \in B(x, \delta)$ . Find large  $N$  such that  $N\delta > f(x)$ . We have when  $y \notin B(x, \delta)$  that

$$f(y) + Nd(x, y) \geq N\delta > f(x) - \varepsilon.$$

Similarly, when  $y \in B(x, \delta)$ , we have

$$f(y) + Nd(x, y) \geq f(x) - \varepsilon.$$

Thus  $\lim_{n \rightarrow \infty} f_n(x) \geq f_N(x) \geq f(x) - \varepsilon$ . We send  $\varepsilon \rightarrow 0$  to conclude the proof.  $\square$

**Corollary 1.4.** *The functional  $\mathcal{K}$  is lower semi-continuous in the weak topology if the cost  $c$  is lower semi-continuous.*

**Proof.** Let  $c_n$  be the sequence of bounded continuous functions constructed in Lemma 1.3 converging pointwise monotonically to  $c$ . For each  $n$ , the functional

$$\mathcal{K}_n(\gamma) := \int c_n(x, y) d\gamma(x, y)$$

is continuous and in particular lower semi-continuous in the weak topology. The Monotone Convergence Theorem implies that  $\mathcal{K}_n \rightarrow \mathcal{K}$  pointwise. Since  $\mathcal{K}_n$  is nondecreasing in  $n$ , we actually have  $\mathcal{K} = \sup_n \mathcal{K}_n$ . As the pointwise supremum of lower semi-continuous functions,  $\mathcal{K}$  is lower semi-continuous.  $\square$

We are now equipped to prove the existence of an optimal transport plan.

add ref

**Theorem 1.5** (Kantorovich). *Let  $X$  and  $Y$  be Polish,  $\mu \in \mathcal{P}(X)$  and  $\nu \in \mathcal{P}(Y)$ , and  $c$  be a lower semi-continuous cost function.*

Do  $\mathbb{R}^d$  instead?

**Proof.** Let  $(\gamma_n)_{n \in \mathbb{N}} \subseteq \Pi(\mu, \nu)$  be a minimizing sequence for the Kantorovich problem. Up to extraction of a subsequence, we may assume that  $\gamma_n \rightharpoonup \gamma$  for some  $\gamma \in \Pi(\mu, \nu)$ . By Corollary 1.4,  $\inf \mathcal{K} = \liminf_n \mathcal{K}(\gamma_n) \geq \mathcal{K}(\gamma)$ . In particular,  $\gamma$  achieves the infimum.  $\square$

Conditions for the existence of optimal transport maps for the Monge problem are much more delicate. We state below a set of sufficient conditions and omit the proof:

**Theorem 1.6** (Gangbo, McCann). *Let  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$  and  $c$  be a strictly convex, superlinear cost function such that the total transport cost from  $\mu$  to  $\nu$  is not always infinite. Assume moreover that  $\mu$  is absolutely continuous with respect to the Lebesgue measure. Then, the optimal transport plan is unique and induced by a transport map. In particular, the Monge problem admits a unique solution.*

**Proof.** See [Villani2003].  $\square$

**1.2. Kantorovich as a Relaxation of Monge.** Under suitable conditions, we can show that  $\mathcal{K}$  is a relaxation of  $\mathcal{M}$  in the following sense:

**Definition 1.7** (Relaxation). Let  $F : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be a functional bounded from below. Its **relaxation**  $\bar{F}$  is the maximal lower semi-continuous functional such that  $\bar{F} \leq F$ . The relaxation is well-defined since the pointwise supremum of any family of lower semi-continuous functionals is lower semi-continuous.

The following observation gives the significance of the relaxation:

**Proposition 1.8.** *Let  $\bar{F}$  be the relaxation of  $F$ . Then,  $\inf F = \inf \bar{F}$ .*

**Proof.** From the definition above, we have  $\bar{F} \leq F$  pointwise and thus  $\inf \bar{F} \leq \inf F$ . For the opposite inequality, note that the constant functional  $x \mapsto \inf F$  is lower semi-continuous and bounded above by  $F$ . Since  $\bar{F}$  is the maximal such functional, we have  $\bar{F} \geq \inf F$  pointwise and so  $\inf \bar{F} \geq \inf F$ .  $\square$

We will see that under suitable conditions, the functional  $\mathcal{K}$  is the relaxation of  $\mathcal{M}$ . In particular, when the optimal transport map  $T$  for  $\mathcal{M}$  exists, then by the result above we have

$$\mathcal{K}(\gamma_T) = \mathcal{M}(T) = \inf \mathcal{M} = \inf \mathcal{K}.$$

And so the optimal transport plan for the Kantorovich problem is given by  $\gamma_T$ .

**[TODO: Prove relaxation]**

A set of sufficient conditions for Kantorovich is the following:  $X, Y$  Polish.  $c$  lower semi-continuous.  $\mu$  atomless.

**1.3. Kantorovich Duality and the Kantorovich-Rubinstein Theorem.**

**Theorem 1.9** (Kantorovich Duality). *Let  $X, Y$  be Polish and  $c$  be a lower semi-continuous cost function. Let  $\Phi_c$  consist of all pairs  $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$  such that*

$$\varphi(x) + \psi(y) \leq c(x, y)$$

*for  $\mu$ -a.e.  $x \in X$  and  $\nu$ -a.e.  $y \in Y$ . Then,*

$$\mathcal{C}_c(\mu, \nu) = \sup_{(\varphi, \psi) \in \Phi_c} \left( \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y) \right).$$

I want to mention the sketch proofs of these two results since they are used in later on.

A formal proof here.

**Theorem 1.10** (Kantorovich-Rubinstein). *Let  $X = Y$  be Polish and endowed with a lower semi-continuous metric  $d$ . Let  $\text{Lip}(X)$  denote the space of all Lipschitz functions on  $X$ , and*

$$\|\varphi\|_{\text{Lip}} := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x, y)}.$$

Then,

$$\mathcal{C}_d(\mu, \nu) = \sup \left\{ \int_X \varphi \, d(\mu - \nu); \quad \varphi \in L^1(d|\mu - \nu|); \quad \|\varphi\|_{\text{Lip}} \leq 1 \right\}.$$

**Proof.**

## 2. BASIC PROPERTIES OF THE WASSERSTEIN SPACE

**2.1. The Wasserstein Distance.** Starting from this section, we will mainly focus on the case where  $X = Y = \Omega \subset \mathbb{R}^d$  and costs of the form  $c(x, y) = |x - y|^p$  for  $p \geq 1$ . Given two probability measures  $\mu, \nu \in \mathcal{P}(\Omega)$ , the optimal transport cost  $\mathcal{C}_c(\mu, \nu)$  may be viewed as a measure of distance between  $\mu$  and  $\nu$ . When  $\Omega$  is unbounded, we will restrict our attention to the subset of  $\mathcal{P}(\Omega)$  consisting of all probability measures with finite  $p$ -th moment:

$$\mathcal{P}_p(\Omega) := \left\{ \mu \in \mathcal{P}(\Omega); \quad \int_{\Omega} |x|^p \, d\mu(x) < +\infty \right\}.$$

This, as we will see in the next proposition, will ensure the finiteness of the **Wasserstein distance of order  $p$** , defined by

$$W_p(\mu, \nu) := (C_{d^p}(\mu, \nu))^{1/p} = \left( \min_{\gamma \in \Pi(\mu, \nu)} \int_{\Omega \times \Omega} |x - y|^p \, d\gamma(x, y) \right)^{1/p}.$$

Here, the existence of the min is established by Theorem 1.5. The space  $\mathcal{P}_p(\Omega)$  endowed with the (to be proven) metric  $W_p$  is called the **Wasserstein space of order  $p$** .

A few immediate observations can be made at this point:

### Proposition 2.1.

- (i) If  $\mu, \nu \in \mathcal{P}_p(\Omega)$ , then  $W_p(\mu, \nu) < +\infty$ .
- (ii) If  $1 \leq p \leq q$ , then  $\mathcal{P}_q(\Omega) \subset \mathcal{P}_p(\Omega)$  and  $W_p(\mu, \nu) \leq W_q(\mu, \nu)$ .
- (iii) When  $\Omega$  is bounded,  $\mathcal{P}_p(\Omega) = \mathcal{P}(\Omega)$  for all  $p \geq 1$ . Moreover, there exists a constant  $C > 0$  such that  $W_p^p(\mu, \nu) \leq CW_1(\mu, \nu)$ .

**Proof.**

- (i) Let  $\gamma \in \Pi(\mu, \nu)$  be any transport plan. From  $|x - y|^p \leq C(|x|^p + |y|^p)$  we have

$$\int |x - y|^p \, d\gamma \leq C \left( \int |x|^p \, d\mu + \int |y|^p \, d\nu \right) < +\infty.$$

- (ii) Since  $x \mapsto |x|^{q/p}$  is convex, we have by Jensen's inequality that

$$\left( \int |x|^p \, d\mu \right)^{\frac{q}{p}} \leq \int |x|^q \, d\mu.$$

Higher moment finiteness implies lower moment finiteness.

Similarly, for the second result, let  $\gamma$  be an optimal transport plan for  $W_q(\mu, \nu)$  and we have

$$W_p^p(\mu, \nu) \leq \int |x - y|^p \, d\gamma \leq \left( \int |x - y|^q \, d\gamma \right)^{\frac{p}{q}} = W_q(\mu, \nu)^p.$$

(iii) From (ii) we have that  $\mathcal{P}_p(\Omega) \subset \mathcal{P}(\Omega)$ . The reverse inclusion is given by  $\int |x|^p d\mu \leq \text{diam}(\Omega)^p$ . Now, if  $\gamma$  is an optimal transport plan for  $W_1(\mu, \nu)$ , then

$$W_p^p(\mu, \nu) \leq \int |x - y|^p d\gamma \leq \text{diam}(\Omega)^{p-1} \int |x - y| d\gamma = \text{diam}(\Omega)^{p-1} W_1(\mu, \nu).$$

□

We next turn our attention to establishing that  $W_p$  is indeed a metric on  $\mathcal{P}_p(\Omega)$ . Positive definiteness and symmetry are quite easy:

**Proposition 2.2.** *Let  $p \in [1, +\infty)$ . The function  $W_p$  is symmetric and nonnegative. Moreover,  $W_p(\mu, \nu) = 0$  if and only if  $\mu = \nu$ .*

**Proof.** The nonnegativity of  $W_p$  is inherited from that of the cost function. Symmetry is obtained by noting that for each transport plan  $\gamma \in \Pi(\mu, \nu)$ , we can define a “reverse plan”  $S_\# \gamma \in \Pi(\nu, \mu)$ , where  $S : \Omega \times \Omega \rightarrow \Omega \times \Omega$  is defined by  $S(x, y) := (y, x)$ . We have

$$W_p^p(\nu, \mu) \leq \int |x - y|^p dS_\# \gamma(x, y) = \int |y - x|^p d\gamma(x, y).$$

Taking the infimum over all  $\gamma \in \Pi(\mu, \nu)$  gives  $W_p(\nu, \mu) \leq W_p(\mu, \nu)$ . The reverse inequality is obtained by interchanging  $\mu$  and  $\nu$ .

Finally, suppose

$$W_p(\mu, \nu) = \left( \int |x - y|^p d\gamma(x, y) \right)^{\frac{1}{p}} = 0.$$

Then  $\gamma$  is supported on  $\{x = y\}$ . Thus for any measurable set  $A \subset \Omega$  we have

$$\begin{aligned} \mu(A) &= \gamma(A \times \Omega) = \gamma((A \times \Omega) \cap \{x = y\}) \\ &= \gamma(\Omega \times A) = \nu(A), \end{aligned}$$

which shows that  $\mu = \nu$ . □

Establishing the triangle inequality is more involved. We will deal with the cases  $p = 1$  and  $p > 1$  separately. For  $p = 1$ , the Kantorovich-Rubinstein theorem will be our main point of attack. For  $p > 1$ , we will leverage the strict convexity of the cost.

## 2.2. Triangular Inequality: The case $p = 1$ . [TODO: TODO]

**2.3. Triangular Inequality: The case  $p \in (1, +\infty)$ .** When  $\mu$  and  $\rho$  are both absolutely continuous with respect to the Lebesgue measure, optimal transport maps exist by Theorem 1.6, and we can rather easily reduce the problem to triangle inequality in the  $L^p$  space. When this is not true, we can approximate  $\mu$  and  $\rho$  with absolutely continuous measures obtained by convolution, and pass to the limit with aid from the following lemma:

**Lemma 2.3.** *Consider  $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$  and  $\chi_\varepsilon$  an even mollifier in  $L^1$  such that  $\int_{\mathbb{R}^d} \chi_\varepsilon(x) dx = 1$ ,  $\chi_\varepsilon(x) = \varepsilon^{-d} \chi_1(x/\varepsilon)$ , and  $\chi_1$  is compactly supported on the unit ball. Write  $\mu_\varepsilon := \mu * \chi_\varepsilon$  and  $\nu_\varepsilon := \nu * \chi_\varepsilon$ . We have*

- (i)  $W_p(\mu_\varepsilon, \nu_\varepsilon) \leq W_p(\mu, \nu)$ ,
- (ii)  $\lim_{\varepsilon \rightarrow 0} W_p(\mu_\varepsilon, \nu_\varepsilon) = W_p(\mu, \nu)$ .

I guess I don't need this part of the result?

**Proof.** (i)  $W_p(\mu_\varepsilon, \nu_\varepsilon) \leq W_p(\mu, \nu)$ . Let  $\gamma$  be the optimal transport plan for  $W_p(\mu, \nu)$ . We use  $\gamma$  to define a candidate transport plan between  $\mu_\varepsilon$  and  $\nu_\varepsilon$ . The functional

$$L : \varphi(x, y) \mapsto \int_{\mathbb{R}^d \times \mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(x - z, y - z) \chi_\varepsilon(z) dz d\gamma(x, y)$$

is positive when  $\varphi \geq 0$ , linear, and continuous with respect to the sup norm. By the Riesz representation theorem, it defines a unique measure  $\gamma_\varepsilon$  by  $L(\varphi) = \int \varphi d\gamma_\varepsilon$ . We check that it lives in  $\Pi(\mu_\varepsilon, \nu_\varepsilon)$ : for any  $\varphi \in C_c(\mathbb{R}^d)$ ,

$$\begin{aligned} \int \varphi(x) d\gamma_\varepsilon &= \iint \varphi(x-z) \chi_\varepsilon(z) dz d\mu(x) \\ &= \iint \varphi(z) \chi_\varepsilon(x-z) d\mu(x) dz \\ &= \int \varphi(z) \int \chi_\varepsilon(z-x) d\mu(x) dz = \int \varphi d\mu_\varepsilon, \end{aligned}$$

where the second line comes from interchanging the two integrals and a change of variable, and the third line comes from  $\chi_\varepsilon$  being even. This shows that the first marginal of  $\gamma_\varepsilon$  is  $\mu_\varepsilon$ . An analogous argument shows that the second marginal is  $\nu_\varepsilon$ . Using this transport plan, we have the bound

$$W_p^p(\mu_\varepsilon, \nu_\varepsilon) \leq \iint |x-z-y+z|^p \chi_\varepsilon(z) dz d\gamma_\varepsilon = \int |x-y|^p d\gamma_\varepsilon = W_p^p(\mu, \nu).$$

(ii)  $\lim_{\varepsilon \rightarrow 0} W_p(\mu_\varepsilon, \nu_\varepsilon) = W_p(\mu, \nu)$ . The inequality above gives  $\limsup W_p(\mu_\varepsilon, \nu_\varepsilon) \leq W_p(\mu, \nu)$ . It thus suffices to show

$$\liminf W_p(\mu_{\varepsilon_n}, \nu_{\varepsilon_n}) \geq W_p(\mu, \nu)$$

for an arbitrary sequence  $\varepsilon_n \rightarrow 0$ . This will be done using the lower semi-continuity of  $\mathcal{K}$ .

We start with showing that the families  $\{\mu_{\varepsilon_n}\}$  and  $\{\nu_{\varepsilon_n}\}$  are tight. Find for each  $\delta > 0$  a compact  $K_0$  such that  $\mu(K_0^c) < \delta$ . Then, enlarge  $K_0$  by 1, the radius of the support of  $\chi_1$ , to obtain

$$K := \bigcup_{x \in K_0} \overline{B}(x, 1) = (x \mapsto d(x, K_0))^{-1}([0, 1]).$$

It is easy to verify that  $K$  is compact (being closed and bounded). We have

$$\mu_\varepsilon(K) = \int \mu(K-x) \chi_\varepsilon(x) dx \leq \int \mu(K_0) \chi_\varepsilon(x) dx = \mu(K_0),$$

where the inequality follows from the fact that  $K_0 + x \subset K$  for each  $x \in \text{supp } \chi_1 \subset B(0, 1)$ . The same argument shows that  $\{\nu_{\varepsilon_n}\}$  is tight.

Now let  $\gamma_{\varepsilon_n}$  be the optimal transport plan for  $W_p(\mu_{\varepsilon_n}, \nu_{\varepsilon_n})$  and note that we have  $\mu_{\varepsilon_n} \rightarrow \mu$  and  $\nu_{\varepsilon_n} \rightarrow \nu$ . We pick a subsequence that attains  $\liminf W_p(\mu_{\varepsilon_n}, \nu_{\varepsilon_n})$ , and, by Proposition 1.2, a further subsequence  $\varepsilon_{n_k}$  such that  $\gamma_{\varepsilon_{n_k}}$  converges weakly to some  $\gamma$ . We check that  $\gamma \in \Pi(\mu, \nu)$ : for each  $\varphi \in C_b(\Omega)$ , we have by a change of variable that

$$\int \varphi d(\pi_{1\#} \gamma_{\varepsilon_{n_k}}) = \int \varphi \circ \pi_1 d\gamma_{\varepsilon_{n_k}},$$

where  $\pi_1 : (x, y) \mapsto x$  projects to the first coordinate. As we send  $n \rightarrow \infty$ , the left side converges to  $\int \varphi d\mu$  by the fact that  $\pi_{1\#} \gamma_{\varepsilon_{n_k}} = \mu_{\varepsilon_{n_k}} \rightarrow \mu$ . Similarly, from  $\gamma_{\varepsilon_{n_k}} \rightarrow \gamma$  we know that the right side converges to  $\int \varphi \circ \pi_1 d\gamma = \int \varphi d(\pi_{1\#} \gamma)$ . We then have by uniqueness of limits that  $\int \varphi d\mu = \int \varphi d(\pi_{1\#} \gamma)$ . Since  $\varphi \in C_b(\Omega)$  is arbitrary,  $\gamma$  has first marginal  $\mu$ . Its second marginal is by the same argument  $\nu$ .

We now have

$$\liminf W_p(\mu_{\varepsilon_n}, \nu_{\varepsilon_n}) = \lim \mathcal{K}(\gamma_{\varepsilon_{n_k}}) \geq \mathcal{K}(\gamma) \geq W_p(\mu, \nu),$$

where the equality comes from how we selected the subsequence  $\varepsilon_{n_k}$ , the first inequality from Corollary 1.4, and the second inequality from  $\gamma$  being an admissible transport plan.

Since  $W_p(\mu_\varepsilon, \nu_\varepsilon) \leq W_p(\mu, \nu)$  for each  $\varepsilon$ , we conclude that  $\lim W_p(\mu_{\varepsilon_n}, \nu_{\varepsilon_n}) = W_p(\mu, \nu)$ .  $\square$

**Theorem 2.4.**  $W_p$  is a metric on  $\mathcal{P}_p(\Omega)$  for each  $p \geq 1$ .

**Proof.** Finally, to show triangular inequality, we suppose first that  $\mu$  and  $\rho$  are absolutely continuous. By Theorem 1.6 there exists optimal transport maps  $S$  and  $T$  such that  $\rho = S_\# \mu$  and  $\nu = T_\# \rho$ . Since  $(T \circ S)_\# \mu = T_\# \rho = \nu$ ,  $T \circ S$  is a transport map from  $\mu$  to  $\nu$ . Thus we have

$$\begin{aligned} W_p(\mu, \nu) &\leq \left( \int |T \circ S - \text{Id}|^p \, d\mu \right)^{1/p} \leq \left( \int |S - \text{Id}|^p \, d\mu \right)^{1/p} + \left( \int |T \circ S - S|^p \, d\mu \right)^{1/p} \\ &= \left( \int |S - \text{Id}|^p \, d\mu \right)^{1/p} + \left( \int |T - \text{Id}|^p \, d\rho \right)^{1/p} \\ &= W_p^p(\mu, \rho) + W_p^p(\rho, \nu), \end{aligned}$$

where the second inequality follows from the triangular inequality for  $L^p(\mu)$ , and the second line follows from a change of variable.

When  $\mu$  or  $\rho$  is not absolutely continuous, the discussion above gives

$$W_p(\mu_\varepsilon, \nu_\varepsilon) \leq W_p(\mu_\varepsilon, \rho_\varepsilon) + W_p(\rho_\varepsilon, \nu_\varepsilon),$$

where  $\mu_\varepsilon, \nu_\varepsilon, \rho_\varepsilon$  are as defined in Lemma 2.3. The same Lemma then allows us to pass to the limit.  $\square$

## 2.4. Topology of the Wasserstein Space.

### 3. CURVES AND GEODESICS IN THE METRIC SPACE

#### 4. AC CURVES IN THE WASSERSTEIN SPACE AND THE BENAMOU-BRENIER FORMULA