ADEN CHEN

Contents

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 γ has marginals μ and ν. Each measure with such properties will be called a **coupling** of μ and ν. The set of all couplings of μ and ν will be denoted by $\Pi(\mu, \nu)$.

Now, let the measurable **cost function** $c : X \times Y \to [0, +\infty]$ 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

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

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

Problem (Kantorovich Problem). Minimize

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

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

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 $T : X \to Y$ 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 μ by T. For a transport map T with this property, we have the associated transport plan $\gamma_T = (\mathrm{Id}, T)_{\#}\mu$ and cost

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

The Monge problem can be summarized as follows:

Problem (Monge). Minimize

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

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

Since for each transport map T, there is an associated transport plan γ_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

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

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 $\mathcal{K}(\gamma_n) \to \inf_{\gamma\in\Pi(\mu,\nu)} \mathcal{K}(\gamma)$, extract by compactness a subsequence that converges to some $\gamma \in \Pi(\mu, \nu)$, and show that γ 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 $\sup_{\mu\in\mathcal{M}} \mu(K^{\mathsf{c}}) < \epsilon$.*

check if can improve to Polish

Proof. See [2, Theorem 7.9]. [TODO: 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 $\mu(K_X^{\mathsf{c}}), \nu(K_Y^{\mathsf{c}}) < \varepsilon/2$. 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

add ref

$$\gamma\left((K_X \times K_Y)^{\mathsf{c}}\right) \le \gamma(K_X^{\mathsf{c}} \times Y) + \gamma(X \times K_Y^{\mathsf{c}}) \le \mu(K_X^{\mathsf{c}}) + \nu(K_Y^{\mathsf{c}}) \le \varepsilon.$$

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 $\mathscr{K}$ is lower semicontinuous (l.s.c.), is an easy consequence of the following approximation theorem:

Lemma 1.3. *Let f be a nonnegative l.s.c. function on a metric space X. Then, there exists a sequence $(f_n)_{n\in\mathbb{N}}$ of bounded nonnegative Lipschitz 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.

Finally, we show that f_n converges to f pointwise. Fix any $x \in X$. By f being l.s.c. for each $\varepsilon > 0$, there exists some $\delta > 0$ such that $f(y) \ge f(x) - \epsilon$ 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

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

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

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

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

Corollary 1.4. *If f is a l.s.c. bounded below function on a metric space X, then the functional $F : \mu \mapsto \int f \,\mathrm{d}\mu$ is lower semi-continuous on $\mathcal{P}(X)$. In particular, if c is l.s.c. and bounded below, then the functional $\mathscr{K}$ is l.s.c.*

Proof. Let f_n be the sequence of bounded continuous functions constructed in Lemma 1.3 converging pointwise monotonically to f. For each n, the functional

$$F_n(\mu) := \int f_n(x,y)\,\mathrm{d}\mu$$

is continuous and in particular l.s.c. in the weak topology. Using the Monotone Convergence Theorem we may deduce that $F_n \to F$ pointwise. Since F_n is nondecreasing in n, we actually have $F = \sup_n F_n$. As the pointwise supremum of l.s.c. functions, $\mathscr{K}$ is lower semi-continuous. □

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

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 l.s.c. cost function.*

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, $\inf \mathscr{K} = \liminf_n \mathscr{K}(\gamma_n) \geq \mathscr{K}(\gamma)$. In particular, γ achieves the infimum. □

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 μ to ν is not always infinite. Assume moreover that μ 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 [1, Theorem 2.44]. □

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

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

The following observation gives the significance of the relaxation:

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

Proof. From the definition above, we have $\overline{F} \leq F$ pointwise and thus $\inf \overline{F} \leq \inf F$. For the opposite inequality, note that the constant functional $x \mapsto \inf F$ is l.s.c. and bounded above by F. Since $\overline{F}$ is the maximal such functional, we have $\overline{F} \geq \inf F$ pointwise and so $\inf \overline{F} \geq \inf F$. □

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

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

And so the optimal transport plan for the Kantorovich problem is given by γ_T.
[TODO: Prove relaxation]
A set of sufficient conditions for Kantorovich is the following: X, Y Polish. c l.s.c. μ atomless.

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

Theorem 1.9 (Kantorovich Duality). *Let X, Y be Polish and c be a l.s.c. cost function. Let φ_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 μ-a.e. $x \in X$ and ν-a.e. $y \in Y$. Then,*

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

We provide in the following a formal proof, which shows how the sup appears via a minimax principle. The full proof of this theorem is quite involved and can be found in [1, pp. 26–33].

Proof *(A Formal Proof)***.** We note first the following: for a measure γ on $X \times Y$, we have

$$\sup_{\varphi} \int \varphi(x)\,\mathrm{d}\mu - \int \varphi(x)\,\mathrm{d}\gamma = \begin{cases} 0, & \gamma \text{ has first marginal } \mu, \\ +\infty, & \text{otherwise,} \end{cases}$$

and a similar result for the second marginal of γ. We can thus rephrase the Kantorovich problem as

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c\,\mathrm{d}\gamma = \inf_{\gamma} \sup_{(\varphi,\psi)} \int c\,\mathrm{d}\gamma + \int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu - \int \varphi(x) + \psi(y)\,\mathrm{d}\gamma,$$

and then by formally exchanging the inf and sup as

$$\sup_{(\varphi,\psi)} \left[\int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu + \inf_{\gamma} \int c\,\mathrm{d}\gamma - \int \varphi(x) + \psi(y)\,\mathrm{d}\gamma \right].$$

Now, it remains to note that

$$\inf_{\gamma} \int c\,\mathrm{d}\gamma - \int \varphi(x) + \psi(y)\,\mathrm{d}\gamma = \begin{cases} 0, & (\varphi,\psi) \in \Phi_c, \\ -\infty, & \text{otherwise.} \end{cases}$$

□

Theorem 1.10 (Kantorovich-Rubinstein)**.** *Let $X = Y$ be Polish and endowed with a l.s.c. metric d. Let $\mathrm{Lip}_1(X)$ denote the space of all 1-Lipschitz functions on X. Then,*

$$\mathscr{C}_d(\mu,\nu) = \sup\left\{ \int_X \varphi\,\mathrm{d}(\mu-\nu) : \quad \phi \in L^1(\mathrm{d}\mu) \cap L^1(\mathrm{d}\nu) \cap L^1(\mathrm{d}|\mu-\nu|) \cap \mathrm{Lip}_1 \right\}.$$

add conditions for existence of optimizers

[TODO: Add 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 $\mathscr{C}_c(\mu,\nu)$ may be viewed as a measure of distance between μ and ν. When Ω is unbounded, we will restrict our attention to the subset of $\mathcal{P}(\Omega)$ consisting of all probability measures with finite p-th moment:

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

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

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

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** and denoted $\mathbb{W}_p(\Omega)$.

Before proving that W_p is a distance, we first state a few immediate observations:

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 Ω is bounded, $\mathcal{P}_p(\Omega) = \mathcal{P}(\Omega)$ for all $p \geq 1$, and

$$W_1(\mu,\nu) \leq W_p(\mu,\nu) \leq \operatorname{diam}(\Omega)^{\frac{p-1}{p}} W_1(\mu,\nu)^{\frac{1}{p}}.$$

Remark 2.2. In particular, once we establish the status of W_p as a metric, Item (iii) shows that each W_p induce the same topology on $\mathcal{P}(\Omega)$ when Ω is bounded. ⋄

Proof.

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

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

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

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

Higher moment finiteness implies lower moment finiteness.

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

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

(iii) In light of (ii), we need only prove $\mathcal{P}(\Omega) \subset \mathcal{P}_p(\Omega)$ and the second equality. The inclusion is given by $\int |x|^p \, \mathrm{d}\mu \leq \operatorname{diam}(\Omega)^p$. For the inequality, let γ be an optimal transport plan for $W_1(\mu,\nu)$ and note that

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

□

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 2.3. *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. Symmetric 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 $S : \Omega \times \Omega \to \Omega \times \Omega$ is defined by $S(x,y) := (y,x)$. We have

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

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 μ and ν.

Finally, suppose

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

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

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

which shows that $\mu = \nu$. □

Establishing the triangle inequality is more involved. We will deal with the cases $p = 1$ and $p > 1$ separately. The case $p = 1$ follows from the Kantorovich-Rubinstein Theorem quite easily. For $p > 1$, we will leverage the strict convexity of the cost and use Theorem 1.6.

2.1.1. *Triangular Inequality: The case $p = 1$.*

Proposition 2.4. *The function W_1 satisfies the triangle inequality.*

Proof. Starting from the identity

$$\int_\Omega \varphi \, \mathrm{d}(\mu - \nu) = \int_\Omega \varphi \, \mathrm{d}(\mu - \rho) + \int_\Omega \varphi \, \mathrm{d}(\rho - \nu)$$

we have

$$\begin{aligned} W_1(\mu, \nu) &= \sup_{\|\varphi\| \le 1} \int_\Omega \varphi \, \mathrm{d}(\mu - \nu) \\ &\le \sup_{\|\varphi\| \le 1} \int_\Omega \varphi \, \mathrm{d}(\mu - \rho) + \sup_{\|\varphi\| \le 1} \int_\Omega \varphi \, \mathrm{d}(\rho - \nu) \\ &= W_1(\mu, \rho) + W_1(\rho, \nu), \end{aligned}$$

where the equalities follow from the Kantorovich-Rubinstein Theorem (Theorem 1.10). □

We note that in the case where Ω is bounded, the triangle inequality for W_p with $p > 1$ can be deduced from that of W_1 using Proposition 2.1. The general case, however, requires a different argument.

2.1.2. *Triangular Inequality: The case $p \in (1, +\infty)$.* When μ and ρ 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 μ and ρ with absolutely continuous measures obtained by convolution, and pass to the limit with aid from the following lemma:

Lemma 2.5. *Consider $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ and χ_ε an even mollifier in L^1 such that $\int_{\mathbb{R}^d} \chi_\epsilon(x) \, \mathrm{d}x = 1$, $\chi_\epsilon(x) = \epsilon^{-d} \chi_1(x/\epsilon)$, and χ_1 is compactly supported on the unit ball. Write $\mu_\varepsilon := \mu * \chi_\varepsilon$ and $\nu_\varepsilon := \nu * \chi_\varepsilon$. We have*

(i) $W_p(\mu_\varepsilon, \nu_\epsilon) \le W_p(\mu, \nu)$,
(ii) $\lim_{\varepsilon \to 0} W_p(\mu_\epsilon, \nu_\epsilon) = W_p(\mu, \nu)$.

Proof. (i) $\boldsymbol{W_p(\mu_\varepsilon, \nu_\varepsilon) \le W_p(\mu, \nu)}$. Let γ be the optimal transport plan for $W_p(\mu, \nu)$. We use γ to define a candidate transport plan between μ_ϵ and ν_ϵ. The functional

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

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

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

where the second line comes from interchanging the two integrals and a change of variable, and the third line comes from χ_ε being even. This shows that the first marginal of γ_ε is μ_ε. An analogous argument shows that the second marginal is ν_ε. Using this transport plan, we have the bound

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

(ii) $\lim_{\varepsilon\to 0} W_p(\mu_\varepsilon, \nu_\varepsilon) = W_p(\mu, \nu)$. 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 $\varepsilon_n \to 0$. This will be done using the lower semi-continuity of $\mathscr{K}$.

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 $\mu(K_0^\mathrm{c}) < \delta$. Then, enlarge K_0 by 1, the radius of the support of χ_1, to obtain

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

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

$$\mu_\varepsilon(K) = \int \mu(K - x)\chi_\varepsilon(x)\, \mathrm{d}x \leq \int \mu(K_0)\chi_\varepsilon(x)\, \mathrm{d}x = \mu(K_0),$$

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

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

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

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

We now have

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

where the equality comes from how we selected the subsequence ε_{n_k}, the first inequality from Corollary 1.4, and the second inequality from γ being an admissible transport plan. Since $W_p(\mu_\varepsilon, \nu_\varepsilon) \leq W_p(\mu, \nu)$ for each ε, we conclude that $\lim W_p(\mu_{\varepsilon_n}, \nu_{\varepsilon_n}) = W_p(\mu, \nu)$. □

Theorem 2.6. *W_p is a metric on $\mathcal{P}_p(\Omega)$ for each $p \geq 1$.*

Proof. The nonnegativity, symmetry, and positive definiteness of W_p are given by Proposition 2.3. Triangle inequality for $p = 1$ is given by Proposition 2.4. We deal with the case $p \in (1, +\infty)$.

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

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

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

When μ or ρ 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 2.5. The same Lemma then allows us to pass to the limit. □

2.2. **Topology of the Wasserstein Space.** The punchline of this section is a characterization of convergence in the Wasserstein space in terms of weak convergence of measures and convergence of moments.

Let us start by recalling that the weak topology is the coarsest topology that makes $\mu \mapsto \int \varphi \, \mathrm{d}\mu$ continuous for each $\varphi \in C_b$. For an arbitrary class of functions $\mathcal{F}$ we can similarly define a "$\mathcal{F}$-weak topology" to be the coarsest topology that makes $\mu \mapsto \int \varphi \, \mathrm{d}\mu$ continuous for each $\varphi \in \mathcal{F}$. We have the following:

Lemma 2.7. *Let X be a metric space. Convergence in the* Lip*-weak topology implies convergence in the weak topology. In particular, to establish $\mu_n \rightharpoonup \mu$, we need only check that $\int \varphi \, \mathrm{d}\mu_n \to \int \varphi \, \mathrm{d}\mu$ for each each $\varphi \in \mathrm{Lip}(X)$.*

Proof. Let $\psi \in C_b$ be arbitrary and find large M such that $\psi + M \geq 0$. Using Lemma 1.3, find a sequence of nonnegative Lipschitz functions φ_n such that $\varphi_n \nearrow \psi + M$. The functionals

$$F_n : \mu \longmapsto \int \varphi_n \, \mathrm{d}\mu$$

are continuous and in particular l.s.c. in the Lip-weak topology. By the Monotone Convergence Theorem, F_n converges pointwise monotonically to

$$F : \mu \longmapsto \int \psi + M \, \mathrm{d}\mu = M + \int \psi \, \mathrm{d}\mu.$$

As the pointwise supremum of a sequence of continuous functions, F and in particular $F_0 : \mu \mapsto \int \psi \, \mathrm{d}\mu$ is l.s.c. in the Lip-weak topology. Repeating the same argument on $-\psi \in C_b$, we see that F_0 is also u.s.c. Being both l.s.c. and u.s.c., it must be continuous. Since $\psi \in C_b$ is arbitrary, we may conclude. □

We are now ready to characterize convergence in the Wasserstein space. We start with the bounded case:

Theorem 2.8. *If $\Omega \subset \mathbb{R}^d$ is bounded and $p \in [1, +\infty)$, then $W_p(\mu_n, \mu) \to 0$ if and only if $\mu_n \rightharpoonup \mu$.*

Proof. Proposition 2.1 shows that all W_p induce the same topology on $\mathcal{P}(\Omega)$. Thus it suffices to consider the case $p = 1$. Consider first a sequence of measures μ_n converging to μ in $\mathbb{W}_1(\Omega)$. The Kantorovich-Rubinstein Theorem implies that

$$\lim_n \sup \left\{ \int \varphi \, \mathrm{d}(\mu_n - \mu) : \varphi \in \mathrm{Lip}_1 \right\} = 0$$

and, in particular,

$$\int \varphi \, \mathrm{d}\mu_n \longrightarrow \int \varphi \, \mathrm{d}\mu, \quad \forall \varphi \in \mathrm{Lip}_1 .$$

By linearity, the same is true for any Lipschitz function. By Lemma 2.7, we have $\mu_n \rightharpoonup \mu$.

For the opposite implication, fix $\mu_n \rightharpoonup \mu$ and let $\varphi_n \in \mathrm{Lip}_1$ be such that

$$W_1(\mu_n, \mu) = \int \varphi_n \, \mathrm{d}(\mu_n - \mu).$$

Fix an arbitrary $x_0 \in \Omega$. Since the value of the integral above is unchanged when we add a constant to φ_n, the functions φ_n can be chosen such that $\varphi_n(x_0)$ is the same for each n. Using the facts that Ω is bounded and φ_n are 1-Lipschitz, we know that the sequence φ_n are equibounded and equicontinuous. By Arzela-Ascoli, we can pick a subsequence of φ_n that attains $\limsup W_1(\mu_n, \nu)$, and then choose a further subsequence say φ_{n_k} that converges uniformly say to φ. Note that

$$\begin{aligned} 0 \le \liminf W_1(\mu_n, \mu) &\le \limsup W_1(\mu_n, \mu) \\ &= \lim_k W_1(\mu_{n_k}, \mu) = \lim_k \int \varphi_{n_k} \, \mathrm{d}(\mu_{n_k} - \mu). \end{aligned}$$

It is thus sufficient to show that the last limit is 0. For this, we use the following estimate:

$$\begin{aligned} \left| \int \varphi_{n_k} \, \mathrm{d}(\mu_{n_k} - \mu) \right| &\le \left| \int \varphi \, \mathrm{d}(\mu_{n_k} - \mu) \right| + \left| \int \varphi_{n_k} - \varphi \, \mathrm{d}(\mu_{n_k} - \mu) \right| \\ &\le \left| \int \varphi \, \mathrm{d}(\mu_{n_k} - \mu) \right| + \left\| \varphi_{n_k} - \varphi \right\|_\infty \left[\mu_{n_k}(\Omega) + \mu(\Omega) \right] . \end{aligned}$$

As we send $k \to \infty$, the first term converges to 0 by the weak convergence of μ_{n_k}, and the second term converges to 0 by the uniform converges of φ_{n_k}. □

Theorem 2.9. *Let $p \in [1, +\infty)$. We have $W_p(\mu_n, \mu) \to 0$ if and only if $\mu_n \rightharpoonup \mu$ and $\int |x|^p \, \mathrm{d}\mu_n \to \int |x|^p \, \mathrm{d}\mu$.*

Proof. First let μ_n converge to μ in $\mathbb{W}_p(\Omega)$. Proposition 2.1 gives $W_1(\mu_n, \mu) \to 0$. Then the same argument in Theorem 2.8 can be used to show that $\mu_n \rightharpoonup \mu$. The convergence in p-moments follows from noting that

$$\int |x|^p \, \mathrm{d}\mu_n = W_p(\mu_n, \delta_0) \longrightarrow W_p(\mu, \delta_0) = \int |x|^p \, \mathrm{d}\mu.$$

For the opposite direction, fix a sequence $\mu_n \rightharpoonup \mu$ such that $\int |x|^p \, \mathrm{d}\mu_n \to \int |x|^p \, \mathrm{d}\mu$. Denote as K_k the closed ball of radius k centered around 0 and let π_k be the projection map onto K_k. Note that π_k is 1-Lipschitz and in particular continuous.

Let $\mu^k := (\pi_k)_\# \mu$ and define μ_n^k analogously. Note that since $\mu^k, \mu_n^k \in \mathcal{P}(K_k)$, Theorem 2.8 may be used once we check $\mu_n^k \rightharpoonup \mu^k$: for each $\varphi \in C_b(K_k)$ we have

$$\int \varphi \, \mathrm{d}\mu_n^k = \int \varphi \circ \pi_k \, \mathrm{d}\mu_n \longrightarrow \int \varphi \circ \pi_k \, \mathrm{d}\mu = \int \varphi \, \mathrm{d}\mu^k.$$

Thus we have $W(\mu_n^k, \mu^k) \to 0$.

We will use this to bound $W_p(\mu_n, \mu)$ via the estimate

$$W_p(\mu_n, \mu) \le W_p(\mu_n, \mu_n^k) + W_p(\mu_n^k, \mu^k) + W_p(\mu, \mu^k).$$

For fixed k, the discussion above shows that the middle term converges to 0 when $n \to \infty$. We need only bound the first and last term. Consider first the term $W(\mu, \mu^k)$. Using the transport map $x \mapsto \pi_k(x)$, we have

$$W_p(\mu, \mu^k) \le \int_\Omega |\pi_k - \mathrm{Id}|^p \, \mathrm{d}\mu \le \int_{K_k^c} |\mathrm{Id}|^p \, \mathrm{d}\mu \le \int_\Omega |x|^p \, \mathrm{d}\mu - \int_{K_k} |x|^p \, \mathrm{d}\mu.$$

Since $\int |x|^p \, \mathrm{d}\mu < +\infty$, we have by the Monotone Convergence Theorem that the last term converges to 0 when we send $k \to \infty$. In particular, $\lim_{k\to\infty} W_p(\mu, \mu^k) = 0$.

Next, we bound the term $W_p(\mu_n, \mu_n^k)$. Again using $x \mapsto \pi_k(x)$ as a transportation map gives

$$\begin{aligned} W_p(\mu_n, \mu_n^k) \le \int |\pi_k - \mathrm{Id}|^p \, \mathrm{d}\mu_n &\le \int |x - \min\{x, k\}|^p \, \mathrm{d}\mu_n \\ &\le \int |x|^p - |\min\{x, k\}|^p \, \mathrm{d}\mu_n, \end{aligned}$$

where the last inequality follows from the fact that $x \mapsto x^p$ is superadditive on $x \ge 0$. Since $\min\{x, k\}$ is continuous and bounded and $\lim_n \int |x|^p \, \mathrm{d}\mu_n = \int |x|^p \, \mathrm{d}\mu$, the last term converges to $\int |x|^p - |\min\{x, k\}|^p \, \mathrm{d}\mu$ as we send $n \to \infty$. On the other hand, from the Monotone Convergence Theorem we know that $\int |x|^p - |\min\{x, k\}|^p \, \mathrm{d}\mu$ converges to 0 when $k \to \infty$. In particular, for any $\varepsilon > 0$, we can find large k_0 and n_0 such that $W_p(\mu_n, \mu_n^k) < \varepsilon$ whenever $k \ge k_0$ and $n \ge n_0$.

We now put the separate estimates together. First, fix ε and choose large k_0 and n_0 as above. Next, increase k_0 if necessary so that $W(\mu, \mu^{k_0}) < \varepsilon$. We have

$$\begin{aligned} \limsup W_p(\mu_n, \mu) &\le \limsup_n \left[W_p(\mu_n, \mu_n^k) + W_p(\mu_n^k, \mu^k) + W_p(\mu, \mu^k) \right] \\ &\le 2\varepsilon + \limsup_n W_p(\mu_n^k, \mu^k) = 2\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $W_p(\mu_n, \mu) \to 0$. □

3. Curves and Geodesics in the Metric Space

3.1. Curves.

Definition 3.1 (Curve, Speed)**.** A **curve** ω is a continuous function defined on $[0, 1]$ and valued in a metric space (X, d).

- Its **length** is defined as

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

- Its **speed** $|\omega'|$ is defined as

$$|\omega'|(t) := \lim_{h\to 0} \frac{d(\omega(t+h), \omega(t))}{|h|},$$

  if the limit exists.

- The curve ω is said to be **absolutely continuous** if there exists a $g \in L^1$ such that
$$d(\omega(s), \omega(t)) \leq \int_s^t g(r)\, \mathrm{d}r, \quad \forall s, t \in [0,1], s \leq t.$$
The set of all absolutely continuous curves is denoted by $\mathrm{AC}(X)$.

Remark 3.2.

- The Rademacher theorem implies that the speed exists for a.e. t.
- If ω is absolutely continuous, then it has finite length bounded above by $\int_0^1 g(r)\, \mathrm{d}r$.
- For any curve $\omega \in \mathrm{AC}(X)$, we have $\mathrm{Length}(\omega) = \int_0^1 |\omega'|(t)\, \mathrm{d}t$.

$\diamond$

Proposition 3.3. *Absolutely continuous curves can be reparametrized in time to be Lipschitz.*

3.2. **Geodesics.**

Definition 3.4 (Geodesic, Geodesic Space)**.** A curve $\omega : [0,1] \to X$ is a **geodesic** between $x_0, x_1 \in X$ if it minimizes the length among all curves such that $\omega(0) = x_0$ and $\omega(1) = x_1$. It is said to be a **constant speed geodesic** if for each $s, t \in [0,1]$ we have
$$d(\omega(s), \omega(t)) = |s - t| d(\omega(0), \omega(1)).$$
A metric space is a **geodesic space** if for each $x, y \in X$ there exists a geodesic connecting x and y.

Proposition 3.5. *Fix $p > 1$ and consider curves connecting x_0 to x_1. The following are equivalent:*

(i) *ω is a constant speed geodesic,*
(ii) *$\omega \in \mathrm{AC}(X)$ and $|\omega'|(t) = d(x_0, x_1)$ for a.e. t,*
(iii) *ω solves* $\min \left\{ \int_0^1 |\omega'|(t)^p\, \mathrm{d}t : \omega(0) = x_0, \omega(1) = x_1 \right\}$.

3.3. **The Continuity Equation.**

Definition 3.6 (Weak Solutions of the Continuity Equation)**.**

Proposition 3.7. *Equivalence of weak and distributional solutions to the CE.*

4. AC curves in the Wasserstein Space and the Benamou-Brenier Formula

Theorem 4.1. *Let $(\mu_t)_{t\in[0,1]}$ be an absolutely continuous curve in $\mathbb{W}(\Omega)$, where $p > 1$ and $\Omega \subset \mathbb{R}^d$ is compact. For a.e. t, there exists a vector field $v_t \in L^p(\mu_t; \mathbb{R}^d)$ such that the continuity equation $\partial_t \mu_t + \mathrm{div}(v_t \mu_t) = 0$ is satisfied in the weak sense, and for a.e. t we have $\|v_t\|_{L^p(\mu_t)} \leq |\mu'|(t)$.*

Conversely, if $(\mu_t)_{t\in[0,1]}$ is a family of measures in $\mathcal{P}_p(\Omega)$ and for each t we have a vector field $v_t \in L^p(\mu_t; \mathbb{R}^d)$ with $\int_0^1 |v_t|_{L^p(\mu_t)}\, \mathrm{d}t < +\infty$ solving $\partial_t \mu_t + \mathrm{div}(v_t \mu_t) = 0$, then $(\mu_t)_t$ is absolutely continuous in $\mathbb{W}(\Omega)$ and for a.e. t, we have $|\mu'|(t) \leq \|v_t\|_{L^p(\mu_t)}$.

Define $\pi_t : \Omega \times \Omega \to \Omega$ by
$$\pi_t(x, y) := (1 - t)x + ty.$$

Theorem 4.2 ($\mathbb{W}_p(\Omega)$ is a Geodesic Space)**.** *Let Ω be convex and $p \geq 1$. Let $\mu, \nu \in \mathcal{P}_p(\Omega)$ and let γ be the optimal transport plan corresponding to $W_p(\mu, \nu)$. The curve $\mu_t := (\pi_t)_\# \gamma$ is a constant-speed geodesic in $\mathbb{W}_p$ between μ and ν.*

Theorem 4.3 (Benamou-Brenier Formula)**.** *Let $p > 1$ and Ω be convex and compact. For each $\mu, \nu \in \mathcal{P}_p(\Omega)$, we have*

$$W_p^p(\mu, \nu) = \min\left\{\int_0^1 \int_\Omega |v_t|^p \, \mathrm{d}\mu_t \, \mathrm{d}t : \partial_t \mu_t + \mathrm{div}(v_t \mu_t) = 0, \quad \mu_0 = \mu, \quad \mu_1 = \nu\right\}.$$