Math 245A: Introduction to Optimal Transport

Jonathan Woo

Abstract

Optimal transport considers the problem of efficiently moving a quantity from one place to another. Mathematically, we can model this problem using probability measures and cost functions to define optimality. In this final project, we will motivate the related yet subtly different Monge and Kantorovich formulations and examine the existence of solutions under modest constraints. We will conclude with some discussion towards uniqueness of solutions.

Notation

1. A *Borel measure* μ on a locally compact, separable, complete metric space (X, d) is a measure defined on the Borel σ-algebra $\mathcal{B}(X)$ of X. Throughout this project, we will assume metric spaces are locally compact, separable, and complete.

2. A *probability space* $(X, \mathcal{B}_X, \mu)$ is a measure space such that $\mu(X) = 1$. We call μ a *probability measure*.

3. If (X, d) is a locally compact, separable, complete metric space, then $\mathcal{P}(X)$ denotes the set of all Borel probability measures on (X, d).

4. Given sets X and Y, we let $\pi_X : X \times Y \to X$ by $(x, y) \mapsto x$ and $\pi_Y : X \times Y \to Y$ by $(x, y) \mapsto y$ be the *coordinate projections*.

1 The Optimal Transport Problem

How does one most efficiently transport resources from one place to another? What does efficiency mean? If a transport strategy is just a rule of allocating resources, how can we formalize the notion of transport?

1.1 Monge's Formulation

In 1781, Gaspard Monge considered a problem similar to the question: what is the optimal way to transport a volume of earth to a void or given shape? We shall present a measure theoretic formulation of this problem.

First, we must define what it means to transport. Suppose we have a volume of earth located at some points in a metric space (X, d_X) and a shape to be filled at some points in another metric space (Y, d_Y). Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ represent how much volume of earth or void is contained at a given location, respectively. We will assume that the volume of earth and void volume are the equal, that is $\mu(X) = \nu(Y)$. Thus, we can normalize μ and ν to be probability measures on their respective spaces. Then when using some transport strategy to move some volume $\mu(A)$ to some void $\nu(B)$, we would like to preserve volume. More formally, we consider:

Definition 1.1 (Transport Maps)**.** Let (X, d_X) and (Y, d_Y) be locally compact, separable complete metric spaces with $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. We say a measurable map $T : X \to Y$ is a **transport map** from μ to ν if

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

for every $B \in \mathcal{B}(Y)$. We will denote the set of all transport maps from μ to ν by $\mathcal{T}(\mu, \nu)$.

Note that if $T : X \to Y$ is a transport map from μ to ν, then ν is just the pushforward measure of μ by T:

Definition 1.2 (Pushforward measure)**.** Let $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu)$ be measure spaces and let $T : X \to Y$ be a measurable map. Then the **pushforward measure** $T_\# \mu : \mathcal{B}_Y \to [0, +\infty]$ of μ by T is given by

$$T_\# \mu(E) = \mu(T^{-1}(E))$$

for all $E \in \mathcal{B}_Y$. Note that $T_\# \mu$ is indeed a measure because ν is.

Proposition 1.1 (Change of variables[1])**.** Let $(X, \mathcal{B}_X, \mu)$ and $(Y, \mathcal{B}_Y, \nu)$ be measure spaces and let $T : X \to Y$ be a measurable map. Suppose $\nu = T_\# \mu$. If $f : Y \to [0, +\infty]$ is measurable, then

$$\int_Y f(y) d\nu(y) = \int_X (f \circ T)(x) d\mu(x)$$

Proof. If f is simple, then write $f = \sum_{i=1}^n c_i 1_{E_i}$ for some $E_i \in \mathcal{B}_Y$. Then we have

$$\int_Y f(y) d\nu(y) = \sum_{i=1}^n c_i \mu(T^{-1}(E_i)) = \int_X \sum_{i=1}^n c_i 1_{T^{-1}(E_i)}(x) d\mu(x) = \int_X (f \circ T)(x) d\mu(x)$$

since $1_{T^{-1}(E_i)} = 1_{E_i} \circ T$. In the general case, there exists simple $f_n : Y \to [0, +\infty]$ such that $f_n \uparrow f$ pointwise and thus $f_n \circ T \uparrow f \circ T$ pointwise. Two applications of Monotone Convergence theorem combined with the result for simple functions gives

$$\int_Y f(y) d\nu(y) = \lim_{n \to \infty} \int_Y f_n(y) d\nu(y) = \lim_{n \to \infty} \int_X (f_n \circ T)(x) d\mu(x) = \int_X (f \circ T)(x) d\mu(x)$$

as desired. $\square$

It turns out a similar condition characterizes the pushforward measure:

Lemma 1.1 (Characterization of Pushforward Measure)**.** Let (X, d_X) and (Y, d_Y) be locally compact, separable, complete metric spaces with $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Let $T : X \to Y$ be a measurable map. Then $\nu = T_\# \mu$ if and only if for any bounded measurable map $f : Y \to \mathbb{R}$, we have

$$\int_Y f(y) d\nu(y) = \int_X (f \circ T)(x) d\mu(x)$$

[1] Compare to Exercise 1.4.38 from [4]

Proof sketch. If $\nu = T_{\#}\mu$, then we show the other condition by a similar strategy to the proof of Proposition 1.1. The other direction is follows by letting $f = 1_A$ for $A \in \mathcal{B}(Y)$. See Lemma 1.2.5 in [3] for more details. $\square$

Next, to consider the optimality of transporting volumes, we need to specify a (measurable) cost function $c : X \times Y \to [0, +\infty]$. We can think of c as the cost of transporting a unit of volume at $x \in X$ to $y \in Y$. Then the cost of some transport map $T : X \to Y$ is reasonably given by the quantity

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

Monge's formulation to the transport problem is then:

Definition 1.3 (Monge's Formulation)**.** Given locally compact, separable, complete metric spaces (X, d_X) and (Y, d_Y) with $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and a measurable cost function $c : X \times Y \to [0, +\infty]$, **Monge's formulation** is the optimization problem

$$\underset{T \in \mathcal{T}(\mu,\nu)}{\text{minimize}} \int_X c(x, T(x))d\mu(x).$$

When given any optimization problem, it is natural to examine the existence and uniqueness of solutions. To this end, we will consider some example settings for Monge's formulation.

Example 1.1 (Volume Splitting)**.** A traditional issue actually arises in the definition of a transport map $T \in \mathcal{T}(\mu, \nu)$. In particular, given any $\mathcal{B}(Y)$-measurable set B, we require that

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

which is to say that all the volume contained at B (after transporting) must come from $T^{-1}(B)$. This means we are not allowed to split up volumes during the transport process. Aside from being physically restrictive, this creates problems with the existence of solutions to Monge's formulation.

Indeed, suppose all our initial volume of earth is concentrated at a single point, that is $\mu = \delta_{x_0}$ for some $x_0 \in X$, and our target void consists of two distinct holes of equal volume, that is $\nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$ for some $y_1, y_2 \in Y$ such that $y_1 \neq y_2$. We use δ_x to denote the Dirac measure at a point x. Then intuitively speaking, if we want to transport the earth, we must split it up among the two holes. However, we have

$$\nu(\{y_1\}) = \tfrac{1}{2} \notin \{0, 1\} = \mu(\mathcal{B}(X))$$

which implies that no transport map from μ to ν can exist.

Example 1.2 (Non-convexity)**.** Another issue with the Monge formulation is that the constraint set $\mathcal{T}(\mu, \nu)$ is not convex, which is to say that there exists $S, T \in \mathcal{T}(\mu, \nu)$ and $\theta \in (0, 1)$ such that $\theta S + (1 - \theta)T \notin \mathcal{T}(\mu, \nu)$. In general optimization, non-convexity is problematic because we do not get any guarantees for global optimality, existence of minimizers, etc.

To see non-convexity, consider $X = [0, 1]$ and $Y = [0, 1]$ equipped with the Euclidean metric. Let μ and ν be the uniform measure on $[0, 1]$. Then the maps $T_1(x) = x$ and $T_2(x) = 1 - x$ are transport maps from μ to ν, but $\frac{1}{2}T_1(x) + \frac{1}{2}T_2(x) = \frac{1}{2}$, so

$$\nu(\{\tfrac{1}{2}\}) = 0 \neq 1 = \mu([0, 1]) = \mu((\tfrac{1}{2}T_1 + \tfrac{1}{2}T_2)^{-1}(\{\tfrac{1}{2}\})).$$

Thus, $\frac{1}{2}T_1 + \frac{1}{2}T_2 \notin \mathcal{T}(\mu, \nu)$.

1.2 Kantorovich's Formulation

In the 1940s, Leonid Kantorovich studied some variant of the optimal transport problem under more relaxed conditions. Suppose we have N ore mines and M factories such that each ore mine produces a certain number of resources and each factory demands a certain number of resources. If the supply and demand are equal, what is the cheapest (relative to some cost function) way to transport resources to factories such that the demand by the factories is fully met? As with the previous formulation of the problem, we can model our source and destination as Borel probability spaces.

In this problem, any transport strategy is allowed to divide resources from one mine to multiple factories. This addresses the volume splitting issue with transport maps. Note that each transport strategy is entirely determined by how much of a given mine's resources are transported to a given factory. Thus, we want to model each transport strategy as a measure:

Definition 1.4 (Transport plan)**.** Let (X, d_X) and (Y, d_Y) be locally compact, separable, complete metric spaces. We say a probability measure $\gamma \in \mathcal{P}(X \times Y)$ is a **transport plan** from μ to ν if

(i) $\mu = (\pi_X)_\# \gamma$,

(ii) $\nu = (\pi_Y)_\# \gamma$.

We will let $\Gamma(\mu, \nu)$ denote the set of all transport plans from μ to γ.

Remark 1.1. We can informally view a transport plan γ from μ to ν as the normalized quantity of resources transported from a particular source to destination. Using definitions, we have that Definition 1.4(i) is equivalent to

$$\mu(A) = \gamma(\pi_X^{-1}(A)) = \gamma(A \times Y)$$

for all $A \in \mathcal{B}(X)$. Similarly, Definition 1.4(ii) is equivalent to

$$\nu(B) = \gamma(\pi_Y^{-1}(B)) = \gamma(X \times B)$$

for all $B \in \mathcal{B}(Y)$. In particular, Definition 1.4(i) says that all of the resources located at A are transported. Similarly, Definition 1.4(ii) says that all of the demand located at B is satisfied.

Having defined this notion of transporting resources, we now introduce a (measurable) cost function $c : X \times Y \to [0, +\infty]$ as before. Then the cost of some transport plan $\gamma \in \Gamma(\mu, \nu)$ is reasonably given by the quantity

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

Thus, Kantorovich's formulation is:

Definition 1.5 (Kantorovich's Formulation)**.** Given locally compact, separable, complete metric spaces (X, d_X) and (Y, d_Y) with $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and a measurable cost function $c : X \times Y \to [0, +\infty]$, **Kantororvich's formulation** is the optimization problem

$$\underset{\gamma \in \Gamma(\mu,\nu)}{\text{minimize}} \int_{X \times Y} c(x, y) d\gamma(x, y).$$

Example 1.3 (Existence of Transport Plans)**.** Unlike Monge's formulation where transport maps need not exist, Kantorovich's formulation always has at least one transport map, namely the product measure $\mu \times \nu$. Note that since we are in the setting of probability spaces, which are σ-finite, we have that the product measure is unique.

Indeed, for all $A \in \mathcal{B}(X)$ and all $B \in \mathcal{B}(Y)$, we have

$$(\pi_X)_\#(\mu \times \nu)(A) = (\mu \times \nu)(\pi_X^{-1}(A)) = (\mu \times \nu)(A \times Y) = \mu(A)\nu(Y) = \mu(A)$$

and

$$(\pi_Y)_\#(\mu \times \nu)(B) = (\mu \times \nu)(\pi_Y^{-1}(B)) = (\mu \times \nu)(X \times B) = \mu(X)\nu(B) = \nu(B).$$

Example 1.4 (Convexity)**.** If $\gamma_1, \gamma_2 \in \Gamma(\mu, \nu)$, then one can show that $\theta\gamma_1 + (1-\theta)\gamma_2 \in \Gamma(\mu, \nu)$ for all $\theta \in [0, 1]$. Thus, Kantorovich's formulation is fundamentally a convex optimization problem. Thus, we can utilize the notion of duality.

1.3 Relating the two formulations

Because of the similarities of the two formulations, one might ask if we can move between in the two problems. Moreover, since Kantorovich's formulation relaxes the conditions of Monge's formulation, we can also ask if the former generalizes the latter.

Proposition 1.2. Let (X, d_X) and $Y, d_Y)$ be locally compact, separable, complete metric spaces with $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. If $c : X \times Y \to [0, +\infty]$ is measurable, then

$$\inf_{T \in \mathcal{T}(\mu,\nu)} \int_{X \times Y} c(x, T(x)) d\mu(x) \geq \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c(x, y) d\gamma(x, y)$$

Proof. If $\mathcal{T}(\mu, \nu) = \varnothing$, then we have the result since $\inf \varnothing = +\infty$. If $\mathcal{T}(\mu, \nu) \neq \varnothing$, then for any $T \in \mathcal{T}(\mu, \nu)$, we have that

$$\gamma_T \coloneqq (1_X \times T)_\# \mu \in \mathcal{P}(X \times Y).$$

Moreover,

$$(\pi_X)_\#(\gamma_T) = (\pi_X)_\#(1_X \times T)_\# \mu = (\pi_X \circ (1_X \times T))_\# \mu = (1_X)_\# \mu = \mu$$

and (similarly)

$$(\pi_Y)\#(\gamma_T) = \nu,$$

Thus, $\gamma_T \in \Gamma(\mu, \nu)$. Then we have by Proposition 1.1

$$\int_X c(x, T(x) d\mu(x) = \int_X (c \circ \gamma_T)(x) d\mu(x) = \int_{X \times Y} c(x, y) d\gamma_T(x, y) \geq \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c(x, y) d\gamma(x, y)$$

as desired. □

2 Existence of Optimal Transport Plans

We will now see that under modest constraints to the cost function, we are guaranteed to have an optimal transport plan and thus a solution to Kantorovich's formulation.

More specifically, we will assume that the cost function is continuous. Then, the strategy for demonstrating the existence of an optimal transport map will be to extract a convergent subsequence from a sequence of transport maps whose costs tend to the optimal value. The limit of this convergent subsequence will be our optimal transport map. To begin this argument, we need to define the type of convergence we will be using:

Definition 2.1 (Weak convergence)**.** Let (X, d) be a locally compact, separable, complete metric space. Let $\mu, \mu_k \in \mathcal{P}(X)$, for $k \geq 1$. We say that μ_k **converges weakly** to μ, write $\mu_k \rightharpoonup \mu$, if

$$\int_X f(x) d\mu_k(x) \to \int_X f(x) d\mu(x)$$

as $k \to \infty$ for all continuous, bounded functions f on X.

To extract the convergent subsequence, we will use a sequential compactness argument. In particular, we will use the following result due to Prokhorov:

Theorem 2.1 (Prokhorov's Theorem)**.** Let (X, d) be a separable, complete metric space with $(X, \mathcal{B}_X)$ being a measurable space. Suppose $A \subset \mathcal{P}(X)$ is tight: for every $\varepsilon > 0$, there exists compact $K_\varepsilon \subset X$ such that $\mu(X \smallsetminus K_\varepsilon) \leq \varepsilon$ for all $\mu \in A$. Then, A is sequentially compact with respect to weak convergence: for any sequence $(\mu_k)_{k \in \mathbb{N}} \subset A$, there exists a subsequence $(\mu_{k_j})_{j \in \mathbb{N}}$ such that $\mu_{k_j} \rightharpoonup \mu$ as $j \to \infty$ for some $\mu \in \mathcal{P}(X)$.

Remark 2.1 (Prokhorov's Theorem)**.** See Theorem 2.1.11 in [3] for a proof of Theorem 2.1. The result presented can actually be upgraded to "$A \subset \mathcal{P}(X)$ is tight if and only if A is sequentially compact with respect to weak convergence", but we only need the result as stated.

The condition of tightness intuitively means that a set of probability measures does not have mass which escapes to infinity. This allows us to extract convergent subsequences whose limits are probability measures.

An alternative way to extract convergent subsequences is to attach a metric π to $\mathcal{P}(X \times Y)$ which makes $\Gamma(\mu, \nu)$ sequentially compact with respect to weak convergence. Indeed, we can define the so-called Prokhorov metric $\pi : \mathcal{P}(X \times Y)^2 \to \mathbb{R}$ by

$$\pi(\alpha, \beta) := \inf\{\varepsilon > 0 : \alpha(A) \leq \beta(A^\varepsilon) + \varepsilon; \beta(A) \leq \alpha(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B}(X \times Y)\}$$

where $A^\varepsilon = \{z \in X \times Y : d_{X \times Y}(z, t) < \varepsilon \text{ for some } t \in A\}$. Theorem 6.8 from [1] shows that if $X \times Y$ is separable and complete, then weak convergence is equivalent to convergence in the Prokhorov metric. Thus a sequentially compact set in $\mathcal{P}(X \times Y)$ is sequentially compact with respect to weak convergence.

To use Theorem 2.1, we need to show that the tightness condition holds. The following lemma will be useful:

Lemma 2.1 (Tightness of a single probability measure)**.** Let (X, d) be a locally compact, separable, complete metric space. If $\mu \in \mathcal{P}(X)$, then there exists a compact set $K_\varepsilon \subset X$ such that $\mu(X \smallsetminus K_\varepsilon) \leq \varepsilon$.

Proof. Let $\varepsilon > 0$. Since X is separable, there exists a countable dense subset $(x_n)_{n \in \mathbb{N}} \subset X$. In particular, for any $k \geq 1$, we have that

$$X = \bigcup_{n=1}^{\infty} \bar{B}(x_n, \tfrac{1}{k}),$$

where $\bar{B}(x,r) = \{y \in X : d(x,y) \leq r\}$. By upward monotone convergence, we have that

$$\lim_{N\to\infty} \mu\left(\bigcup_{n=1}^{N} \bar{B}(x_n, \tfrac{1}{k})\right) = 1 = \mu(X),$$

so for each $k \geq 1$, there exists N_k such that

$$\mu\left(X \setminus \bigcup_{n=1}^{N_k} \bar{B}(x_n, \tfrac{1}{k})\right) = 1 - \mu\left(\bigcup_{n=1}^{N_k} \bar{B}(x_n, \tfrac{1}{k})\right) \leq \varepsilon \cdot 2^{-k}.$$

Define

$$K_\varepsilon \coloneqq \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{N_k} \bar{B}(x_n, \tfrac{1}{k}).$$

Then we have that

$$\mu(X \setminus K_\varepsilon) = \mu\left(\bigcup_{k=1}^{\infty}\left(X \setminus \bigcup_{n=1}^{N_k} \bar{B}(x_n, \tfrac{1}{k})\right)\right) \leq \sum_{k=1}^{\infty} \varepsilon \cdot 2^{-k} = \varepsilon$$

To finish, we show that K_ε is compact. Note that K_ε is closed as it is the intersection of closed sets. Then since X is complete, K_ε is complete. Moreover, for any $\delta > 0$, there exists $k \geq 1$ such that $\frac{1}{k} < \delta$ and thus

$$K_\varepsilon \subset \bigcup_{n=1}^{N_k} \bar{B}(x_n, \tfrac{1}{k}) \subset \bigcup_{n=1}^{N_k} B(x_n, \delta).$$

In particular, K_ε is totally bounded. Hence, K_ε is compact. $\square$

Once we have our weakly convergent subsequence, we will show the optimality of the limit using a result that should remind the reader of Fatou's Lemma and its proof:

Lemma 2.2 (Fatou-like Lemma with Weakly-Converging Measures)**.** Let (X, d) be a locally compact, separable, complete metric space and suppose $f : X \to [0, +\infty]$ is continuous. If $\mu \in \mathcal{P}(X)$ and $\mu_k \in \mathcal{P}(X)$, for $k \geq 1$, such that $\mu_k \rightharpoonup \mu$, then

$$\int_X f(x)d\mu(x) \leq \liminf_{k\to\infty} \int_X f(x)d\mu_k(x)$$

Proof. Letting $f_n \coloneqq \min(f, n)$ gives us a sequence of continuous bounded functions such that $f_n \uparrow f$ pointwise. Then by definition of weak convergence,

$$\int_X f_n(x)d\mu_k(x) \to \int_X f_n(x)d\mu(x)$$

as $k \to \infty$, for any $n \geq 1$. Then by monotonicity, we have

$$\int_X f_n(x)d\mu_k(x) \leq \int_X f(x)d\mu_k(x)$$

for all $k \geq 1$. Thus

$$\int_X f_n(x) d\mu(x) = \liminf_{k \to \infty} \int_X f_n(x) d\mu_k(x) \leq \liminf_{k \to \infty} \int_X f(x) d\mu_k(x)$$

Then by Monotone Convergence Theorem, we have

$$\int_X f_n d\mu(x) \to \int_X f(x) d\mu(x),$$

as $n \to \infty$, so we have the result. $\square$

We are now ready to formally prove the existence of an optimal transport plan:

Theorem 2.2 (Existence of Optimal Transport Plan)**.** Let (X, d_X) and (Y, d_Y) be locally compact, separable, complete metric spaces. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. If $c : X \times Y \to [0, +\infty]$ is a continuous cost function. Then, there exists $\gamma_\star \in \Gamma(\mu, \nu)$ such that

$$\int_{X \times Y} c(x, y) d\gamma_\star(x, y) = \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y).$$

In particular, there exists a solution to Kantorovich's formulation.

Proof. If $\inf_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) = +\infty$, then $\gamma_\star = \mu \times \nu$ works. If

$$\inf_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) < +\infty,$$

then we can get a sequence $\gamma_k \in \Gamma(\mu, \nu)$, $k \geq 1$, such that

$$\lim_{k \to \infty} \int_{X \times Y} c(x, y) d\gamma_k(x, y) = \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y)$$

For example, we can choose $\gamma_k \in \Gamma(\mu, \nu)$ such that

$$\int_{X \times Y} c(x, y) d\gamma_k(x, y) \leq \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) \right) + 2^{-k}$$

Then we need to extract a weakly converging subsequence using Theorem 2.1. In particular, we show that $\Gamma(\mu, \nu)$ is tight. By Lemma 2.1, for each $\varepsilon > 0$, there exists a compact set $K_1 \subset X$ and a compact set $K_2 \subset Y$ such that

$$\mu(X \setminus K_1) \leq \tfrac{\varepsilon}{2} \quad \text{and} \quad \nu(Y \setminus K_2) \leq \tfrac{\varepsilon}{2}$$

Then $K_\varepsilon = K_1 \times K_2 \subset X \times Y$ is compact and

$$(X \times Y) \setminus K_\varepsilon = (X \setminus K_1) \times Y \cup X \times (Y \setminus K_2)$$

So, for any $\gamma \in \Gamma(\mu, \nu)$, we have

$$\begin{aligned} \gamma((X \times Y) \setminus K_\varepsilon) &\leq \gamma((X \setminus K_1) \times Y) + \gamma(X \times (Y \setminus K_2)) \\ &= \mu(X \setminus K_1) + \nu(Y \setminus K_2) \\ &\leq \varepsilon \end{aligned}$$

Thus, by Theorem 2.1, there exists a subsequence γ_{k_j} of γ_k such that $\gamma_{k_j} \rightharpoonup \gamma_\star$ as $j \to \infty$, for some $\gamma_\star \in \mathcal{P}(X \times Y)$. Then, Lemma 2.2 gives us

$$\begin{aligned}\int_{X\times Y} c(x,y)d\gamma_\star(x,y) &\leq \liminf_{j\to\infty} \int_{X\times Y} c(x,y)d\gamma_{k_j}(x,y) \\ &= \lim_{j\to\infty} \int_{X\times Y} c(x,y)d\gamma_{k_j}(x,y) \\ &= \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{X\times Y} c(x,y)d\gamma(x,y)\end{aligned}$$

To finish, it suffices to show that $\gamma_\star \in \Gamma(\mu,\nu)$. Indeed, for all continuous bounded functions f on X, we have by Proposition 1.1 and the definition of weak convergence:

$$\begin{aligned}\int_X f(x)d\mu(x) &= \int_{X\times Y} f(\pi_X(x,y))d\gamma_k(x,y) \\ &= \int_{X\times Y} f(x)d\gamma_k(x,y) \\ &\to \int_{X\times Y} f(x)d\gamma_\star(x,y)\end{aligned}$$

as $k \to \infty$. By Lemma 1.1, we have that $(\pi_X)_\#\gamma_\star = \mu$. Similarly, we have that $(\pi_Y)_\#\gamma_\star = \nu$. Hence $\gamma_\star \in \Gamma(\mu,\nu)$. □

Theorem 2.2 gives us some optimal transport plan $\gamma \in \Gamma(\mu,\nu)$. If there exists a measurable map $T : X \to Y$ such that $\gamma = (1_X \times T)\#\mu$, then it is straightforward to see that $T \in \mathcal{T}(\mu,\nu)$. Combining this fact with the proof of Proposition 1.2 shows that T is an optimal solution to Monge's formulation.

3 Towards Uniqueness

Recall that Kantorovich's formulation is a convex optimization problem. Then one can show a powerful duality result:

Theorem 3.1 (Kantorovich's Duality Theorem)**.** In the context of Kantorovich's formulation, let $c : X \times Y \to [0, +\infty]$ be continuous and bounded from below. Assuming the optimal value is finite, that is

$$\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{X\times Y} c(x,y)d\gamma(x,y) < +\infty,$$

we have

$$\inf_{\gamma\in\Gamma(\mu,\nu)} \int_{X\times Y} c(x,y)d\gamma(x,y) = \max\left\{\int_X -f(x)d\mu(x) + \int_Y -g(y)d\nu(y) : f(x)+g(y)+c(x,y) \geq 0\right\}$$

Using Theorem 2.2 and the structure of minimizers from Theorem 3.1, one can guarantee existence and uniqueness of optimal transport plans for quadratic costs:

Theorem 3.2 (Brenier's Theorem)**.** Let $X = \mathbb{R}^d$ and $Y = \mathbb{R}^d$ equipped with the Euclidean metric and $c(x, y) = \frac{1}{2}|x - y|^2$. Suppose

$$\int_X |x|^2 d\mu(x) < +\infty \text{ and } \int_Y |y|^2 d\nu(y) < +\infty$$

and that μ is absolutely continuous with respect to the Lebesgue measure. Then there exists a unique optimal transport plan $\gamma_\star \in \Gamma(\mu, \nu)$. Moreover, $\gamma_\star = (1_X \times T)_\# \mu$ and $T = \nabla \phi$ for some convex function ϕ.

4 Acknowledgments

The project idea was inspired by Alessio Figalli's lecture series (see [2]). Section 2 was modeled after Chapter 2 from [3].

References

[1] Patrick Billingsley. *Convergence of Probability Measures.* 1999.

[2] Alessio Figalli. *An Introduction to Optimal Transport and Wasserstein Gradient Flows.* 2022.

[3] Alessio Figalli and Federico Glaudo. *An Invintation to Optimal Transport, Wasserstein Distances, and Gradient Flows.* 2021.

[4] Terence Tao. *An Introduction to Measure Theory.* 2011.