MATH 8803: Optimal Transport: Theory and Applications

Rui Gong

September 12, 2025

Acknowledgements

Contents

1 Introduction

1.1 Monge's Original Formulation of Optimal Transport

Consider a measure μ, another measure ν, and x, y in the supports of μ and ν respectively, *what is the optimal way of moving x to y?*

1. Pile and hole

   - Pile and hole should have the same volume $\rightarrow$ normalize to 1.

   - modern way of thinking about pile and hole: probability measure on some metric space X and Y respectively, $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$. It could be point cloud: $\mu = \sum_i \alpha_i \delta_{x_i}$; or continuous densities: $\mu(dx) = f(x)dx$.

2. Transport pile to hole region: Transport described by a map $T : X \rightarrow Y$. Notice that T may be discontinuous. We need T to be measurable.

3. Transport Cost: $c : X \times Y \rightarrow [0, \infty) \cup \{\infty\}$, where $c(x, y)$ represents the cost of moving one unit of mass from x to y (how it is transported does not matter). Implicit Assumption: cost only depends on initial and final. Typical cost: $c(x, y) = |x - y|$; $c(x, y) = |x - y|^2$; $c(x, y) = \frac{1}{|x-y|}$.

4. Filling the hole completely: $\mu(T^{-1}(B)) = \nu(B)$ for every measurable $B \subseteq Y$.

> **Definition 1.1: Push-Forward Measure**
>
> *Let $\mu \in \mathcal{X}$ be a probability measure on X, $T : X \rightarrow Y$ be a measurable map between metric space X, Y. Then* push-forward (or image) measure *of μ under T is the measure $T_\# \mu$ on Y defined by $\mu(T^{-1}(B)) = T_\# \mu(B)$ for every $B \subseteq Y$ measurable.*

A little bit of functional analysis

- $C_b(X)$: the Banach space of bounded continuous function on X endowed with the norm $\|f\|_\infty := \sup_{x \in X} |f(x)|$.

- $C_0(X) \subseteq C_b(X)$: closed subspace (w.r.t. $\|\cdot\|_\infty$), which is the space of continuous functions vanishing at ∞: $f \in C_0(X)$ if $f \in C_b(X)$ and for every $\epsilon > 0$, there exists a compact set $K_\epsilon \subseteq X$, such that $|f| < \epsilon$ on $X \setminus K_\epsilon$.

- $\mathcal{M}(X)$: space of finite signed measures on X. $\lambda \in \mathcal{M}(X)$ if

   (a) $\lambda(A) \in \mathbb{R}$ for any (Borel) measurable $A \subseteq X$.

   (b) for every countable disjoint union $A = \cup_{i \in \mathbb{N}} A_i$, $A_i \cap A_j = \emptyset$ for $i \neq j$, these holds
      - $\sum_{i \in \mathbb{N}} |\lambda(A_i)| < \infty$
      - $\sum_{i \in \mathbb{N}} \lambda(A_i) = \lambda(A)$.

   To every $\lambda \in \mathcal{M}(X)$, we can associate a unique non-negatable measure $|\lambda| \in \mathcal{M}_+(X)$ via $|\lambda|(A) := \sup\{\sum_{i \in \mathbb{N}} |\lambda(A_i)|$: $A = \cup_{i \in \mathbb{N}} A_i, A_i \cap A_j = \emptyset$ for $i \neq j\}$, the total variation measure of λ. $\|\lambda\| := |\lambda|(X)$ is a norm on $\mathcal{M}(X)$.

> **Theorem 1.2: Riesz Representation Theorem**
>
> *Suppose X is separable, and locally compact. Then $\mathcal{M}(X) \cong [C_0(X)]^*$ (the dual space of $C_0(X)$). That is, every continuous linear functional $L : C_0(X) \rightarrow \mathbb{R}$ is represented in a unique way by an element of $\mathcal{M}(X)$, i.e., there exists a unique measure $\mu_L \in \mathcal{M}(X)$ s.t. $L(\varphi) = \int_X \varphi d\mu_L$.*

Remark. Consider a special case of $T_\# \mu = \nu$; assume that T is a C^1-diffeomorphism between X, Y and $X, Y \subseteq \mathbb{R}^d$ open, and that $\mu(dx) = f(x)dx, \nu(dy) = g(y)dy$. Then for any $B \subseteq \mathbb{R}^d$ measurable:

$$(T_\# \mu)(B) = \mu(T^{-1}(B)) = \int_{\mathbb{R}^d} \mathbb{1}_{T^{-1}(B)}(x) f(x) dx = \int_{T^{-1}(B)} f(x) dx.$$

Write $y = T(x), dy = |\det DT(x)|dx$, we can write

$$(T_{\#}\mu)(B) = \nu(B) = \int_B g(y)dy = \int_{T^{-1}(B)} g(T(x))|\det DT(x)|dx,$$

which implies $f(x) = g(T(x))|\det DT(x)|$ for almost every $x \in X$ (technically remark: $f \geq \alpha$ for some $\alpha > 0$ on X).

Reminder: $T_{\#}\mu = \nu$ means:

1. $(T_{\#}\mu)(B) = \mu(T^{-1}(B)) = \nu(B)$ for any measurable subset $B \subseteq Y$.

2. $\int_y \varphi d(T_{\#}\mu) = \int_X \varphi \circ T d\mu = \int_y \varphi d\nu, \ \forall \varphi \in C_0(y)$.

A quick remark on the change of variables formula:

$$\int_y \varphi d(T_{\#}\mu) = \int_X \varphi(T(x))\mu(dx).$$

Definition 1.3: Monge's Optimal Tranposrt Problem

Given $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$,

$$\min \ I[T] = \int_X c(x, T(x))\mu(dx) \tag{$\mathcal{M}$}$$

over all transport maps $T : X \to Y$ (i.e., all measurable maps from X to Y such that $T_{\#}\mu = \nu$).

Remark.   • I is a highly nonlinear functional of T subject to the nonlinear constraint $T_{\#}\mu = \nu$.

- Functional relatively simple: depends only locally on T (or its pointwise values)

  – no coupling between different values of T.

  – without constraint could just minimize pointwise, i.e. find minimum $y_{\min}(x)$ of $y \mapsto c(x, y)$ for each x and get $T(x) = y_{\min}(x)$.

- constraint complicated: nonlocal, couples values of T. If we could restrict to smooth diffeomorphisms, problem requires solving highly nonlinear PDE. Also, it is not even clear whether T such that $T_{\#}\mu = \nu$ exists for given μ, ν.

1.2   The Kantorovich Optimal Transport Problem

Comparing to Monge's OT problem, Kantorovich OT problem allows measure splitting, so we are looking for probability measure on $X \times Y$.

- <u>Pile and hole:</u> $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$.

- <u>Transport:</u> probability measure $\gamma \in \mathcal{P}(X \times Y)$ (transport plan).

$$\gamma(A \times B) = \int_{A \times B} \gamma(dxdy)$$

  is the amount of mass moved from measurable $A \subseteq X$ to measurable $B \subseteq Y$. All the mass of μ has to be transported somewhere, hence, $\gamma(A \times Y) = \mu(A)$ for all $A \subseteq X$ measurable.

- <u>Transport cost:</u> Let $c(x, y)$ be the cost of moving one unit of mass from x to y, then the total cost is

$$\int_{X \times Y} c(x, y)\gamma(dxdy) = c[\gamma].$$

- <u>Filling hole completely</u>: $\gamma(X \times B) = \nu(B)$ for all $B \subseteq Y$ measurable. That is, the amount of mass transported to B has to be the volume of the hole in region B.

Remark. Note that, from above, $\gamma(X \times Y) = \mu(X) = \nu(Y) = 1$, so $\gamma \in \mathcal{P}(X \times Y)$, and $\gamma(A \times Y), \gamma(X \times B)$ defined marginals.

Definition 1.4: Marginals

Let $\gamma \in \mathcal{P}(X \times Y)$.

- *Marginal w.r.t. X: $M_X\gamma \in \mathcal{P}(X)$ defined via*

$$(M_X\gamma)(A) = \gamma(A \times Y) = \int_{A \times Y} \gamma(dxdy), \ \forall \ measurable \ A \subseteq X,$$

- *Marginal w.r.t. Y: $M_Y\gamma \in \mathcal{P}(Y)$ defined via*

$$(M_Y\gamma)(B) = \gamma(X \times B) = \int_{X \times B} \gamma(dxdy), \ \forall \ measurable \ B \subseteq Y.$$

Remark. Transport plans are probability measures on $X \times Y$ with marginals $M_X\gamma = \mu$, $M_Y\gamma = \nu$, γ is a *coupling* of the probability measure μ and ν.

Let $\Pi(\mu, \nu)$ be the set of all couplings between μ and ν.

Lemma 1.5

Let $\varphi \in L^1(X, \mu)$ and $\psi \in L^1(Y, \nu)$. Then for any coupling $\gamma \in \Pi(\mu, \nu)$. These hold

(M1) $\int_{X \times Y} \varphi(x)\gamma(dxdy) = \int_X \varphi(x)(M_X\gamma)(dx) = \int_X \varphi(x)\mu(dx)$.

(M2) $\int_{X \times Y} \psi(y)\gamma(dxdy) = \int_Y \psi(y)(M_Y\gamma)(dy) = \int_Y \psi(y)\nu(dy)$.

Proof sketch. Any function $\varphi \in L^1(X, \mu)$ can be approximated by simple functions: $\varphi = \lim_{n \to \infty} \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}$, for $A_j \subseteq X$ measurable.

$$\int_{X \times Y} \varphi(x)\gamma(dxdy) = \lim_{n \to \infty} \sum_{j=1}^n \alpha_j \int_{A_j \times Y} \gamma(dxdy) = \lim_{n \to \infty} \sum_{j=1}^n \alpha_j \mu(A_j)$$

$$= \lim_{n \to \infty} \int_X \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}(x)\mu(dx) = \int_X \varphi(x)\mu(dx).$$

$\square$

Definition 1.6: Couplings

Let $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. A probability measure $\gamma \in \mathcal{P}(X \times Y)$ is called coupling of μ and ν if $M_X\gamma = \mu$, $M_Y\gamma = \nu$. The set of all couplings between μ and ν is called $\Pi(\mu, \nu)$.

Definition 1.7: Kantorovich Optimal Transport Problem

Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$,

$$\min \ C[\gamma] = \int_{X \times Y} c(x, y)\gamma(dxdy) \tag{$\mathcal{K}$}$$

over all couplings $\gamma \in \Pi(\mu, \nu)$.

Structure of ($\mathcal{K}$):

(1) $\gamma \mapsto C[\gamma]$ linear function.

(2) $M_X \gamma = \mu$, $M_Y \gamma = \nu$ linear constraints.

(3) $\Pi(\mu, \nu)$ is a convex set: if $\gamma_1, \gamma_2 \in \Pi(\mu, \nu)$, $\lambda \in (0,1)$, then

    (a) $\lambda\gamma_1 + (1-\lambda)\gamma_2 \in \mathcal{P}(X \times Y)$, since $\lambda\gamma_1(X \times Y) + (1-\lambda)\gamma_2(X \times Y) = 1$, and $\lambda\gamma_1(Z) + (1-\lambda)\gamma_2(Z) \geq 0$ for any $Z \subseteq X \times Y$ measurable.

    (b) $(\lambda\gamma_1 + (1-\lambda)\gamma_2)(A \times Y) = \lambda\gamma_1(A \times Y) + (1-\lambda)\gamma_2(A \times Y) = \lambda\mu(A) + (1-\lambda)\mu(A) = \mu(A)$; analogously, $(\lambda\gamma_1 + (1-\lambda)\gamma_2)(X \times B) = \nu(B)$, for any $A \subseteq X, B \subseteq Y$ measurable.

    $\implies$ ($\mathcal{K}$) is a linear programming problem. *But*: it is in inifinte dimensions.

(4) Existence of couplings is trivial: independent coupling (product measure) $\gamma = \mu \otimes \nu$ defined via $\gamma(A \times B) = \mu(A)\nu(B)$ for every measurable $A \subseteq X, B \subseteq Y$, is a coupling of μ and ν.

(5) ($\mathcal{K}$) is higher-dimensional than ($\mathcal{M}$) in the following sense: consider transport plan given by a density $\tilde{\gamma} : \mathbb{R}^{2d} \to \mathbb{R}$. Discretize $\mathbb{R}^d$ by ℓ gridpoints then $\tilde{\gamma}$ corresponds to ℓ^2 real numbers. Transport $T : \mathbb{R}^d \to \mathbb{R}^d$ however only corresponds to ℓd real numbers. e.g. ℓ = number of pixels in a 2D picture, say $\ell = 500 \times 500$, then $\ell d = 500000$, but $\ell^2 = 62.5^9$.

1.3   Monge VS Kantorovich

Kantorovish problem is a relaxation of Monge Problem in the following sense: Monge = restriction of Kantorovish to sparse plan.

$$\gamma_T(dxdy) = \delta_{T(x)}(dy)\mu(dx)$$

for $T : X \to Y$ measurable such that $T_{\#}\mu = \nu$. For any $\varphi \in C_0(X \times Y)$,

$$\int_{X \times Y} \varphi d\gamma_T = \int_{X \times Y} \varphi(x,y)\gamma_T(dx \times dy) = \int_{X \times Y} \varphi(x,y)\delta_{T(x)}(dy)\mu(dx) = \int_X \varphi(x, T(x))\mu(dx)$$

$$= \int_X \varphi((id,T)(x))\mu(dx) = \int_X \varphi \circ (id,t)d\mu = \int_{X \times Y} \varphi(x,y)[(id,T)_{\#}\mu](dx \times dy).$$

In other words, Monge measures that

$$\operatorname{supp} \gamma = \{(x,y) \in X \times Y : \gamma(B_\epsilon(x,y)) > 0 \text{ for every } \epsilon > 0\}$$
$$= \cap\{Z \subseteq X \times Y \text{ closed } : \gamma(Z) = 1\}$$
$$\subseteq \operatorname{graph} T = \{(x, T(x)) \in X \times Y : x \in X\}$$

Lemma 1.8

Let $T : X \to Y$ be measurable such that $T_{\#}\mu = \nu$. Then

    (i) $C[\gamma_T] = I[T]$.

    (ii) $\gamma_T \in \Pi(\mu, \nu)$

    (iii) ($\mathcal{K}$) $\leq$ ($\mathcal{M}$).

Proof.   (i) $C[\gamma_T] = \int_{X \times Y} c(x,y)\gamma_T(dx \times dy) = \int_{X \times Y} c(x,y)((id,T)_{\#}\mu)(dxdy) = \int_X c(x,T(x))\mu(dx) = I[T]$.

(ii) Let $A \subseteq X$, $B \subseteq Y$ measurable. Then

$$(M_X \gamma_T)(A) = \gamma_T(A \times Y) = \int_{A \times Y} d\gamma_T = \int_X \mathbb{1}_{A \times Y}(x, y) \underbrace{\gamma_T(dx \times dy)}_{((id, T)_\# \mu)(dxdy)}$$

$$= \int_X \underbrace{\mathbb{1}_{A \times Y}(x, T(x))}_{\mathbb{1}_A(x)} \mu(dx)$$

$$= \int_A d\mu = \mu(A)$$

$$(M_Y \gamma_T)(B) = \int_X \underbrace{\mathbb{1}_{X \times B}(x, T(x))}_{\mathbb{1}_B(x)} \mu(dx) = \int_Y \mathbb{1}_B(y)(T_\# \mu)(dy)$$

$$= \int_A d\nu = \mu(A)$$

(iii) By (i) and (ii),

$$(\mathcal{M}) = \inf_{T : X \to Y \text{ measurable}, T_\# \mu = \nu} I[T] = \inf_{T : X \to Y \text{ measurable}, T_\# \mu = \nu} C[\gamma_T] \geq \inf_{\gamma \in \Pi(\mu, \nu)} C[\gamma] = (\mathcal{K}),$$

by $\gamma_T \in \Pi(\mu, \nu)$.

$\square$

1.4 Basic questions and examples

1. <u>Existence</u> Do optimal plans/maps exist? For optimal plans, yes, under reasonable assumptions on c, inf in $(\mathcal{K})$ is min. For optimal maps:

   Example 1.1. Let $X = Y = \mathbb{R}^d$, $\mu = \delta_a, a \in \mathbb{R}^d$, $\nu = \frac{1}{2}(\delta_b + \delta_C)$ for $b \neq c \in \mathbb{R}^d$. Let $T : \mathbb{R}^d \to \mathbb{R}^D$ be measurable for all $A \subseteq \mathbb{R}^d$ open.

   $$(T_\# \delta_a)(A) = \delta_a(T^1(A)) = \begin{cases} 1 & \text{if } a \in T^{-1}(A) \\ 0 & \text{otherwise} \end{cases} = \delta_{T(a)}(A).$$

   i.e. $T_\# \delta_a = \delta_{T(a)}$, so I cannot map the mass to two different points so $T_\# \mu = \nu$ is only possible if $b = c$.

   The class of all transport plans $\Pi(\mu, \nu)$ consist of a single measure

   $$\Pi(\mu, \nu) = \left\{ \mu \otimes \nu = \frac{1}{2}(\delta_a \otimes \delta_b + \delta_a \otimes \delta_c) \right\}.$$

   Indeed, note that

   $$\gamma(\underbrace{\mathbb{R}^{2d} \setminus \{(a, b), (a, c)\}}_{=(\mathbb{R}^d \setminus \{a\} \times \mathbb{R}^d) \cup (\mathbb{R}^d \times \mathbb{R}^d \setminus \{b, c\})})$$

   $$\leq \gamma\left((\mathbb{R}^d \setminus \{a\} \times \mathbb{R}^d)\right) + \gamma\left((\mathbb{R}^d \times \mathbb{R}^d \setminus \{b, c\})\right) = \mu(\mathbb{R}^d \setminus \{a\}) + \nu(\mathbb{R}^d \setminus \{b, c\})$$

   $$= \delta_a(\mathbb{R}^d \setminus \{a\}) + \frac{1}{2}(\delta_b + \delta_c)(\mathbb{R}^d \setminus \{b, c\}) = 0.$$

   For any $\gamma \in \Pi(\mu, \nu)$,

   $\implies$ Any $\gamma \in \Pi(\mu, \nu)$ is supported on the points (a, b) and (a, c).

   $\implies$ $\gamma = \lambda \delta_{(a,b)} + (1 - \lambda)\delta_{(a,c)} = \lambda \delta_a \otimes \delta_b + (1 - \lambda)\delta_a \otimes \delta_c$, for some $\lambda \in [0, 1]$.
   Since $M_Y \gamma = \lambda \delta_b + (1 - \lambda)\delta_c$, $\nu = M_Y \gamma$ implies that $\lambda = \frac{1}{2}$.

2. <u>Monge VS Kantorovich</u>: discussed above.

3. <u>Uniqueness</u> Are minimizers unique? If not, can we characterize the set of minimizers?

Example 1.2. Consider $X = Y = \mathbb{R}^2$, $a = (-1, 0), b = (1, 0), a' = (0, -1), b' = (0, 1)$; $\mu = \frac{1}{2}(\delta_a + \delta_b), \nu = \frac{1}{2}(\delta_{a'} + \delta_{b'})$.

a) Consider Monge problem with quadratic distance cost

$$\inf_{T \in \mathcal{J}(\mu,\nu)} \int_{\mathbb{R}^2} |T(x) - x|^2 \mu(dx),$$

where $\mathcal{J}(\mu, \nu)$ is the set of push-forward maps, has two minimizers, defined on the support of μ:

$$T^{(1)}(a) = a', T^{(1)}(b) = b'; \quad T^{(1)}(a) = b', T^{(1)}(b) = a'.$$

Indeed, $\mathcal{J}(\mu, \nu) = \{T^{(1)}, T^{(2)}\}$, and

$$\int_{\mathbb{R}^2} |T^{(1)}(x) - x|^2 \mu(dx) = \frac{1}{2} \left(|T^{(1)}(a) - a|^2 + |T^{(1)}(b) - b|^2 \right)$$

$$= \frac{1}{2} \left(|a' - a|^2 + |b' - b|^2 \right) = \frac{1}{2}(2 + 2) = 2.$$

$$\int_{\mathbb{R}^2} |T^{(2)}(x) - x|^2 \mu(dx) = \frac{1}{2} \left(|T^{(2)}(a) - a|^2 + |T^{(2)}(b) - b|^2 \right)$$

$$= \frac{1}{2} \left(|b' - a|^2 + |a' - b|^2 \right) = \frac{1}{2}(2 + 2) = 2.$$

b) Non-uniqueness in Kantorovich is even bigger:

$$\Pi(\mu, \nu) = \left\{ \text{ convex combinations of } \gamma_{T^{(1)}} \text{ and } \gamma_{T^{(2)}} \right\}$$

$$= \left\{ \frac{1}{2} \left[(1 - \lambda)\delta_a \otimes \delta_{a'} + \lambda\delta_a \otimes \delta_{b'} + \lambda\delta_b \otimes \delta_{a'} + (1 - \lambda)\delta_b \otimes \delta_{b'} \right] : \lambda \in [0, 1] \right\}$$

4. <u>Exact solutions</u>

Example 1.3. Optimality of translation Let $X = Y = \mathbb{R}^d$, μ be a compactly supported probability measure on $\mathbb{R}^d$, and $\nu = (\tau_a)_{\#}\mu$, $\tau_a : \mathbb{R}^d \to \mathbb{R}^d$ translation by $a \in \mathbb{R}^d$, where $\tau_a(x) = x + a$. e.g., $\mu(dx) = f(x)dx$, then $\nu(dy) = g(y)dy$ with $g(y) = f(y - a)$ for all $y \in \mathbb{R}^d$.

$$\min I[T] = \int_{\mathbb{R}^d} |T(x) - x|^p \mu(dx), \ 1 < p < \infty$$

over $\mathcal{J}(\mu, \nu)$.

- Best to move each piece of mass by some distance?

- or better to move right side of pile to left side of hole (shorter distance) and make up for it by moving left side of pile to right side of hole (longer distance)?

The answer lies in *convexity* of the cost in displacement $T(x) - x$.

Definition 1.9

$\Phi : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ convex *if*

$$\Phi((1 - t)z + tz') \le (1 - t)\Phi(z) + t\Phi(z')$$

for all $z \ne z' \in \mathbb{R}^d, t \in (0, 1)$. *It is* strictly convex *if the inequality is always strict.*

> ### Theorem 1.10: Jensen's Inequality
>
> $\Phi : \mathbb{R}^d \to \mathbb{R}$ *convex and continuout. If μ is a probability measure on $\mathbb{R}^d$ and $v : \mathbb{R}^d \to \mathbb{R}^d$ integrable w.r.t. μ, then*
>
> $$\Phi\left(\int_{\mathbb{R}^d} u d\mu\right) \leq \int_{\mathbb{R}^d} \Phi(u) d\mu.$$
>
> *If Φ is strictly convex, inequality is strict, unless $u(x) = \bar{u} \in \mathbb{R}^d$ for μ-almost-everywhere x.*

Let us prove (using convexity) that uniform transition, $T = \tau_a$ is the best:

Step 1 Introduce centers of mass of μ and ν,

$$R_\mu = \int_{\mathbb{R}^d} x\mu(dx), R_\nu = \int_{\mathbb{R}^d} y\nu(dy),$$

then $R_\nu = \int_{\mathbb{R}^d} y((\tau_a)_{\#}\mu)dy = \int_{\mathbb{R}^d} \tau_a(x)\mu(dx) = \int_{\mathbb{R}^d} x\mu(dx) + \int_{\mathbb{R}^d} a\mu(dx) = R_\mu + a$.

Step 2 Average displacement of any $T \in \mathcal{J}(\mu, \nu)$

$$\int_{\mathbb{R}^d} (T(x) - x)\mu(dx) = \underbrace{\int_{\mathbb{R}^d} T(x)\mu dx}_{\int_{\mathbb{R}^d} y d(T_{\#}\mu)(y) = R_\nu} - \underbrace{\int_{\mathbb{R}^d} x\mu(dx)}_{R_\mu} = R_\nu - R_\mu = a$$

Step 3 Strict convexity of $\Phi_p : \mathbb{R}^d \to \mathbb{R}, z \to |z|^p$ for $1 < p < \infty$. By Jensen's inequality:

$$I[T] = \int_{\mathbb{R}^d} |T(x) - x|^p \mu(dx) = \int_{\mathbb{R}^d} \Phi_p(T(x) - x)\mu(dx)$$

$$\geq \Phi\underbrace{\left(\int_{\mathbb{R}^d} (T(x) - x)\mu(dx)\right)}_{a} = \Phi_p(a) = |a|^p$$

for every $T \in \mathcal{J}(\mu, \nu)$.

- $T = \tau_a$ achieves equality by $\tau_a(x) - x = x + a - x = a$, so $I[\tau_a] = \int_{\mathbb{R}^d} |a|^p \mu(dx) = |a|^p$.
- Φ_p is strictly convex for $p > 1$: so equality holds if and only if $T(x) - x$ is a constant μ-a.e., which implies that $T = \tau_a$ is a unique minimizer.

Example 1.4 (Book Shifting (Gangbo-McCann 1996)). Consider $\mu(dx) = \frac{4}{3}\mathbb{1}_{[0,\frac{3}{4}]}(x)dx$ and $\nu(dx) = \frac{4}{3}\mathbb{1}_{[\frac{1}{4},1]}(y)dy$. Consider the problem

$$\inf \int_{[0,1]} |T(x) - x|\mu(dx)$$

where define $\Phi_1(x) := |x|$. Since Φ_1 is convex, a solution is given by $T_1(x) = x + \frac{1}{4}$ for $x \in [0, 3/4]$, which is minimal. Its transportation cost is $\frac{3}{4} * \frac{4}{3} * \frac{1}{4} = \frac{1}{4}$. Consider another shift:

$$T_2(x) = \begin{cases} x + \frac{3}{4}, & \text{if } x \in [0, \frac{1}{4}] \\ x, & \text{otherwise.} \end{cases}$$

The transport cost is $\frac{4}{3}\int_{[0,\frac{1}{4}]} \frac{3}{4}dx + \frac{4}{3}\int_{[\frac{1}{4},\frac{3}{4}]} 0 = \frac{1}{4}$, so T_2 is also minimal. Notce Φ_1 is convex but not strictly.

Example 1.5 (A dam and two ditches). We now consider a problem where Monge's problem has an optimal value but the inf is not attained.

Let $X = Y = \mathbb{R}^2$, and let $\mu(dx) = dx_2|_{x_1=0, x_2 \in [0,1]}$, $\nu(dy) = \frac{1}{2}dy_2|_{y_1=-1, y_2 \in [0,1]} + \frac{1}{2}dy_2|_{y_1=1, y_2 \in [0,1]}$. The goal is

$$\min I[T] = \int_{\mathbb{R}^2} |T(x) - x|^2 \mu(dx) \text{ among all } T \in \mathcal{J}(\mu, \nu).$$

- $|T(x) - x| \geq d(\operatorname{supp}\mu, \operatorname{supp}\nu) = 1$ for all $x \in \operatorname{supp}\mu$ and for all $T \in \mathcal{J}(\mu, \nu)$., which implies that $I[T] \geq \int_{\mathbb{R}^2} 1\mu(dx) = 1$.

- Now we show $\inf_{T \in \mathcal{J}(\mu,\nu)} I[T] = 1$. Split up the dam into even number of segments of length $\epsilon > 0$ and consider the transport map T_2 depicted on the left. The maximal displacement of any point in the support is ϵ. Hence

$$|T_\epsilon(x) - x| \leq \sqrt{1 + \epsilon^2} \implies I(T_\epsilon) = \int_{\mathbb{R}^2} |T_\epsilon(x) - x|^2 \mu(dx) \leq 1 + \epsilon^2.$$

Since ϵ can be chosen arbitrarily, $\inf I = 1$.

Claim. No $T \in \mathcal{J}(\mu, \nu)$ achieves $I[T] = 1$.

Proof. Assume there exists $T \in \mathcal{J}(\mu, \nu)$ such that $I(T) = 1$. This means $T(x) \in \operatorname{supp}\nu$ for μ-a.e. x; $|T(x) - x| = 1$ for μ-a.e. x. Hence, $T(0, x_2) = (\pm 1, x_2)$ for every Lesbegue-a.e. $x_2 \in [0, 1]$. Set $\Omega_+ := \{x_2 \in [0, 1] : [T(0, x_2)]_1 = 1\}$ and $\Omega_- := \{x_2 \in [0, 1] : [T(0, x_2)]_1 = -1\}$. Then

$$T_\# \mu = dx_2|_{x_1 = -1, x_2 \in \Omega_-} + dx_2|_{x_1 = 1, x_2 \in \Omega_+} \neq \nu$$

so $T \notin \mathcal{J}(\mu, \nu)$. Hence $\inf_{T \in \preceq, \succeq} I[T]$ is not attained. $\square$

Everything holds if change $|T(x) - x|^2$ to $|T(x) - x|^p, p > 0$.
Problem: minimizing sequence T_ϵ exhibits faster and faster oscillations as $\epsilon \downarrow 0$. Hence, T_ϵ converges weakly, but not strongly in any $L^p(\mathbb{R}^2; \mu)$,

$$T_\epsilon(0, x_k) \xrightarrow[L^p]{\epsilon \downarrow 0} T_0(0, x_2) = \begin{pmatrix} 0 \\ x_2 \end{pmatrix}, \ \forall p > 1, \text{ but } T_0 \notin \mathcal{J}(\mu, \nu).$$

- Corresponding Kantororvich ProblemL

$$\inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^2 \times \mathbb{R}^2} |y - x|^2 \gamma(dxdy)$$

has simple explicit solution. Let $\gamma(dxdy) = \delta_0(x_1) \mathbb{1}_{[0,1]}(x_2) dx_2 \cdot \frac{\delta_1(y_1) + \delta_{-1}(y_1)}{2} \delta_{x_2}(y_2)$, which splits half of the mass to the left ditch and another half to the right one. It is not hard to see that the objective value of this γ is 1.

2   Multi-Marginal Optimal Transport (MMOT)