# Optimal transport

## I Monge and Kantorovich problems

Seongho, Joo

August 7, 2025

Seoul National University

# Notation

- In the following, we assume that $X$ is a complete and separable metric space.

- We denote $\mathcal{C}(X)$ the space of continuous functions, $\mathcal{C}_0(X)$ be the space of continuous function vanishing at infinity and $\mathcal{C}_b(X)$ be the space of bounded continuous functions.

- We denote $\mathcal{M}(X)$ the space of Borel regular measures on $X$ with finite total mass and

$$\mathcal{M}^+(X) := \{\mu \in \mathcal{M}(X) \,|\, \mu \leq 0\}$$

$$\mathcal{P}(X) := \big\{\mu \in \mathcal{M}^+(X) \,|\, \mu(X) = 1\big\}$$

# Reminders

**Definition 0.1** (Lower semi-continuous function). On a metric space $\Omega$, a function $f : \Omega \to \mathbb{R} \cup \{+\infty\}$ is said to be lower semi-continuous (l.s.c) is for every sequence $x_n \to x$ we have $f(x) \leq \liminf_n f(x_n)$.

**Definition 0.2** A metric space $\Omega$ is said to be compact if from any sequence $x_n$, we can extract a converging subsequence $x_{n_k} \to x \in \Omega$.

**Theorem 0.3** (Weierstrass). If $f : \Omega \to \mathbb{R} \cup \{+\infty\}$ is l.s.c and $\Omega$ is compact, then there exists $x^* \in \Omega$ such that $f(x^*) = \min \{f(X) \,|\, x \in \Omega\}$.

**Definition 0.4** (Weak and weak-* convergence) A sequence $x_n$ is a Banach space $\mathcal{X}$ is said to weakly converging to $x$ and we write $x_n \to x$, if for every $\eta \in \mathcal{X}'$ we have $\langle \eta, x_n \rangle \to \langle \eta, x \rangle$. A sequence $\eta_n \in \mathcal{X}'$ is a said to be weakly-* converging to $\eta \in \mathcal{X}'$, and we write $\eta_n \xrightarrow{w} \eta$, if for every $x \in \mathcal{X}$ we have $\langle \eta_n, x \rangle \to \langle \eta, x \rangle$.

**Theorem 0.5** (Banach-Alaoglu) If $\mathcal{X}'$ is separable and $\varphi_n$ is bounded sequence in $\mathcal{X}$, there exists a subsequence $\varphi_{n_k}$ weakly-* converging to some $\varphi \in \mathcal{X}'$.

# Reminders

**Theorem 0.6** (Riesz) Let $X$ be a compact metric space and $\mathcal{X} = C(X)$ then every element of $X$ is represented in a unique way as an element of $\mathcal{M}^+(X)$, this if for every $\eta \in \mathcal{X}$ there exists a unique $\lambda \in \mathcal{M}^+(X)$ such that $\langle \eta, \varphi \rangle = \int_X \varphi \,\mathrm{d}\lambda$ for every $\varphi \in \mathcal{X}$.

**Definition 0.7** (Narrow convergence) A sequence of finite measures $(\mu_n)_{n\geq 1}$ on $X$ narrowly converges to $\mu \in \mathcal{M}(X)$ if

$$\forall \varphi \in C_b(X), \quad \lim_{n\to\infty} \int_X \varphi_n \,\mathrm{d}\mu_n = \int_X \varphi \,\mathrm{d}\mu_n$$

# Monge problem

**Definition 1 (Push-forward and transport map)**

Let $X, Y$ be metric spaces, $\mu \in \mathcal{M}(X)$ and $T : X \to Y$ be a measurable map, The push-forward of $\mu$ by $T$ is the measure $T_{\#}\mu$ on $Y$ defined by

$$\forall B \subseteq Y,\ T_{\#}\mu(B) = \mu(T^{-1}(B))$$

or equivalently, if the following change-of-variable formula holds for all measurable and bounded $\varphi : Y \to \mathbb{R}$:

$$\int_Y \varphi(y)\,\mathrm{d}T_{\#}\mu(y) = \int_X \varphi(T(x))\,\mathrm{d}\mu x$$

A measurable map $T : X \to Y$ such that $T_{\#}\mu = \nu$ is also called a *transport map* between $\mu$ and $\nu$.

# Monge problem

**Example.** If $Y = \{y_1, \dots y_n\}$, then $T_{\#}\mu = \sum_{1 \le i \le n} \mu(T^{-1}(\{y_i\}))\delta_{y_i}$

**Example.** Assume that $T$ is $C^1$ diffeomorphism between open sets $X, Y$ of $\mathbb{R}^d$, and assume also that the probability measures $\mu, \nu$ have continuous densities $\rho, \sigma$ with respect to the Lebesgue measure. Then,

$$\int_Y \varphi(y)\sigma(y)\,\mathrm{d}y = \int_X \varphi\left(T(x)\right)\sigma(T(x))\mathsf{det}(\mathrm{D}T(x))\,\mathrm{d}x$$

Hence, $T$ is a transport map between $\mu$ and $\nu$ iff

$$\forall \varphi \in C_b(X),\ \int_X \varphi(T(x))\sigma(T(x))\mathsf{det}(\mathrm{D}T(x))\,\mathrm{d}x = \varphi(T(x))\rho(x)\,\mathrm{d}x$$

Hence, $T$ is a transport map iff the non-linear Jacobian equation holds

$$\rho(x) = \sigma(T(x))\mathsf{det}(\mathrm{D}T(x)).$$

# Monge problem

**Definition 2 (Monge problem)**

Consider two metric spaces $X, Y$, two probability measures $\mu \in \mathcal{P}(X).\nu \in \mathcal{P}(Y)$ and a cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. *Monge's problem* is the following optimization problem

$$\text{(MP)} := \inf \left\{ \int_X c(x, T(x)) \, \mathrm{d}\mu(X) \,|\, T : X \to Y \text{ and } T_\# \mu = \nu \right\}$$

This problem exhibits several difficulties, one of which is that both the constraint $(T_\# \mu \nu)$ and the functional are non-convex.

**Example.** There might exist no transport map between $\mu and \nu$. For instance, consider $\mu = \delta_x$ for some $x \in X$. Then $T_\# \mu(B) = \mu(T^{-1}(B)) = \delta_{T(x)}$. In particular if card(spt($\nu$))$> 1$ , there exists no transport map between $\mu$ and $\nu$.

**Example.** The infimum might not be attained even if $\mu$ is atomless. Consider for instance $\mu = \frac{1}{2}\lambda|_{\{\pm 1\} \times [-1,1]}$ on $\mathbb{R}^2$ and $\nu = \lambda|_{\{0\} \times [-1,1]}$, where $\lambda$ is the Lebesgue measure. One solution is to allow mass to split, leading to Kantrovich's relaxation of Monge's problem.

# Kantorovich problem

**Definition 3 (Marginals)**

The *Marginals* of a measure $\gamma$ on a product space $X \times Y$ are the measures $\pi_{X\#}\gamma$ and $\pi_{Y\#}\gamma$, where $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are the projection maps.

**Definition 4 (Transport plan)**

A transport plan between two probability measure $\mu, \nu$ on two metric spaces $X$ and $Y$ is a probability measure $\gamma$ on the product space $X \times Y$ whose marginals are $\mu$ and $\nu$. The space of transport plans is denotes as $\Pi(\mu, \nu)$, i.e.

$$\Pi(\mu, \nu) = \left\{\gamma \in \mathcal{P}(X \times Y) \,|\, \pi_{X\#}\gamma = \mu,\ \pi_{Y\#}\gamma = \nu\right\}$$

Note that $\Pi(\mu, \nu)$ is a convex set.

**Example. (Tensor product)** Note that the set of transport plans $\Pi(\mu, \nu)$ is never empty, as it contains the measure $\mu \otimes \nu$.

**Example. (Transport plan associated to a map)** Let $T$ be a transport map between $\mu, \nu$, and define $\gamma_T = (id, T)_{\#}\nu$. Then, $\gamma_T$ is a transport plan between $\mu$ and $\nu$.

# Kantorovich problem

**Definition 5 (Kantorovich problem)**

Consider two metric spaces $X, Y$, two probability measures $\mu \in \mathcal{P}(X).\nu \in \mathcal{P}(Y)$ and a cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. *Kantorovich's problem* is the following optimization problem

$$\text{(KP)} := \inf \left\{ \int_{X \times Y} c(x, y) \,\mathrm{d}\gamma(x, y) \,|\, \gamma \in \Pi(\mu, \nu) \right\}$$

**Remark.** The infimum in Kantorovich problem is less than the infimum in Monge problem. Indeed, consider a transport map satisfying $T_{\#}\mu = \nu$ and the associated transport plan $\gamma_T$. Then, by the change of variable one has

$$\int_{X \times Y} c(x, y) \,\mathrm{d}(id, T)_{\#}\mu(x, y) = \int_X c(x, T(x)) \,\mathrm{d}\mu$$

**Example. (Finite support)** Assume that $X = Y = \{1 \dots N\}$ and that $\mu, \nu$ are the uniform probability measures over $X$ and $Y$. Then monge's problem can be rewritten as a minimization problem over bijections between $X$ and $Y$:

$$\min \left\{ \frac{1}{N} \sum_{1 \le i \le N} c(i, \sigma(i)) \,|\, \sigma \in \mathfrak{S}_N \right\}$$

# Cont.

In kantorovich's relaxation, the set of transport plans $\Pi(\mu, \nu)$ agrees with the set of bijection stochastic matrices:

$$\gamma \in \Pi(\mu, \nu) \iff \gamma \geq 0, \sum_i \gamma(i, j) = \frac{1}{N} = \sum_j \gamma(i, j)$$

By Birkhoff's theorem, any extremal bi-stochastic matrix is induced by a permutation. This shows that, in this case, the solution to Monge's and Kantorovich's problem agrees.

# Cont.

In kantorovich's relaxation, the set of transport plans $\Pi(\mu, \nu)$ agrees with the set of bijection stochastic matrices:

$$\gamma \in \Pi(\mu, \nu) \iff \gamma \geq 0, \sum_i \gamma(i,j) = \frac{1}{N} = \sum_j \gamma(i,j)$$

By Birkhoff's theorem, any extremal bi-stochastic matrix is induced by a permutation. This shows that, in this case, the solution to Monge's and Kantorovich's problem agrees.

**Definition 6 (Support)**

Let $\Omega$ be a separable metric space. The *support* of a non-negative measure $\mu$ is the smallest closed set on which $\mu$ is concentrated

$$\text{spt}(\mu) := \bigcap \{A \subseteq \Omega \,|\, A \text{ is closed and } \mu(\Omega \setminus A) = 0\}$$

A point $x$ belongs to spt($\mu$) iff for every $r > 0$ one has $\mu(B(x,r)) > 0$.

# Cont(2).

**Proposition 1**

Let $\gamma \in \Pi(\mu, \nu)$ and $T : X \to Y$ measurable be such that $\gamma\left(\{(x, y) \in X \times Y \mid T(x) \neq y\}\right) = 0$. Then $\gamma = \gamma_T$.

**Proof.**

# Cont(2).

**Proposition 1**

Let $\gamma \in \Pi(\mu, \nu)$ and $T : X \to Y$ measurable be such that $\gamma\left(\{(x, y) \in X \times Y \mid T(x) \neq y\}\right) = 0$. Then $\gamma = \gamma_T$.

**Proof.**

# Solutions to Kantorovich's problem

**Theorem 1**

Let $X, Y$ be two compact spaces, and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous cost function, which is bounded from below. Then kantorovich's problem admits a minimizer.

**Lemma 1 (Lower semi-continuity of measure )**

Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous function, which is also bound from below. Define $\mathcal{F} : \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\}$ through $\mathcal{F}(\mu) = \int_X f \, \mathrm{d}\mu$. Then, $\mathcal{F}$ is lower-semicontinuous for the narrow convergence, i.e.

$$\forall \mu_n \rightharpoonup \mu, \ \liminf_{n \to \infty} \mathcal{F}(\mu_n) \geq \mathcal{F}(\mu)$$

**Theorem Proof.**

# Kantorovich as a relaxation of Monge

The question that we consider here is the equality between the infimum in Monge problem and the minimum in Kantorovich problem.

**Theorem 2**

Let $X = Y$ be a compact subset of $\mathbb{R}^d$, $c \in C(X \times Y)$ and $\mu \in \mathcal{P}(X)$, $\mu \in \mathcal{P}(Y)$. Assume that $\mu$ is atomelss. Then,

$$\inf(\mathsf{MP}) = \min(\mathsf{KP})$$

**Example 3.3** Consider $\mu_i = \frac{1}{2}(\delta_{x_i} + \alpha\lambda_{B(y_i,1)})$ with $\alpha = \frac{1}{\lambda(B(y_i,1))}$ on $\mathbb{R}^2$ with $c(x,y) = \|x-y\|$. Then, any transport map must transport the Dirac to the Dirac and the ball to the ball, so that is cost is $\|x_1 - x_2\| + \|y_1 - y_2\|$. On the other hand, a transport plan can transport $\delta_{x_1}$ to $\alpha\lambda|_{B(y_2,1)}$ with cost $\leq \|x_1 - y_2\| + 1$. The total cost of this transport plan is $2 = \|x_1 - y_2\| + \|x_2 - y_1\|$, which can be lower that $\|x_1 - x_2\| + \|y_1 - y_2\|$.

# Cont.

**Lemma 2**

If $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and $\mu$ has no atoms, then $\exists T : \mathbb{R}^d \to \mathbb{R}^d$ measurable such that $T_\# \mu = \nu$.

**Lemma 3**

Let $K$ be a compact metric space. For any $\varepsilon > 0$, there exists a (measurable) partition $K_1, \dots K_N$ of $K$ such that for every $i$, diam$(K_i) \leq \varepsilon$

**Proof.** By compactness, there exists $N$ points $x_1, \dots x_N$ such that $K \subseteq \bigcup_i \mathrm{B}(x_i, \varepsilon)$. The partition $K_1, \dots, K_N$ of $K$ defined recursively by

$$K_i = \{x \in K \setminus (K_1 \cup \dots \cup K_{i-1}) \,|\, \forall j \in [N]_{-i},\ d(x, x_i) \leq d(x, x_j)\}$$

so that $K_i \subseteq \mathrm{B}(x_i, \varepsilon)$

**Theorem ?? Proof.**

# The dual problem

Let write down the constraint $\gamma \in \Pi(\mu, \nu)$ as follows: if $\gamma \in \mathcal{M}^+(X \times Y)$ we have

$$\sup_{\varphi,\psi} \int_X \varphi \,\mathrm{d}\mu + \int_Y \varphi \,\mathrm{d}\nu + \int_{X\times Y} (\varphi(x) + \psi(y))\,\mathrm{d}\gamma = \begin{cases} 0 & \text{if } \gamma \in \Pi(\mu,\nu), \\ +\infty & \text{otherwise }, \end{cases}$$

where the supremum is taken on $C_b(X) \times C_b(Y)$. We can now remove the constraint on $\gamma$ in (KP)

$$\inf_{\gamma \in \mathcal{M}^+(X\times Y)} \int_{X\times Y} c \,\mathrm{d}\gamma + \sup_{\varphi,\psi} \int_X \varphi \,\mathrm{d}\mu + \int_Y \psi \,\mathrm{d}\nu - \int_{X\times Y} (\varphi(x) + \psi(y))\,\mathrm{d}\gamma$$

and by interchanging sup and inf we get

$$\sup_{\varphi,\psi} \int_X \varphi \,\mathrm{d}\mu + \int_Y \psi \,\mathrm{d}\nu + \inf_{\gamma \in \mathcal{M}^+(X\times Y)} \int_{X\times Y} (c(x,y) - \varphi(x) - \psi(y))\,\mathrm{d}\gamma$$

One can now rewrite the inf in $\gamma$ as constraint on $\varphi$ and $\psi$ as

$$\inf_{\gamma \in \mathcal{M}^+(X\times Y)} \int_{X\times Y} (c - \varphi \oplus \psi)\,\mathrm{d}\gamma = \begin{cases} 0 & \text{if } \varphi \oplus \psi \le c \text{ on } X \times Y \\ -\infty & \text{otherwise} \end{cases}$$

where $\varphi \oplus \psi(x,y) := \varphi(x) + \psi(y)$

# The dual problem

**Definition 7 (Dual Problem)**

Given $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$ and a cost function $c \in C(X \times Y)$. The dual problem is the following optimization problem

$$\text{(DP)} := \sup \left\{ \int_X \varphi \,\mathrm{d}\mu + \int_Y \psi \,\mathrm{d}\nu \,|\, \varphi \in C_b(X), \psi \in C_b(Y), \varphi \oplus \psi \leq c \right\}$$

**Remark.** One trivially has the weak duality inequality (KP) $\geq$ (DP). Indeed, denote

$$L(\gamma, \varphi, \psi) = \int_{X \times Y} (c - \varphi \oplus \psi) \,\mathrm{d}\gamma) + \int_X \varphi \,\mathrm{d}\mu + \int_Y \psi \,\mathrm{d}\nu$$

one has for any $(\varphi, \psi, \gamma) \in C_b(X) \times C_b(X) \times \mathcal{M}^+(X \times Y)$,

$$\inf_{\tilde{\gamma} \geq 0} L(\tilde{\gamma}, \varphi, \psi) \leq L(\gamma, \varphi, \psi) \leq \sup_{\tilde{\varphi}, \tilde{\psi}} L(\gamma, \tilde{\varphi}, \tilde{\psi})$$

Taking the supremum with respect to $(\varphi, \psi)$ on the left and the infimum with respect to $\gamma$ on the right gives $\inf$ (KP) $\geq \sup$ (DP). When $\sup(\text{DP}) = \int(\text{KP})$, one talks of strong duality.

We now focus on the existence of a pair $(\varphi, \psi)$ which solves (DP).

# The dual problem

**Definition 8 ($c$-transform and $\bar{c}$-transform)**

Given a function $f : x \to \bar{\mathbb{R}}$, we define its $\bar{c}$-transform $f^c : Y \to \mathbb{R}bar\mathbb{R}$ by

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

We also define the $\bar{c}$-transform of $g : Y \to \bar{\mathbb{R}}$ by

$$g^{\bar{c}}(x) = \inf_{y \in Y} c(x, y) - g(y)$$

We also way that a function $\psi$ on $Y$ is $\bar{c}$-concave if there exists $f$ such that $\psi = f^c$.

If we consider $f^c$, we have that $f^c(y) = \inf_x \tilde{f}_x(y)$ with $\tilde{f}_x(y) = c(x, y) - f(x)$, and the functions $\tilde{f}_x$ satisfy $\tilde{f}_x(y) - \tilde{f}_x(y) \leq w(d_Y(y, y'))$. This implies that $f^c$ actually shares the same continuity modulus of $c$.

# The dual problem

**Theorem 3**

Suppose $X$ and $Y$ are compact and $c \in C(X \times Y)$. Then there exists a pair $(\varphi^{c\bar{c}}, \varphi^{c})$ which solves (DP).

**Proof.** Let us first denote by $J(\varphi, \psi)$ the following functional

$$J(\varphi, \psi) = \int_X \varphi \, \mathrm{d}\mu + \int_Y \psi \, \mathrm{d}\nu$$

note that for every constant $\lambda$, $J(\varphi - \lambda, \psi + \lambda) = J(\varphi, \psi)$. Given now a maximizing sequence $(\varphi_n, \psi_n)$ we can improve it by means of the c- and $\bar{c}$-transform obtaining a new one $\varphi_n^{c\bar{c}}, \varphi_n^{c})$. Notice that by the consideration above the sequences $\varphi_n^{c\bar{c}}, \varphi_n^{c}$ are uniformly equicontinuous. Since $\varphi_n^{c}$ is continuous on a compact set we can always subtract is minimum and assume that $\min \varphi_n^{c} = 0$. This implies that the sequence $\varphi_n^{c}$ is also equibounded as $0 \leq \varphi_n^{c} \leq w(\text{diam}(Y))$. We also deduce uniform bounds on $\varphi_n^{c\bar{c}}$ as $\varphi_n^{c\bar{c}} = \inf_Y c(x, y) - \varphi^{c}(y)$. This let us apply Ascoli-Arzela's theorem and extract two uniformly converging subsequences $\varphi_{n_k}^{c\bar{c}} \to \bar{\varphi}$ an $\varphi_{n_K}^{c} \to \bar{\psi}$ where the pair $(\bar{\varphi}, \bar{\psi})$ satisfies the inequality constraint. Moreover, since $(\varphi_n^{c\bar{c}}, \varphi_n^{c})$ is a maximising sequence we get that the pair $(\bar{\varphi}, \bar{\psi})$ is optimal. Now one can apply again the $c-$ and $\bar{c}-$-transforms obtaining an optimal pair of the form $(\bar{\varphi}^{c\bar{c}}, \bar{\varphi}^{c})$