Optimal Transport

Abteilung für Mathematische Stochastik

www.stochastik.uni-freiburg.de
thorsten.schmidt@stochastik.uni-freiburg.de

SS 2024

Monge's Transport Problem

Gaspard Monge (1781). „Mémoire sur la théorie des déblais et des remblais“. In: **Mem. Math. Phys. Acad. Royale Sci.** Pp. 666–704

- **Monge's problem:** transport soil from extraction sites to construction sites.
- **Goal:** minimize total transport cost.
- **Cost:** product of mass and distance.

Monge's problem is the search of an optimal coupling; He was looking for a deterministic optimal coupling.

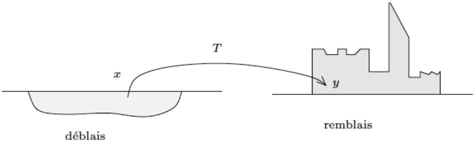

Figure: Illustration of Monge's Problem

- The following explanations largely adhere to Villani et al. (2009).

Coupling

Definition (Coupling)

Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be two probability spaces. Coupling μ and ν means constructing two random variables X and Y on some probability space $(\Omega, \mathbb{P})$, such that $\mathrm{law}(X) = \mu$, $\mathrm{law}(Y) = \nu$. The couple (X, Y) is called a coupling of (μ, ν). The law of (X, Y) is also called a coupling of (μ, ν).

Definition (Deterministic Coupling)

A coupling (X, Y) is said to be deterministic if there exists a measurable function $T : \mathcal{X} \to \mathcal{Y}$ such that $Y = T(X)$.

Unlike couplings, deterministic couplings do not always exist. To say that (X, Y) is a deterministic coupling of μ and ν is strictly equivalent to any one of the four statements below:

- (X, Y) is a coupling of μ and ν whose law π is concentrated on the graph of a measurable function $T : \mathcal{X} \to \mathcal{Y}$;
- X has law μ and $Y = T(X)$, where $T_{\#}\mu = \nu$;
- X has law μ and $Y = T(X)$, where T is a change of variables from μ to ν: for all ν-integrable (resp. nonnegative measurable) functions φ,

$$\int_Y \varphi(y)\, d\nu(y) = \int_X \varphi\big(T(x)\big)\, d\mu(x); \tag{1}$$

- $\pi = (\mathrm{Id}, T)_{\#}\mu$.

It is common to call T the transport map: Informally, one can say that T transports the mass represented by the measure μ, to the mass represented by the measure ν.

Existence of optimal couplings

Definition (Upper Semicontinuous Function)

A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is called upper semicontinuous at a point $x_0 \in \mathcal{X}$ if

$$\limsup_{x \to x_0} f(x) \leq f(x_0).$$

Theorem (Existence of an optimal coupling)

Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be two Polish probability spaces; let $a : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ and $b : \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be two upper semicontinuous functions such that $a \in L^1(\mu)$, $b \in L^1(\nu)$. Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous cost function, such that $c(x, y) \geq a(x) + b(y)$ for all x, y. Then there is a coupling of (μ, ν) which minimizes the total cost $E[c(X, Y)]$ among all possible couplings (X, Y).

Existence of optimal couplings - Proof

Lemma (Lower semicontinuity of the cost functional)

Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces, and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ a lower semicontinuous cost function. Let $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be an upper semicontinuous function such that $c \geq h$. Let $(\pi_k)_{k \in \mathbb{N}}$ be a sequence of probability measures on $\mathcal{X} \times \mathcal{Y}$, converging weakly to some $\pi \in \mathcal{P}(X \times Y)$, in such a way that $h \in L^1(\pi_k)$, $h \in L^1(\pi)$, and

$$\int_{\mathcal{X} \times \mathcal{Y}} h \, d\pi_k \xrightarrow{k \to \infty} \int_{\mathcal{X} \times \mathcal{Y}} h \, d\pi.$$

Then

$$\int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi \leq \liminf_{k \to \infty} \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi_k.$$

In particular, if c is nonnegative, then $F : \pi \mapsto \int c \, d\pi$ is lower semicontinuous on $\mathcal{P}(X \times Y)$, equipped with the topology of weak convergence.

Lemma (Tightness of transference plans)

Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces. Let $P \subset \mathcal{P}(\mathcal{X})$ and $Q \subset \mathcal{P}(\mathcal{Y})$ be tight subsets of $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ respectively. Then the set $\Pi(P, Q)$ of all transference plans whose marginals lie in P and Q respectively, is itself tight in $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$.

Existence of optimal couplings - Proof

Since $\mathcal{X}$ is Polish, $\{\mu\}$ is tight in $\mathcal{P}(\mathcal{X})$; similarly, $\{\nu\}$ is tight in $\mathcal{P}(\mathcal{Y})$. By Lemma (Tightness of transference plans), $\Pi(\mu, \nu)$ is tight in $\mathcal{P}(X \times Y)$, and by Prokhorov's theorem this set has a compact closure. By passing to the limit in the equation for marginals, we see that $\Pi(\mu, \nu)$ is closed, so it is in fact compact.

Let $(\pi_k)_{k\in\mathbb{N}}$ be a sequence of probability measures on $X \times Y$, such that

$$\int c\, d\pi_k \to \inf_{\pi\in\Pi(\mu,\nu)} \int c\, d\pi.$$

Extracting a subsequence if necessary, we may assume that π_k converges to some $\pi \in \Pi(\mu, \nu)$. The function $h : (x, y) \mapsto a(x) + b(y)$ lies in $L^1(\pi_k)$ and in $L^1(\pi)$, and $c \geq h$ by assumption; moreover,

$$\int h\, d\pi_k = \int h\, d\pi = \int a\, d\mu + \int b\, d\nu.$$

So Lemma (Lower semicontinuity of the cost functional) implies

$$\int c\, d\pi \leq \liminf_{k\to\infty} \int c\, d\pi_k.$$

Thus π is minimizing. □

Lower semicontinuity of the cost functional - Proof

Lemma (Lower semicontinuity of the cost functional)

Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces, and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ a lower semicontinuous cost function. Let $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{-\infty\}$ be an upper semicontinuous function such that $c \geq h$. Let $(\pi_k)_{k \in \mathbb{N}}$ be a sequence of probability measures on $\mathcal{X} \times \mathcal{Y}$, converging weakly to some $\pi \in \mathcal{P}(X \times Y)$, in such a way that $h \in L^1(\pi_k)$, $h \in L^1(\pi)$, and

$$\int_{\mathcal{X} \times \mathcal{Y}} h \, d\pi_k \xrightarrow{k \to \infty} \int_{\mathcal{X} \times \mathcal{Y}} h \, d\pi.$$

Then

$$\int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi \leq \liminf_{k \to \infty} \int_{\mathcal{X} \times \mathcal{Y}} c \, d\pi_k.$$

In particular, if c is nonnegative, then $F : \pi \mapsto \int c \, d\pi$ is lower semicontinuous on $\mathcal{P}(X \times Y)$, equipped with the topology of weak convergence.

Proof: Replacing c by $c - h$, we may assume that c is a nonnegative lower semicontinuous function. Then c can be written as the pointwise limit of a nondecreasing family $(c_\ell)_{\ell \in \mathbb{N}}$ of continuous real-valued functions. By monotone convergence,

$$\int c \, d\pi = \lim_{\ell \to \infty} \int c_\ell \, d\pi = \lim_{\ell \to \infty} \lim_{k \to \infty} \int c_\ell \, d\pi_k \leq \liminf_{k \to \infty} \int c \, d\pi_k.$$

□

- **Theorem of Baire:** Assume $\mathcal{X}$ is a metric space. Every lower semicontinuous function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is the limit of a monotone increasing sequence of extended real-valued continuous functions on X; if f does not take the value $-\infty$, the continuous functions can be taken to be real-valued.

Tightness of transference plans - Proof

Lemma (Tightness of transference plans)

Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces. Let $P \subset \mathcal{P}(\mathcal{X})$ and $Q \subset \mathcal{P}(\mathcal{Y})$ be tight subsets of $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ respectively. Then the set $\Pi(P, Q)$ of all transference plans whose marginals lie in P and Q respectively, is itself tight in $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$.

Proof: Let $\mu \in P$, $\nu \in Q$, and $\pi \in \Pi(\mu, \nu)$. By assumption, for any $\epsilon > 0$ there is a compact set $K_\epsilon \subset X$, independent of the choice of μ in P, such that $\mu[X \setminus K_\epsilon] \leq \epsilon$; and similarly, there is a compact set $L_\epsilon \subset Y$, independent of the choice of ν in Q, such that $\nu[Y \setminus L_\epsilon] \leq \epsilon$. Then for any coupling (X, Y) of (μ, ν),

$$\mathbb{P}\left((X, Y) \notin K_\epsilon \times L_\epsilon\right) \leq \mathbb{P}[X \notin K_\epsilon] + \mathbb{P}[Y \notin L_\epsilon] \leq 2\epsilon.$$

The desired result follows since this bound is independent of the coupling, and $K_\epsilon \times L_\epsilon$ is compact in $X \times Y$.

□

Remarks on the Theorem

- The lower bound for c ensures that the expected costs $\mathbb{E}[c(X, Y)]$ are well-defined in $\mathbb{R} \cup \{+\infty\}$. Often, c is non-negative, so one can choose $a = 0$ and $b = 0$.
- This existence theorem does not imply that the optimal cost is finite. It might be that all transport plans lead to an infinite total cost, i.e.,

  $$\int c \, d\pi = +\infty \quad \text{for all } \pi \in \Pi(\mu, \nu).$$

  A simple condition to rule out this annoying possibility is

  $$\int c(x, y) \, d\mu(x) \, d\nu(y) < +\infty,$$

  which guarantees that at least the independent coupling has finite total cost. A stronger assumption is

  $$c(x, y) \leq c_X(x) + c_Y(y), \quad (c_{\mathcal{X}}, c_{\mathcal{Y}}) \in L^1(\mu) \times L^1(\nu),$$

  which implies that any coupling has finite total cost.

Kantorovich-Rubinstein-Duality

Theorem (Kantorovich-Rubinstein-Duality)

Let $c : X \times Y \to [0, \infty]$ *be a lower semicontinuous cost function. Then,*

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c \, d\pi = \sup_{(\varphi,\psi) \in I_c} \left(\int_X \varphi(x) \, d\mu(x) + \int_Y \psi(y) \, d\nu(y) \right),$$

where $I_c := \{(\varphi, \psi) \in \textit{Lip}_b(X) \times \textit{Lip}_b(Y) : \varphi(x) + \psi(y) \leq c(x, y)\}$.

Economic Interpretation: Let $\mathcal{X}$ be a set of bakeries and $\mathcal{Y}$ be a set of cafes. The problem in the Kantorovich formulation corresponds to minimizing the cost of a consortium between bakeries and cafes. Now assume that there is a transportation company that buys a unit from the bakery $x \in \mathcal{X}$ at the price $\varphi(x)$ and sells it to the cafe $y \in \mathcal{Y}$ at the price $\psi(y)$. To be competitive with the direct agreement between bakeries and cafes, it must hold that $\psi(y) - \varphi(x) \leq c(x, y)$. Then the profit is

$$\int_Y \psi \, d\nu - \int_X \varphi \, d\mu,$$

which corresponds to the dual formulation (except for the sign change of φ).

The Wasserstein distances

Definition (Wasserstein distances)

Let $(\mathcal{X}, d)$ be a Polish metric space, and let $p \in [1, \infty)$. For any two probability measures μ, ν on $\mathcal{X}$, the Wasserstein distance of order p between μ and ν is defined by the formula

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p \, d\pi(x, y) \right)^{1/p}$$

$$= \inf \left\{ \mathbb{E}[d(X, Y)^p]^{1/p} \mid \text{law}(X) = \mu, \text{law}(Y) = \nu \right\}. \tag{2}$$

- **Example:** $W_p(\delta_x, \delta_y) = d(x, y)$. In this example, the distance does not depend on p; but this is not the rule.
- At the present level of generality, W_p is still not a distance in the strict sense, because it might take the value $+\infty$; but otherwise it does satisfy the axioms of a distance.

Definition (Wasserstein space)

The Wasserstein space of order p is defined as

$$\mathcal{P}_p(\mathcal{X}) := \left\{ \mu \in \mathcal{P}(X) \mid \int_X d(x_0, x)^p \, \mu(dx) < +\infty \right\},$$

where $x_0 \in \mathcal{X}$ is arbitrary. This space does not depend on the choice of the point x_0. Then W_p defines a (finite) distance on $\mathcal{P}_p(\mathcal{X})$.

Convergence in Wasserstein sense

The notation $\mu_k \xrightarrow{w} \mu$ means that μ_k converges weakly to μ, i.e.

$$\int \varphi \, d\mu_k \to \int \varphi \, d\mu \text{ for any bounded continuous } \varphi.$$

Definition (Weak convergence in $\mathcal{P}_p$)

Let $(\mathcal{X}, d)$ be a Polish space, and $p \in [1, \infty)$. Let $(\mu_k)_{k \in \mathbb{N}}$ be a sequence of probability measures in $\mathcal{P}_p(\mathcal{X})$ and let μ be another element of $\mathcal{P}_p(\mathcal{X})$. Then (μ_k) is said to converge weakly in $\mathcal{P}_p(\mathcal{X})$ if any one of the following equivalent properties is satisfied for some (and then any) $x_0 \in \mathcal{X}$:

(i) $\mu_k \xrightarrow{w} \mu$ and $\int d(x_0, x)^p \, d\mu_k(x) \to \int d(x_0, x)^p \, d\mu(x)$;

(ii) $\mu_k \xrightarrow{w} \mu$ and $\limsup_{k \to \infty} \int d(x_0, x)^p \, d\mu_k(x) \le \int d(x_0, x)^p \, d\mu(x)$;

(iii) $\mu_k \xrightarrow{w} \mu$ and $\lim_{R \to \infty} \limsup_{k \to \infty} \int_{d(x_0, x) \ge R} d(x_0, x)^p \, d\mu_k(x) = 0$;

(iv) For all continuous functions φ with $|\varphi(x)| \le C(1 + d(x_0, x)^p)$, $C \in \mathbb{R}$, one has $\int \varphi(x) \, d\mu_k(x) \to \int \varphi(x) \, d\mu(x)$.

Convergence in Wasserstein sense

Theorem (W_p metrizes $\mathcal{P}_p$)

Let $(\mathcal{X}, d)$ be a Polish space, and $p \in [1, \infty)$; then the Wasserstein distance W_p metrizes the weak convergence in $\mathcal{P}_p(\mathcal{X})$. In other words, if $(\mu_k)_{k\in\mathbb{N}}$ is a sequence of measures in $\mathcal{P}_p(\mathcal{X})$ and μ is another measure in $\mathcal{P}(\mathcal{X})$, then the statements

$$\mu_k \text{ converges weakly in } \mathcal{P}_p(\mathcal{X}) \text{ to } \mu$$

and

$$W_p(\mu_k, \mu) \to 0$$

are equivalent.

Lemma (Continuity of W_p)

If (X, d) is a Polish space, and $p \in [1, \infty)$, then W_p is continuous on $\mathcal{P}_p(\mathcal{X})$. More explicitly, if μ_k (resp. ν_k) converges to μ (resp. ν) weakly in $\mathcal{P}_p(\mathcal{X})$ as $k \to \infty$, then

$$W_p(\mu_k, \nu_k) \to W_p(\mu, \nu).$$

Lemma (Metrizability of the weak topology)

Let (X, d) be a Polish space. If $\tilde{d}$ is a bounded distance inducing the same topology as d (such as $\tilde{d} = \frac{d}{1+d}$), then the convergence in Wasserstein sense for the distance $\tilde{d}$ is equivalent to the usual weak convergence of probability measures in $\mathcal{P}(X)$.

Monge, Gaspard (1781). „Mémoire sur la théorie des déblais et des remblais“. In: **Mem. Math. Phys. Acad. Royale Sci.** Pp. 666–704.

Villani, Cédric et al. (2009). **Optimal transport: old and new**. Vol. 338. Springer.