OPTIMAL TRANSPORT
Lecture I 17/02/2021

THE MONGE-KANTOROVICH PROBLEMS

Evaluations: • report on practical classes
↳ notebook jupyter python

• report on a paper

• Oral presentation

→ book ~~Sa~~ by F Santambrogio
" Optimal transport for applied mathematicians "

→ Ambrosio, Gigli, Savaré

→ Villani
→ topics on OT
↳ Old and New]

• Motivation

WASSERSTEIN

→ OT defines a distance between probability measures

→ interpretation of some parabolic PDE as Wasserstein gradient flows

i.e.
$$\begin{cases} \partial_t \rho + \mathrm{div}(\rho v) = 0 \\ v = -\nabla \log \rho \end{cases}$$

$$\nabla \log \rho = \frac{\nabla \rho}{\rho}$$ → heat equation

W gradient flow $\rho^{k+1} \in \mathrm{argmin}\ W_2^2(\rho, \rho^k) + F(\rho)$

$$F(\rho) = \int \rho \log \rho$$

$\rho^h \longrightarrow \rho$ sol to the heat eq

• → interesting geometry on $\rho(x)$

machine learning, inverse pb.

imaging etc.

→ Quantum physics: electronic configurations of molecules and atoms

X complete sep metric space

$\mathcal{M}(X)$ finite measures

$\mathcal{M}^+(X) = \{ \mu \in \mathcal{M}(X) \mid \mu \geqslant 0 \}$

$\mathcal{P}(X) = \{ \mu \in \mathcal{M}^+(X) \mid \mu(X) = 1 \}$

Some reminders:

Definition 0.1 (Lower semi-continuous function). On a metric space Ω, 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 $f(x) \leqslant \liminf_n f(x_n)$.

Definition 0.2. A metric space Ω 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 Ω is compact, then there exists $x^\star \in \Omega$ such that $f(x^\star) = \min\{f(x) \mid x \in \Omega\}$.*

Definition 0.4 (weak and weak$-\star$ convergence). A sequence x_n in a Banach space $\mathcal{X}$ is said to be weakly converging to x and we write $x_n \rightharpoonup x$, if for every $\eta \in \mathcal{X}'$ ($\mathcal{X}'$ is the topological dual of $\mathcal{X}$ and $\langle \cdot, \cdot \rangle$ is the duality product) we have $\langle \eta, x_n \rangle \to \langle \eta, x \rangle$. A sequence $\eta_n \in \mathcal{X}'$ is said to be weakly-$\star$ converging to $\eta \in \mathcal{X}'$, and we write $\eta_n \overset{\star}{\rightharpoonup} \eta$, 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 η_n is a bounded sequence in $\mathcal{X}'$, then there exists a subsequence η_{n_k} weakly-$\star$ converging to some $\eta \in \mathcal{X}'$*

Theorem 0.6 (Riesz). *Let X be a compact metric space and $\mathcal{X} = \mathcal{C}(X)$ then every element of $\mathcal{X}$ is represented in a unique way as an element of $\mathcal{M}^+(X)$, that is 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 $(\mu_n)_{n \geqslant 1}$ on X narrowly converges to $\mu \in \mathcal{M}(X)$ if

$$\forall \varphi \in \mathcal{C}_b(X), \quad \lim_{n \to \infty} \int_X \varphi \mathrm{d}\mu_n = \int_X \varphi \mathrm{d}\mu.$$

With a slightly abuse of notation we will denote it by $\mu_n \rightharpoonup \mu$.

Remark 0.8. Since we will mostly work on compact set X, then $\mathcal{C}(X) = \mathcal{C}_0(X) = \mathcal{C}_b(X)$. This means that the narrow convergence of measures, that is the notion of convergence in duality with $\mathcal{C}_b(X)$, corresponds to the weak-$\star$ convergence (the convergence in duality with $\mathcal{C}_0(X)$).

Monge problem

x $\mu(x)$ = production of x y $\nu(y)$ = demand

looking for T (which matches a bakery x with a cafés y) such that it minimizes an average ~~co.~~ transportation cost

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

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

eq ~~ca~~ change of variable formula

$$\forall \varphi \qquad \boxed{\int_Y \varphi(y)\, dT_\#\mu(y) = \int_X \varphi(T(x))\, d\mu(x)}$$

A measurable map $T: X \to Y$ such that $T_{\#}\mu = \nu$ is called TRANSPORT MAP between μ and ν.

<u>Ang</u>. T is a C^1 diff. between open sets $X, Y \subseteq \mathbb{R}^d$

- $\mu, \nu \ll \mathcal{L}$ $\quad \frac{d\mu}{d\mathcal{L}} = \bar{\mu} \quad \frac{d\nu}{d\mathcal{L}} = \bar{\nu}$

$$\int_Y \varphi(y)\bar{\nu}(y)dy = \int_X \varphi(T(x))\bar{\nu}(T(x)) \det(DT(x))\,dx$$

$$= \int_X \varphi(T(x))\bar{\mu}(x)\,dx$$

$$\Rightarrow \boxed{\bar{\mu}(x) = \bar{\nu}(T(x)) \det(DT(x))}$$

$\hookrightarrow$ Monge-Ampere equations

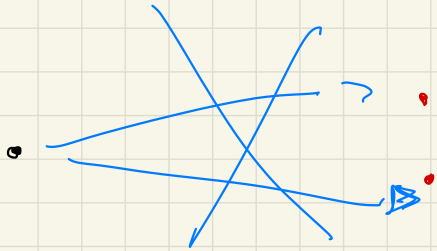
<u>def</u> (Monge problem). X, Y two metric spaces $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ and a cost function $c: X \times Y \to \mathbb{R} \cup \{+\infty\}$. Monge problem is the following optimization problem

(MP) $\inf \left\{ \int_X c(x, T(x))\, d\mu(x) \;\middle|\; \begin{array}{l} T : X \to Y \\ T_\# \mu = \nu \end{array} \right\}$

<u>ex</u> 1 $c(x,y) = \|x-y\|^2$ quadratic cost

Brenier's cost

<u>ex</u> 2



$\nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$

$\mu = \delta_{x_1}$

In general consider $\mu = \delta_x$

then $T_\# \mu(B) = \mu(T^{-1}(B)) = \delta_{T(x)}$

if $\text{card}(\text{spt}(\nu)) > 1 \Rightarrow$ ~~the~~ $\nexists$ transport map between μ and ν

<u>ex</u> 3 T ok ☺

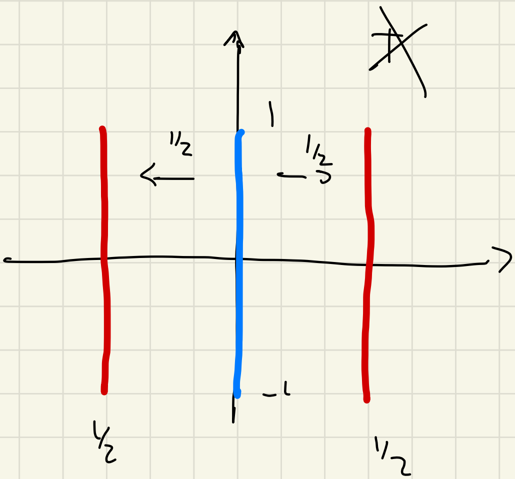
$\nu = \delta_y$

$\mu = \frac{1}{2}\delta_{x_1} + \frac{1}{2}\delta_{x_2}$

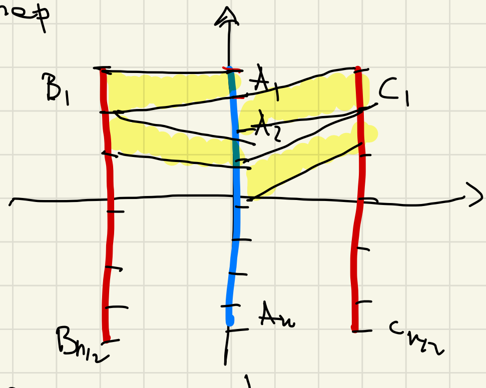
ex 4 (McCann '95)

$\mu = \lambda|_{\{0\} \times [-1,1]}$ $\mathbb{R}^2$

$\nu = \frac{1}{2}\lambda|_{\{\pm 1\} \times [-1,1]}$ $\mathbb{R}^2$



one can have a suboptimal map



$n \uparrow +\infty$

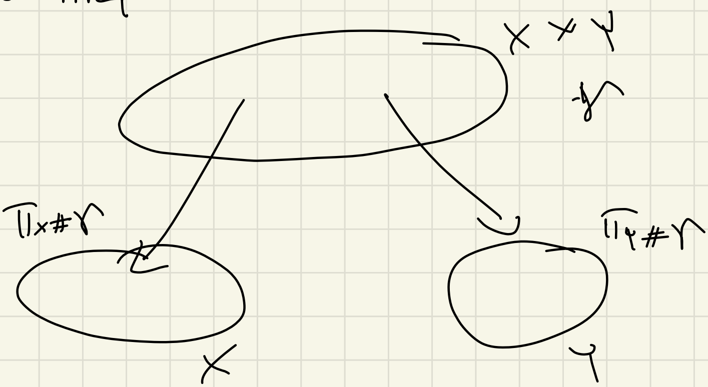
$c = 1 + \frac{1}{n}$

one can do better

KANTOROVICH PROBLEM

def (Marginals) the marginals of a measure γ 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 their projections map

$\Pi_X(x,y) = x$



(transference)

definition (Transport plan) A transport plan between $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ is a prob. measure γ on $X \times Y$ whose marginals are μ and ν. The space of transport plans is denoted $\Pi(\mu, \nu)$

$$\Pi(\mu,\nu) := \left\{ \gamma \in \mathcal{P}(X \times Y) \mid \Pi_X \# \gamma = \mu, \Pi_Y \# \gamma = \nu \right\}$$

exe $\Pi(\mu,\nu)$ is convex!

Rmk 1 $\Pi(\mu,\nu)$ is non-empty $\mu\otimes\nu \in \Pi(\mu,\nu)$

def (transport plan associated to a map)

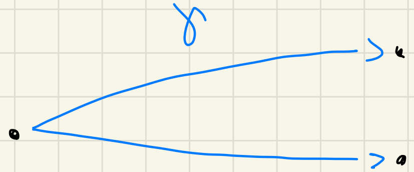
Let T be the transport map between μ and ν and define $\gamma_T = (Id, T)\#\mu$

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

def (Kantorovich problem) Given $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ and a cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ KANTOROVICH pb is the following optimization pb.

$$(KP) \quad \inf \left\{ \int_{X\times Y} c(x,y)\, d\gamma(x,y) \;\middle|\; \gamma \in \Pi(\mu,\nu) \right\}$$

Rmk 2 Mass splitting is allowed



$\nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$

$\mu = \delta_x$

discrete version.

$\min cx$

$Ax = b$

$x \geq 0$

LP

Rmk $(KP) \leq (MP)$

Consider T s.t $T\#\mu = \nu$ and the associated plan γ_T, γ^* opt sol for (KP)

$$\int c\, d\gamma^* \leq \int c\, d\gamma_T = \int_X c(x, T(x))\, d\mu(x)$$

def (Support) The support of a non-negative measure μ is the smallest closed set on which μ is concentrated

$$spt(\mu) := \bigcap \{ A \subseteq X \mid A \text{ closed and } \mu(X \setminus A) = 0 \}$$

proposition let $\gamma \in \Pi(\mu, \nu)$ and $T : X \to Y$ measurable be s.t $\gamma(\{(x,y) \in X \times Y \mid T(x) \neq y\}) = 0$

then $\gamma = \gamma_T$

• EXISTENCE OF SOLUTIONS TO (KP)

<u>thm</u> Let X and Y be two compact spaces and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a l.s.c cost function, which is bounded from below then the (KP) admits a minimizer

<u>Lemma</u> Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a l.s.c. function bounded from below. Define

$$\mathcal{J} : \mathcal{P}(X) \to \mathbb{R} \cup \{+\infty\} \quad \mathcal{J}(\mu) = \int_X f \, d\mu$$

then $\mathcal{J}$ is l.s.c. for the narrow convergence ie,

$$\forall \mu_n \to \mu \quad \liminf_{n \uparrow +\infty} \mathcal{J}(\mu_n) \geq \mathcal{J}(\mu)$$

<u>Rmk</u> Lemma $\Rightarrow$ $\int c \, d\gamma$ is l.s.c. for the $(\rightharpoonup)$

proof (Lemma)

Step 1. We show $\exists$ a family of bounded and continuous f^k s.t. $f^k \nearrow$ is pointwise increasing and $f = \sup_k f^k$

step 2: $\mathcal{G}^k(\mu) = \int f^k d\mu$. Since f^k is cont and bounded the linear form $\mathcal{G}^k$ is narrowly cont. Thus $\mathcal{G} = \sup_k \mathcal{G}^k$ is l.s.c. as a ~~sup~~ sup of l.s.c. functions

We have to prove $\boxed{\text{step 1}}$: We assume $\exists x_0$ s.t. $f(x_0) < +\infty$

Define $g^k(x) = \inf_y f(y) + k\, d(x,y)$

$$\leq f(x_0) + k\, d(x, x_0)$$

$\rightarrow g^k$ is k-lip as minimum of k-lip functions

$\rightarrow g^k \leq g^l \leq f$ for $k \leq l$

~~We~~ let's prove that $\sup_k g^k(x) = f(x)$ $\forall x \in X$

Given x for every k $\exists x_k$ s.t

$$f(x_k) + k\, d(x, x_k) \leq g^k(x) + \frac{1}{k} \qquad (1)$$

$$\leq f(x) + \frac{1}{k}$$

Using that $f \geq M > -\infty$

$$d(x, x_n) \leq \frac{1}{k}\left(f(x) + \frac{1}{k} - M\right)$$

$\Rightarrow \quad x_k \to x$

Taking the limit in (1) and l.s.l. of f we get $f(x) \leq \liminf f(x_k) \leq \sup_{k\to\infty} g^k(x)$

$f^k(x) = \min(g^k(x), k)$ we have that f^k are k-lip., bounded by k and $\sup_k f^k = f$

$\boxtimes$

proof (thm)

Define $J(\gamma) = \int c \, d\gamma$ which l.s.c for the lemma.

We have to show that $\Pi(\mu, \nu)$ is compact for the narrow convergence

Take a sequence $\gamma_n \in \Pi(\mu, \nu)$

since they are probability measures they are bounded in the dual of $C(X \times Y)$

Usual weak-* compactness (B-A)

guarantees the existence of a converging

subsequence $\gamma_{n_k} \to \gamma \in \mathcal{P}(X \times Y)$

We just need to check that $\gamma \in \Pi(\mu, \nu)$

Fix $\varphi \in \mathcal{C}(X)$ then $\int \varphi(x) \, d\gamma_{n_k} = \int \varphi \, d\mu$

$$\begin{array}{ccc} \int \varphi(x) \, d\gamma_{n_k} & = & \int \varphi \, d\mu \\ \downarrow & & \downarrow \\ \int \varphi(x) \, d\gamma & = & \int \varphi \, d\mu \end{array}$$

and same for the second marginal.

• KANTOROVICH AS A RELAXATION OF MONGE

thm Let X and Y be compact set $(X = Y)$ of $\mathbb{R}^d$, $c \in \mathcal{C}(X \times Y)$ and $\mu \in \mathcal{P}(X)$ $\nu \in \mathcal{P}(Y)$

Assume that μ is atomless. then

$$\inf (MP) = \min (KP)$$

Lemma if $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and μ has no atoms then $\exists T : \mathbb{R}^d \to \mathbb{R}^d$ measurable s.t

$T_\# \mu = \nu$ (proof in Santambrogio)

Lemma Let K be a compact metric space
For any $\varepsilon > 0$ $\exists$ a meas. partition
$K_1 \dots K_N$ of K s.t. $\forall i$ $\operatorname{diam}(K_i) \leq \varepsilon$

proof (thm) Using the continuity of $\int c\, d\gamma$
the statement will follow if we prove that
for any, $\gamma \in \Pi(\mu, \nu)$ $\exists$ a sequence of transport
maps $T^n : X \to Y$ s.t. $T^n_{\#}\mu = \nu$
and $\gamma_{T^n} \to \gamma$
(the set of transport plans γ_T is dense in $\Pi(\mu,\nu)$)

$$\int c\, d\gamma = \lim \int c\, d\gamma^n_T$$

$\exists$ a partition $K_1, .., K_N$ of X s.t.
$\operatorname{diam}(K_i) \leq \varepsilon$ $X = Y$
Define $\gamma_i := \gamma \big|_{K_i \times \cancel{X}}$ Now let

$\mu_i = \pi_{X\#}\gamma_i$ and $\underline{\nu_i = \pi_{Y\#}\gamma_i}$

Since $\mu_i \leq \mu$ so μ_i has no atoms

$\Rightarrow \exists S_i \quad S_i \# \mu_i = \nu_i$

by gluing the transports S_i we get a transport T^n sending μ onto ν (we use $\mu = \sum \mu_i \quad \nu = \sum \nu_i$) as the μ_i are concentrated on disjoint sets

Since $\gamma_{S_i} = (Id, S_i)\#\mu_i$ and γ_i have marginals μ_i and ν_i, one has

$\gamma_{S_i}(K_i \times K_j) = \nu_i(K_j) = \gamma_i(K_i \times K_j)$.

We have to prove $\gamma_{T^n} \to \gamma$

$\varphi \in C_b(X \times Y)$ by compactness of $X \times Y$ φ has a uniform continuity modulus ω_φ w.r.t each norm

$$\int \varphi \, d(\gamma - \gamma_{T^n}) = \sum_{ij} \int_{K_i \times K_j} \varphi \, d(\gamma_i - \gamma_{S_i})$$

$$\leq \sum_{ij} \left(\gamma_i (K_i \times K_j) \max \varphi - \gamma_{S_i}(K_i \times K_j) \min \varphi \right)$$

$$\leq \sum_{ij} \gamma_i (K_i \times K_j)\, w_\varphi(\mathrm{diam}(K_i \times K_j)) < \Theta(w_\varphi(2\varepsilon))$$

this holds for ~~each~~ every $\varphi \in C_b(X \times Y)$

$\Rightarrow \gamma_{T^n} \rightharpoonup \gamma$

In particular if γ is the minimizer of (KP) then γ_{T^n} a minimizing sequence

$$\lim_{n\uparrow+\infty} \int_X c(x, T^n(x))\, d\mu(x)$$

$$= \lim_{n\uparrow+\infty} \int_{X\times Y} c(x,y)\, d\gamma_{T^n}$$

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

$\boxtimes$

DUAL PROBLEM

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

$$\underbrace{\sup_{\varphi,\psi \in C_b(X)\times C_b(Y)} \int_X \varphi \, d\mu + \int_Y \psi \, d\nu - \int_{X\times Y} (\varphi + \psi)\, d\gamma}_{\delta(\gamma)} = \begin{cases} 0 & \text{if } \gamma \in \Pi(\mu,\nu) \\ +\infty & \text{othwise} \end{cases}$$

we can now remove the constraint in (KP)

$$\inf_{\gamma \in \mathcal{M}^+(X\times Y)} \int c \, d\gamma + \delta(\gamma)$$

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

by interchanging inf and sup we get

$$= \sup_{\varphi,\psi} \int \varphi\, d\mu + \int \psi\, d\nu + \underbrace{\inf_{\gamma} \int (c - \varphi - \psi)\, d\gamma}$$

$\geq$ ~~$\not>$~~

$$\inf_\gamma \int (c-\varphi-\psi)\,d\gamma = \begin{cases} 0 & \text{if } \varphi+\psi\le c \text{ on } X\times Y \\ -\infty & \text{otherwise} \end{cases}$$

$$(*) = \sup\left\{ \int \varphi\,d\mu + \int \psi\,d\nu \;\middle|\; \begin{array}{l} \varphi,\psi \in C_b(X)\times C_b(Y) \\ \varphi(x)+\psi(y)\le c(x,y) \\ \forall x,y \end{array} \right\}$$

$$(DP)$$

<u>def</u> (Dual pb) Given $\mu\in\mathcal{P}(X)$ and $\nu\in\mathcal{P}(Y)$ and a cost function c, the dual problem is (DP)

<u>Rmk</u> (weak duality)

$$\sup(DP) \le \inf(KP)$$

$$=$$

$\varphi^*, \psi^*, \gamma^*$

$$\int \varphi^* d\mu + \int \psi^* d\nu = \int (\varphi^* + \psi^*)\,d\gamma^*$$

$$\le \int c\,d\gamma^*$$

<u>Rmk</u> the optimal sol to (DP) are called KANTOROVICH POTENTIALS!