Optimal Transport

- Xavier Fernández-Real
- Roberto Colombo

Week 1

References:

A. Figalli, F. Glaudo, "An invitation to optimal transport, Wasserstein distances, and gradient flows".

F. Santambrogio, "Optimal Transport for Applied Mathematicians: ..."

L. Ambrosio, E. Brué, D. Semola, "Lectures on Optimal Transport".

- Exercise classes:
  - → Present a solution from the serie (by a student!)
  - → Discuss one exercise.
  - → Answer questions.

- Hand-in series
  - → 2 hand-in series.
  - → One random exercise to be corrected.

- Exam: → Two parts; 4-6 topics to choose from a list.

  1st: 30' for exercise + topic preparation

  2nd: 30' presentation of the assigned topic

  ↳ statement, examples, ideas, proofs, etc.

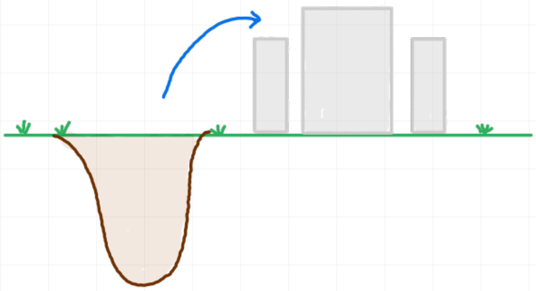
- Grade: Grade of the exam + up to 0.5 from hand-ins.

  + up to 0.25 from exercise presentations.

  (exercises to be presented assigned beforehand).

☐ The Optimal Transport problem and its origin

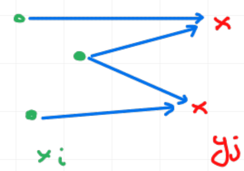
1781 Gaspar Monge: "assume we want to extract soil from the ground to build fortifications. How do we transport the soil in the cheapest way?



Data: • initial and target mass

Unknown: • transport map.

1940 Leonid Kantorovic: allocation of resources.

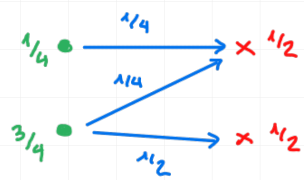


- Bakeries in $x_i \in \mathbb{R}^2$ produce $\alpha_i \geq 0$ of bread.
- Coffee shops in $y_j \in \mathbb{R}^2$ consume $\beta_j \geq 0$ of bread.
- Assume $\sum \alpha_i = \sum \beta_j$, and let γ_{ij} the bread sold from x_i to y_j, to satisfy the demand:

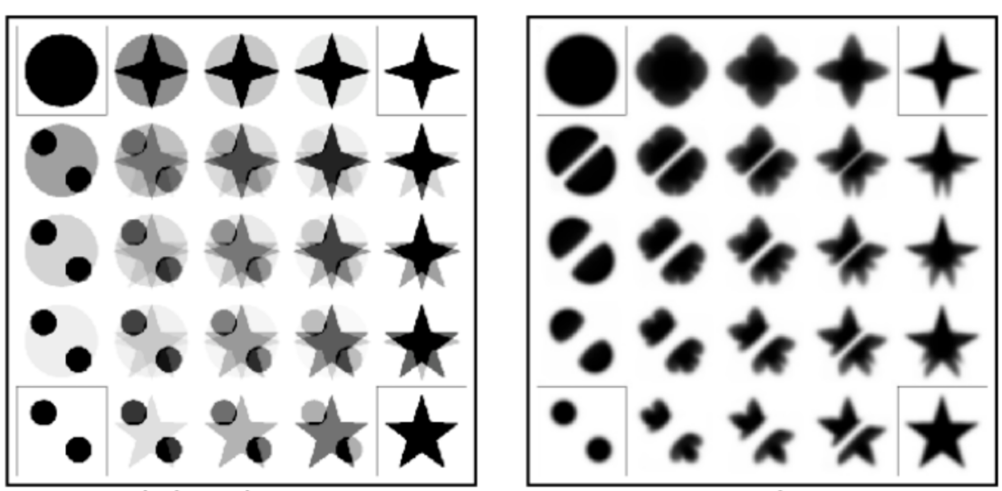
(*) $\alpha_i = \sum_j \gamma_{ij}$

(**) $\beta_j = \sum_i \gamma_{ij}$

Example:



We want: $\min \left\{ \sum_{i,j} \gamma_{ij}\, c(x_i, y_j) \text{ among all } (\gamma_{ij})_{ij} \text{ admissible: } (*) \,\&\, (**) \right\}$.



Euclidean barycenter Wasserstein barycenter

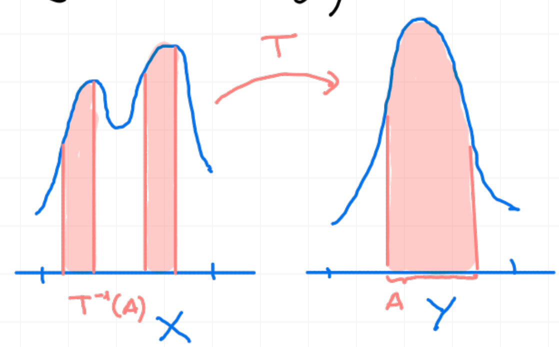
☐ Pushforward of measures and transport maps

- X complete and separable metric space (usually $\mathbb{R}^d$)
- $\mathcal{P}(X)$: probability measures on X.

<u>Def</u> : Given $\mu \in \mathcal{P}(X)$, $T: X \longrightarrow Y$, the <u>image measure of μ through T</u> (<u>pushforward</u>) is $\nu = T_\# \mu$ s.t. :

$$(*) \quad \boxed{\nu(A) = \mu(T^{-1}(A))}$$

for all $A \subseteq Y$ measurable.



<u>Lemma</u> : $T_\# \mu$ is a probability measure. $\left(\text{Use } T^{-1}\left(\bigcup_i A_i\right) = \bigcup_i T^{-1}(A_i)\right)$.

<u>Remark</u> : $(*)$ can be written as:

$$\int_Y \overbrace{\mathbb{1}_A}^{\text{characteristic/indicator function.}} d\nu = \int_X \mathbb{1}_{T^{-1}(A)} d\mu = \int_X \mathbb{1}_A \circ T \, d\mu$$

Hence, if φ is a simple function : $\boxed{\int_Y \varphi \, d\nu = \int_X \varphi \circ T \, d\mu}$ $(**)$.

By monotone convergence, same holds for φ measurable and nonnegative (bounded).

$(*)$ and $(**)$ are equivalent!

<u>Corollary</u> : $\nu = T_\# \mu \iff \int \varphi \, d\nu = \int \varphi \circ T \, d\mu$ for all φ measurable and bounded.

<u>Def</u> : Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, $T: X \longrightarrow Y$ is a <u>transport map</u> between μ and ν if $T_\# \mu = \nu$.

<u>Exercises</u> : are the following true?

1. $T(x) = 2x+1 : \mathbb{R} \to \mathbb{R}$, $\quad T_\# \delta_0 = \delta_1$?
2. $T(x) \equiv 1 : \mathbb{R} \to \mathbb{R}$, $\quad T_\# \mathcal{L}_{[-1,1]} = \delta_1$?
3. Given $\mu, \nu \in \mathcal{P}(X)$, is there always a map T s.t. $T_\# \mu = \nu$?

Remark: Given $T: X \to Y$ & $S: Y \to Z$ measurable,

$$(S \circ T)_{\#}\mu = S_{\#} T_{\#} \mu .$$

Remark: Assume $X = Y = \mathbb{R}^n$, $T: \mathbb{R}^n \to \mathbb{R}^n$ diffeomorphism,

$\mu = f(x)dx$, $\nu = g(y)dy$. If $T_{\#}\mu = \nu$, then:

$$\int \varphi \circ T(x) f(x)\, dx = \int \varphi(y) g(y)\, dy \overset{y = T(x)}{=} \int \varphi(T(x))\, g(T(x))\, |\det DT(x)|\, dx$$

i.e.
$$\boxed{\frac{f(x)}{g(T(x))} = |\det DT(x)|} \qquad (\bullet)$$

- Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, $c: X \times Y \longrightarrow \mathbb{R}$, Monge's minimization problem is:

$$\boxed{(MP) := \inf \left\{ \int c(x, T(x))\, d\mu(x) \ : \ T_{\#}\mu = \nu \right\}}$$

Def: $\gamma \in \mathcal{P}(X \times Y)$ is a coupling (transport plan) of μ and ν if

$\pi_X \# \gamma = \mu$ $\qquad$ $\pi_Y \# \gamma = \nu$

$\pi_X : X \times Y \to X$ $\qquad$ $\pi_Y : X \times Y \to Y$

$(x,y) \mapsto x$ $\qquad$ $(x,y) \mapsto y$

- The Kantorovich minimization problem is:

$$\boxed{(KP) := \inf \left\{ \int c(x,y)\, d\gamma(x,y) \ : \ \gamma \text{ coupling between } \mu \text{ \& } \nu \right\}}$$

Remark: Given μ, ν, $\gamma := \mu \otimes \nu$ is a coupling!

Remark: $(KP) \leq (MP)$.

Main Theorem (Brenier 1987). $X = Y = \mathbb{R}^n$, $c(x,y) = \frac{|x-y|^2}{2}$.

Let $\mu = f(x)dx \in \mathcal{P}(\mathbb{R}^n)$, $\nu = g(y)dy \in \mathcal{P}(\mathbb{R}^n)$. Then, there exists a unique optimal map T in (MP). Moreover, $T = \nabla\varphi$, φ convex, $(Id, T)_\# \mu$ is the optimizer in (KP).

Proof: will be completed around week 7.

Application isoperimetric inequality:

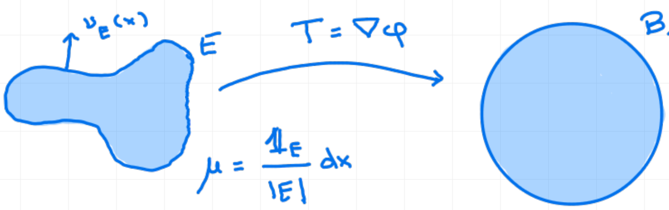
$E \subseteq \mathbb{R}^n$ smooth, open, bounded set.

$P(E) = \text{perimeter of } E = \int_{\partial E} 1$

$|E| = \text{volume of } E = \int_E 1$

Then, $P(E) \geq n |B_1|^{\frac{1}{n}} |E|^{\frac{n-1}{n}}$

Proof:



$\mu = \frac{\mathbb{1}_E}{|E|}dx$

$\nu = \frac{\mathbb{1}_{B_1}}{|B_1|}dy$ $\quad$ ($\mathbb{1}_A(x)$ characteristic function $\begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$) $\quad$ $(|\partial B_1| = n|B_1|)$

By Brenier's theorem, $\exists T = \nabla\varphi$, φ convex, s.t. $T_\#\mu = \nu$. Properties of T:

1. $|T(x)| \leq 1$ a.e. $x \in E$
2. $0 \underset{\varphi \text{ is convex}}{\leq} \det DT(x) = \frac{|B_1|}{|E|}$ a.e. $x \in E$ by $(\bullet)$
3. $\operatorname{div} T = \sum_i \partial_i T^i = \sum_i \partial_i\partial_i\varphi = \Delta\varphi = \operatorname{tr} D^2\varphi = \sum_{i=1}^n \lambda_i = n\left(\frac{1}{n}\sum_{i=1}^n \lambda_i\right) \overset{\text{AM-GM}}{\geq} n\left[\prod_{i=1}^n \lambda_i\right]^{1/n} = n(\det DT)^{1/n}$ (λ_i: eigenvalues of $D^2\varphi$, $\lambda_i \geq 0$)

Hence: $P(E) = \int_{\partial E} 1 \overset{①}{\geq} \int_{\partial E} |T(x)| \geq \int_{\partial E} T(x)\cdot\nu_E(x) \overset{\text{divergence thm./ Stokes}}{=} \int_E \operatorname{div} T(x)\,dx \geq$

$\overset{③}{\geq} \int_E n[\det DT]^{1/n} \overset{②}{=} \int_E n\left(\frac{|B_1|}{|E|}\right)^{1/n} = n|B_1|^{\frac{1}{n}}|E|^{\frac{n-1}{n}}$. $\square$

we are using $T \in C^1$.

Equality: all $\geq$ are $=$ $\longrightarrow$ $\overbrace{\lambda_1(x) = \dots = \lambda_n(x)}^{\text{AM-GM}} \overset{②}{=} \left(\frac{|B_1|}{|E|}\right)^{1/n}$.

Hence, $D^2\varphi = \left(\frac{|B_1|}{|E|}\right)^{1/n} Id \longrightarrow \nabla\varphi(x) = T(x) = c_0 x + c_1$,

$c_0 = \left(\frac{|B_1|}{|E|}\right)^{1/n} = 1$ if they have the same volume.

☐ One-dimensional Optimal Transport $X = Y = \mathbb{R}$

<u>Theorem</u>: $\mu, \nu \in \mathcal{P}(X)$, μ without atoms (i.e., $\mu(\{x\}) = 0 \ \forall x \in \mathbb{R}$). Then:

① $\exists!$ non-decreasing map T s.t. $T_\# \mu = \nu$: the <u>MONOTONE REARRANGEMENT</u>.

② if $c(x,y) = \varphi(|x-y|)$, φ non-decreasing $\to$ T optimal in (MP).

③ φ strictly convex $\Rightarrow$ T is the unique optimal map.

<u>Proof</u>: ② & ③ will be consequences of the theorems in the course ($\varphi(t) = t^2$ is Brenier's).

↳ proof of ①:

Let us consider the repartition functions (cumulative distr.): $\begin{cases} F_\mu(x) := \mu((-\infty, x]) \\ G_\nu(y) := \nu((-\infty, y]). \end{cases}$

Observe that they are non-decreasing, and F_μ is continuous (not G_ν!) $\left[|F_\mu(t_k) - F_\mu(t)| = \left| \int_t^{t_k} d\mu \right| \xrightarrow[t_k \to t]{} 0 \right]$.

We have defined G (and F) so that it is continuous from the right. $\left(G_\nu(y) = \lim_{\varepsilon \downarrow 0} \nu((-\infty, y+\varepsilon]) \right)$

Define the pseudo-inverse $G^{-1}(y) := \inf\{t \in \mathbb{R} : G(t) > y\}$

(G may have constant regions). G^{-1} is continuous from the right.

<u>Claim</u>: $T_\# \mu = \nu$, where we define $T := G^{-1} \circ F$.

└ proof:
- F is continuous
- $F(t) \to 0$ as $t \to -\infty$, $F(t) \to 1$ as $t \to \infty$.
- F surjective in $(0,1)$.

Given $t \in (0,1)$, let x be the largest value s.t. $F(x) = t$:

$$\mu(F^{-1}[0,t]) = \int_{F^{-1}([0,t])} d\mu = \int_{-\infty}^{x} d\mu = F(x) = t.$$

Notice, also, that: $\underset{\overset{\parallel}{t}}{\mu(F^{-1}[0,t))} \geq \mu(F^{-1}[0, t-\varepsilon]) = t - \varepsilon$

Let now $A = (-\infty, a]$, $a \in \mathbb{R}$. Then:

$$T_\#\mu(A) = \mu(T^{-1}(A)) = \mu(F^{-1} \circ \underbrace{(G^{-1})^{-1}((-\infty, a]))}_{(-\infty, G(a)] \text{ or } (-\infty, G(a)) \text{ depending on the point}} =$$

$$= \mu(F^{-1}(-\infty, G(a)]) = G(a)$$

$= \nu(-\infty, a] = \nu(A)$. Use properties of measures for all Borel sets.