Introduction to optimal transport

Nicola Gigli

December 4, 2023

Content

- Formulation of the transport problem
- The notions of c-convexity and c-cyclical monotonicity
- The dual problem
- Optimal maps: Brenier's theorem

Content

- Formulation of the transport problem
- The notions of c-convexity and c-cyclical monotonicity
- The dual problem
- Optimal maps: Brenier's theorem

The setting

We shall work on Polish spaces, i.e. topological spaces which are metrizable by a complete and separable distance.

Given such space X, by $\mathscr{P}(X)$ we mean the space of Borel probability measures on X.

It is perfectly fine to consider just the case $X = \mathbb{R}^d$.

A notion: the push forward

Let X, Y be Polish spaces, $\mu \in \mathscr{P}(\mathsf{X})$ and $T : \mathsf{X} \to \mathsf{Y}$ a Borel map.

The measure $T_*\mu \in \mathscr{P}(\mathsf{Y})$ is defined by

$$T_*\mu(A) := \mu(T^{-1}(A)), \qquad \text{for every Borel set } A \subset \mathsf{Y}$$

The measure $T_*\mu$ is characterized by

$$\int f \, \mathrm{d} T_*\mu = \int f \circ T \, \mathrm{d}\mu,$$

for any Borel function $f : \mathsf{Y} \to \mathbb{R}$.

Monge's formulation of the transport problem

Let $\mu \in \mathscr{P}(\mathsf{X})$, $\nu \in \mathscr{P}(\mathsf{Y})$ be given, and let $c : \mathsf{X} \times \mathsf{Y} \to \mathbb{R}$ be a *cost* function, say continuous and non-negative.

Problem: Minimize

$$\int c(x, T(x))\,\mathrm{d}\mu(x),$$

among all *transport maps* from μ to ν, i.e., among all maps T such that $T_*\mu = \nu$

Why this is a bad formulation

There are several issues with this formulation:

- it may be that no transport map exists at all (eg., if μ is a Delta and ν is not)
- the constraint $T_*\mu = \nu$ is not closed w.r.t. any reasonable weak topology

Kantorovich's formulation

A measure $\gamma \in \mathscr{P}(X \times Y)$ is a *transport plan* from μ to ν if

$$\pi^1_* \gamma = \mu,$$
$$\pi^2_* \gamma = \nu.$$

Problem Minimize

$$\int c(x,y)\,\mathrm{d}\gamma(x,y),$$

among all transport plans from μ to ν.

Why this is a good formulation

- There always exists at least one transport plan: $\mu \times \nu$,
- Transport plans 'include' transport maps: if $T_*\mu = \nu$, then $(Id, T)_*\mu$ is a transport plan
- The set of transport plans is compact w.r.t. the weak topology of measures.
- The map $\gamma \mapsto \int c(x,y)\,\mathrm{d}\gamma(x,y)$ is linear and weakly lower semi-continuous,

Why this is a good formulation

- There always exists at least one transport plan: $\mu \times \nu$,
- Transport plans 'include' transport maps: if $T_* \mu = \nu$, then $(Id, T)_* \mu$ is a transport plan
- The set of transport plans is compact w.r.t. the weak topology of measures.
- The map $\gamma \mapsto \int c(x, y) \, \mathrm{d}\gamma(x, y)$ is linear and weakly lower semi-continuous,

In particular, minima exist.

Now what?

What can we say about optimal plans?

In particular:

- Do they have any particular structure? If so, which one?
- Are they unique?
- Are they induced by maps?

Content

- Formulation of the transport problem
- The notions of c-cyclical monotonicity and c-convexity
- The dual problem
- Optimal maps: Brenier's theorem

A key example

Let $\{x_i\}_i$, $\{y_i\}_i$, $i = 1, \ldots, N$ be points in X and Y respectively

$$\mu := \frac{1}{N} \sum_i \delta_{x_i},$$

$$\nu := \frac{1}{N} \sum_i \delta_{y_i}.$$

A key example

Let $\{x_i\}_i$, $\{y_i\}_i$, $i = 1, \ldots, N$ be points in X and Y respectively

$$\mu := \frac{1}{N} \sum_i \delta_{x_i},$$

$$\nu := \frac{1}{N} \sum_i \delta_{y_i}.$$

Then a plan γ is optimal iff for any $n \in \mathbb{N}$, permutation σ of $\{1, \ldots, n\}$ and any $\{(x_k, y_k)\}_{k=1,\ldots,n} \subset \mathrm{supp}(\gamma)$ it holds

$$\sum_k c(x_k, y_k) \leq \sum_k c(x_k, y_{\sigma(k)})$$

The general definition

We say that a set $\Gamma \subset X \times Y$ is *c-cyclically monotone* if for any $n \in \mathbb{N}$, permutation σ of $\{1, \ldots, n\}$ and any $\{(x_k, y_k)\}_{k=1,\ldots,n} \subset \Gamma$ it holds

$$\sum_k c(x_k, y_k) \leq \sum_k c(x_k, y_{\sigma(k)})$$

First structural theorem

Theorem A transport plan γ is optimal if and only if its support $\mathrm{supp}(\gamma)$ is c-cyclically monotone.

First structural theorem

Theorem A transport plan γ is optimal if and only if its support $\mathrm{supp}(\gamma)$ is c-cyclically monotone.

In particular, being optimal depends only on the support of γ, and not on how the mass is distributed on the support (!).

Content

- Formulation of the transport problem
- The notions of c-cyclical monotonicity and c-convexity
- The dual problem
- Optimal maps: Brenier's theorem

The dual formulation

Given the measures $\mu \in \mathscr{P}(\mathsf{X})$, $\nu \in \mathscr{P}(\mathsf{Y})$ and the cost function $c : \mathsf{X} \times \mathsf{Y} \to \mathbb{R}$, maximize

$$\int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu,$$

among all couples of functions $\varphi : \mathsf{X} \to \mathbb{R}$, $\psi : \mathsf{Y} \to \mathbb{R}$ continuous and bounded such that

$$\varphi(x) + \psi(y) \leq c(x, y), \qquad \forall x \in \mathsf{X},\ y \in \mathsf{Y}.$$

We call such a couple of functions *admissible potentials*

A simple inequality

Let γ be a transport plan from μ to ν and (φ, ψ) admissible potentials. Then

$$\begin{aligned}\int c(x,y)\,\mathrm{d}\gamma(x,y) &\geq \int \varphi(x) + \psi(y)\,\mathrm{d}\gamma(x,y) \\ &= \int \varphi(x)\,\mathrm{d}\mu(x) + \int \psi(y)\,\mathrm{d}\nu(y).\end{aligned}$$

Thus

$$\inf\{\text{transport problem}\} \geq \sup\{\text{dual problem}\}$$

A property of admissible potentials

Say that (φ, ψ) are admissible potentials and define

$$\varphi^c(y) := \inf_x c(x, y) - \varphi(x).$$

Then $\varphi^c \geq \psi$ and (φ, φ^c) are admissible as well.

A property of admissible potentials

Say that (φ, ψ) are admissible potentials and define

$$\varphi^c(y) := \inf_x c(x, y) - \varphi(x).$$

Then $\varphi^c \geq \psi$ and (φ, φ^c) are admissible as well.

Similarly, we can define

$$\psi^c(x) := \inf_y c(x, y) - \psi(y),$$

so that $\psi^c \geq \varphi$ and (ψ^c, ψ) are admissible

The process stabilizes

Starting from (φ, ψ), we can consider the admissible potentials (φ, φ^c), $(\varphi^{cc}, \varphi^c)$, $(\varphi^{cc}, \varphi^{ccc})$...

The process stabilizes

Starting from (φ, ψ), we can consider the admissible potentials (φ, φ^c), $(\varphi^{cc}, \varphi^c)$, $(\varphi^{cc}, \varphi^{ccc})$...

This process stops, because $\varphi^{ccc} = \varphi^c$.

The process stabilizes

Starting from (φ, ψ), we can consider the admissible potentials (φ, φ^c), $(\varphi^{cc}, \varphi^c)$, $(\varphi^{cc}, \varphi^{ccc})$...

This process stops, because $\varphi^{ccc} = \varphi^c$. Indeed

$$\varphi^{ccc}(y) = \inf_x \sup_{\tilde{y}} \inf_{\tilde{x}} c(x, y) - c(x, \tilde{y}) + c(\tilde{x}, \tilde{y}) - \varphi(\tilde{x}),$$

and picking $\tilde{x} = x$ we get $\varphi^{ccc} \leq \varphi^c$, and picking $\tilde{y} = y$ we get $\varphi^{ccc} \geq \varphi^c$.

c-concavity and c-superdifferential

A function φ is c-concave if $\varphi = \psi^c$ for some function ψ.

The c-superdifferential $\partial^c\varphi \subset X \times Y$ is the set of (x, y) such that

$$\varphi(x) + \varphi^c(y) = c(x, y).$$

Second structural theorem

For any c-concave function φ, the set $\partial^c\varphi$ is c-cyclically monotone, indeed if $\{(x_k, y_k)\}_k \subset \partial^c\varphi$ it holds

$$\begin{aligned}\sum_k c(x_k, y_k) &= \sum_k \varphi(x_k) + \varphi^c(y_k)\\ &= \sum_k \varphi(x_k) + \varphi^c(y_{\sigma(k)})\\ &\leq \sum_k c(x_k, y_{\sigma(k)})\end{aligned}$$

Second structural theorem

For any c-concave function φ, the set $\partial^c\varphi$ is c-cyclically monotone, indeed if $\{(x_k, y_k)\}_k \subset \partial^c\varphi$ it holds

$$\begin{aligned}\sum_k c(x_k, y_k) &= \sum_k \varphi(x_k) + \varphi^c(y_k) \\ &= \sum_k \varphi(x_k) + \varphi^c(y_{\sigma(k)}) \\ &\leq \sum_k c(x_k, y_{\sigma(k)})\end{aligned}$$

Actually much more holds:
Theorem A set Γ is c-cyclically monotone iff $\Gamma \subset \partial^c\varphi$ for some φ c-concave.

To summarize

Given $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ and a cost function *c*, for an admissible plan γ the following three are equivalent:

- γ is optimal
- $\mathrm{supp}(\gamma)$ is *c*-cyclically monotone
- $\mathrm{supp}(\gamma) \subset \partial^c \varphi$ for some *c*-concave function φ

(this requires some minor technical compatibility conditions between μ, ν, c which we neglect here)

No duality gap

It holds

$$\inf\{\text{transport problem}\} = \sup\{\text{dual problem}\}$$

Indeed, if γ is optimal, then $\mathrm{supp}(\gamma) \subset \partial^c \varphi$ for some c-concave φ. Thus

$$\int c(x,y)\,\mathrm{d}\gamma(x,y) = \int \varphi(x) + \varphi^c(y)\,\mathrm{d}\gamma(x,y) = \int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu$$

Content

- Formulation of the transport problem
- The notions of c-cyclical monotonicity and c-convexity
- The dual problem
- Optimal maps: Brenier's theorem

The case $X = Y = \mathbb{R}^d$ and $c(x, y) = |x - y|^2/2$

φ is c-concave iff $\hat{\varphi}(x) := |x|^2/2 - \varphi(x)$ is convex.
Indeed:

$$
\begin{aligned}
& \varphi(x) = \inf_y \frac{|x-y|^2}{2} - \psi(y) \\
\Leftrightarrow \quad & \varphi(x) = \inf_y \frac{|x|^2}{2} + \langle x, -y \rangle + \frac{|y|^2}{2} - \psi(y) \\
\Leftrightarrow \quad & \varphi(x) - \frac{|x|^2}{2} = \inf_y \langle x, -y \rangle + \left(\frac{|y|^2}{2} - \psi(y) \right) \\
\Leftrightarrow \quad & \hat{\varphi}(x) = \sup_y \langle x, y \rangle - \left(\frac{|y|^2}{2} - \psi(y) \right),
\end{aligned}
$$

The case $X = Y = \mathbb{R}^d$ and $c(x, y) = |x - y|^2/2$

c-superdifferential and subdifferential

$$(x, y) \in \partial^c \varphi \text{ iff } y \in \partial^- \hat{\varphi}(x).$$

The case $X = Y = \mathbb{R}^d$ and $c(x, y) = |x - y|^2/2$

c-superdifferential and subdifferential

$(x, y) \in \partial^c \varphi$ iff $y \in \partial^- \hat{\varphi}(x)$.
Indeed:

$$
\begin{aligned}
&(x, y) \in \partial^c \varphi \\
&\Leftrightarrow \begin{cases} \varphi(x) = |x - y|^2/2 - \varphi^c(y), & \\ \varphi(z) \le |z - y|^2/2 - \varphi^c(y), & \forall z \in \mathbb{R}^d \end{cases} \\
&\Leftrightarrow \begin{cases} \varphi(x) - |x|^2/2 = \langle x, -y \rangle + |y|^2/2 - \varphi^c(y), & \\ \varphi(z) - |z|^2/2 \le \langle z, -y \rangle + |y|^2/2 - \varphi^c(y), & \forall z \in \mathbb{R}^d \end{cases} \\
&\Leftrightarrow \varphi(z) - |z|^2/2 \le \varphi(x) - |x|^2/2 + \langle z - x, -y \rangle \qquad \forall z \in \mathbb{R}^d \\
&\Leftrightarrow -y \in \partial^+(\varphi - |\cdot|^2/2)(x) \\
&\Leftrightarrow y \in \partial^- \hat{\varphi}(x)
\end{aligned}
$$

Reminder: differentiability of convex functions

Let $\hat{\varphi} : \mathbb{R}^d \to \mathbb{R}$ be convex.

Then for a.e. x, $\hat{\varphi}$ is differentiable at x. This is the same as to say that for a.e. x the set $\partial^- \hat{\varphi}(x)$ has only one element.

Brenier's theorem: statement

Let $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$. Assume that $\mu \ll \mathcal{L}^d$.

Then:

- there exists a unique transport plan
- this transport plan is induced by a map
- the map is the gradient of a convex function

Brenier's theorem: proof

- Pick an optimal plan γ.

Brenier's theorem: proof

- Pick an optimal plan γ.
- Then $\mathrm{supp}(\gamma) \subset \partial^c \varphi$ for some c-concave function φ.

Brenier's theorem: proof

- Pick an optimal plan γ.
- Then $\operatorname{supp}(\gamma) \subset \partial^c \varphi$ for some c-concave function φ.
- Then $\operatorname{supp}(\gamma) \subset \partial^- \hat{\varphi}$ for some convex function $\hat{\varphi}$.

Brenier's theorem: proof

- Pick an optimal plan γ.
- Then $\mathrm{supp}(\gamma) \subset \partial^c \varphi$ for some c-concave function φ.
- Then $\mathrm{supp}(\gamma) \subset \partial^- \hat{\varphi}$ for some convex function $\hat{\varphi}$.
- Thus for γ-a.e. (x, y) it holds $y \in \partial^- \hat{\varphi}(x)$.

Brenier's theorem: proof

- Pick an optimal plan γ.
- Then $\mathrm{supp}(\gamma) \subset \partial^c \varphi$ for some c-concave function φ.
- Then $\mathrm{supp}(\gamma) \subset \partial^- \hat{\varphi}$ for some convex function $\hat{\varphi}$.
- Thus for γ-a.e. (x, y) it holds $y \in \partial^- \hat{\varphi}(x)$.
- Therefore for μ-a.e. x there is only one y such that $(x, y) \in \mathrm{supp}(\gamma)$, and this y is given by $y := \nabla \hat{\varphi}(x)$.

Brenier's theorem: proof

- Pick an optimal plan γ.
- Then $\mathrm{supp}(\gamma) \subset \partial^c \varphi$ for some c-concave function φ.
- Then $\mathrm{supp}(\gamma) \subset \partial^- \hat{\varphi}$ for some convex function $\hat{\varphi}$.
- Thus for γ-a.e. (x, y) it holds $y \in \partial^- \hat{\varphi}(x)$.
- Therefore for μ-a.e. x there is only one y such that $(x, y) \in \mathrm{supp}(\gamma)$, and this y is given by $y := \nabla \hat{\varphi}(x)$.
- This is the same as to say that $\gamma = (Id, \nabla \hat{\varphi})_* \mu$.

Brenier's theorem: proof

- Pick an optimal plan γ.
- Then $\operatorname{supp}(\gamma) \subset \partial^c \varphi$ for some c-concave function φ.
- Then $\operatorname{supp}(\gamma) \subset \partial^- \hat{\varphi}$ for some convex function $\hat{\varphi}$.
- Thus for γ-a.e. (x, y) it holds $y \in \partial^- \hat{\varphi}(x)$.
- Therefore for μ-a.e. x there is only one y such that $(x, y) \in \operatorname{supp}(\gamma)$, and this y is given by $y := \nabla \hat{\varphi}(x)$.
- This is the same as to say that $\gamma = (Id, \nabla \hat{\varphi})_* \mu$.
- But this is true for any optimal plan, thus if $\tilde{\gamma}$ is another optimal plan we must also have $\tilde{\gamma} = (Id, \nabla \hat{\varphi})_* \mu$ and therefore $\tilde{\gamma} = \gamma$

Thank you