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 $\mathcal{P}(X)$ 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, $\mu \in \mathcal{P}(X)$ and $T : X \to Y$ a Borel map.

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

$$T_*\mu(A) := \mu(T^{-1}(A)),$$

for every Borel set $A \subset Y$.

The measure $T_*\mu$ is characterized by

$$\int f \, dT_*\mu = \int f \circ T \, d\mu,$$

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

Monge’s formulation of the transport problem

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

**Problem:** Minimize

\[
\int c(x, T(x)) \, d\mu(x),
\]

among all transport maps from \( \mu \) to \( \nu \), 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 (e.g., if $\mu$ is a Delta and $\nu$ is not)

- the constraint $T_*\mu = \nu$ is not closed w.r.t. any reasonable weak topology
Kantorovich’s formulation

A measure $\gamma \in \mathcal{P}(X \times Y)$ is a transport plan from $\mu$ to $\nu$ if

$$\pi_1^* \gamma = \mu, \quad \pi_2^* \gamma = \nu.$$ 

Problem Minimize

$$\int c(x, y) \, d\gamma(x, y),$$

among all transport plans from $\mu$ to $\nu$. 
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 $(\text{Id}, T)_*\mu$ is a transport plan

- The set of transport plans is closed w.r.t. the weak topology of measures.

- The map $\gamma \mapsto \int c(x, y) \, d\gamma(x, y)$ is linear and weakly 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 $(\text{Id}, T)_* \mu$ is a transport plan.
- The set of transport plans is closed w.r.t. the weak topology of measures.
- The map $\gamma \mapsto \int c(x, y) \, d\gamma(x, y)$ is linear and weakly 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 \( \gamma \) is optimal iff for any \( n \in \mathbb{N} \), permutation \( \sigma \) of \( \{1, \ldots, n\} \) and any \( \{(x_k, y_k)\}_{k=1,\ldots,n} \subset \text{supp}(\gamma) \) 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 $\sigma$ 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 \( \gamma \) is optimal if and only if its support \( \text{supp}(\gamma) \) is \( c \)-cyclically monotone.
Theorem A transport plan $\gamma$ is optimal if and only if its support $\text{supp}(\gamma)$ is $c$-cyclically monotone.

In particular, being optimal depends only on the support of $\gamma$, 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 $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and the cost function $c : X \times Y \to \mathbb{R}$, maximize

$$\int \varphi \, d\mu + \int \psi \, d\nu,$$

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

$$\varphi(x) + \psi(y) \leq c(x, y), \quad \forall x \in X, \ y \in Y.$$

We call such a couple of functions *admissible potentials*
A simple inequality

Let $\gamma$ be a transport plan from $\mu$ to $\nu$ and $(\varphi, \psi)$ admissible potentials. Then

$$\int c(x, y) \, d\gamma(x, y) \geq \int \varphi(x) + \psi(y) \, d\gamma(x, y)$$

$$= \int \varphi(x) \, d\mu(x) + \int \psi(y) \, d\nu(y).$$

Thus

$$\inf\{\text{transport problem}\} \geq \sup\{\text{dual problem}\}$$
A property of admissible potentials

Say that $(\varphi, \psi)$ are admissible potentials and define

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

Then $\varphi^c \geq \psi$ and $(\varphi, \varphi^c)$ are admissible as well.
A property of admissible potentials

Say that \((\varphi, \psi)\) are admissible potentials and define

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

Then \(\varphi^c \geq \psi\) and \((\varphi, \varphi^c)\) 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 \((\psi^c, \psi)\) are admissible
The process stabilizes

Starting from \((\varphi, \psi)\), we can consider the admissible potentials 
\((\varphi, \varphi^c), (\varphi^{cc}, \varphi^c), (\varphi^{cc}, \varphi^{ccc})\)...
The process stabilizes

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

This process stops, because \(\varphi^{ccc} = \varphi^c\).
The process stabilizes

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

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

\[
\varphi^{ccc}(y) = \inf_{x} \sup_{\tilde{y}} \inf_{\tilde{\bar{x}}} c(x, y) - c(x, \tilde{y}) + c(\tilde{x}, \tilde{y}) - \varphi(\tilde{\bar{x}}),
\]

and picking \(\tilde{x} = x\) we get \(\varphi^{ccc} \leq \varphi^c\), and picking \(\tilde{y} = y\) we get \(\varphi^{ccc} \geq \varphi^c\).
A function $\varphi$ is $c$-concave if $\varphi = \psi^c$ for some function $\psi$.

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

$$
\sum_k c(x_k, y_k) = \sum_k \varphi(x_k) + \varphi^c(y_k)
$$

$$
= \sum_k \varphi(x_k) + \varphi^c(y_{\sigma(k)})
$$

$$
\leq \sum_k c(x_k, y_{\sigma(k)})
$$
Second structural theorem

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

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

Actually much more holds:

**Theorem** A set $\Gamma$ is $c$-cyclically monotone iff $\Gamma \subset \partial^c \varphi$ for some $\varphi$ $c$-concave.
To summarize

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

- \( \gamma \) is optimal
- \( \text{supp}(\gamma) \) is \( c \)-cyclically monotone
- \( \text{supp}(\gamma) \subset \partial^c \phi \) for some \( c \)-concave function \( \phi \)

(this requires some minor technical compatibility conditions between \( \mu, \nu, c \) which we neglect here)
No duality gap

It holds

\[ \inf \{ \text{transport problem} \} = \sup \{ \text{dual problem} \} \]

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

\[
\int c(x, y) \, d\gamma(x, y) = \int \varphi(x) + \varphi^c(y) \, d\gamma(x, y) = \int \varphi \, d\mu + \int \psi \, d\nu
\]
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 $c(x, y) = |x - y|^2/2$

c-concavity and convexity

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

$$\varphi(x) = \inf_y \frac{|x - y|^2}{2} - \psi(y)$$

$\Leftrightarrow$  

$$\varphi(x) = \inf_y \frac{|x|^2}{2} + \langle x, -y \rangle + \frac{|y|^2}{2} - \psi(y)$$

$\Leftrightarrow$  

$$\varphi(x) - \frac{|x|^2}{2} = \inf_y \langle x, -y \rangle + \left( \frac{|y|^2}{2} - \psi(y) \right)$$

$\Leftrightarrow$  

$$\varphi(x) = \sup_y \langle x, y \rangle - \left( \frac{|y|^2}{2} - \psi(y) \right),$$
The case $X = Y = \mathbb{R}^d$ and $c(x, y) = \frac{|x - y|^2}{2}$

c-superdifferential and subdifferential

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

Indeed:

$(x, y) \in \partial^c \varphi$

$\iff \left\{ \begin{array}{l} \varphi(x) = |x - y|^2/2 - \varphi^c(y), \\
\varphi(z) \leq |z - y|^2/2 - \varphi^c(y), \quad \forall z \in \mathbb{R}^d \end{array} \right.$

$\iff \left\{ \begin{array}{l} \varphi(x) - |x|^2/2 = \langle x, -y \rangle + |y|^2/2 - \varphi^c(y), \\
\varphi(z) - |z|^2/2 \leq \langle z, -y \rangle + |y|^2/2 - \varphi^c(y), \quad \forall z \in \mathbb{R}^d \end{array} \right.$

$\iff \varphi(z) - |z|^2/2 \leq \varphi(x) - |x|^2/2 + \langle z - x, -y \rangle \quad \forall z \in \mathbb{R}^d$

$\iff -y \in \partial^+(\varphi - | \cdot |^2/2)(x)$

$\iff y \in \partial^{-\varphi}(x)$
Let $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ be convex.

Then for a.e. $x$, $\varphi$ is differentiable at $x$. This is the same as to say that for a.e. $x$ the set $\partial^- \varphi(x)$ has only one element.
Brenier’s theorem: statement

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$. 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 $\gamma$. 

- Then $\text{supp}(\gamma) \subset \partial c \phi$ for some $c$-concave function $\phi$.

- Then $\text{supp}(\gamma) \subset \partial -\phi$ for some convex function $\phi$.

- Thus for $\gamma$-a.e. $(x, y)$ it holds $y \in \partial -\phi(x)$.

- Therefore for $\mu$-a.e. $x$ there is only one $y$ such that $(x, y) \in \text{supp}(\gamma)$, and this $y$ is given by $y := \nabla \phi(x)$.

- This is the same as to say that $\gamma = (\text{Id}, \nabla \phi) \ast \mu$.

- Now assume that $\tilde{\gamma}$ is another optimal plan. Then the plan $(\gamma + \tilde{\gamma})/2$ would be optimal and not induced by a map.
Brenier’s theorem: proof

- Pick an optimal plan $\gamma$.
- Then $\text{supp}(\gamma) \subset \partial^c \varphi$ for some $c$-concave function $\varphi$. 
Brenier’s theorem: proof

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

- Therefore for $\mu$-a.e. $x$ there is only one $y$ such that $(x, y) \in \text{supp}(\gamma)$, and this $y$ is given by $y := \nabla \varphi(x)$. 
- This is the same as to say that $\gamma = (\text{Id}, \nabla \varphi) \ast \mu$. 
- Now assume that $\tilde{\gamma}$ is another optimal plan. Then the plan $(\gamma + \tilde{\gamma})/2$ would be optimal and not induced by a map.
Brenier’s theorem: proof

- Pick an optimal plan \( \gamma \).
- Then \( \text{supp}(\gamma) \subset \partial^c \varphi \) for some \( c \)-concave function \( \varphi \).
- Then \( \text{supp}(\gamma) \subset \partial^- \varphi \) for some convex function \( \varphi \).
- Thus for \( \gamma \)-a.e. \((x, y)\) it holds \( y \in \partial^- \varphi(x) \).
Brenier’s theorem: proof

- Pick an optimal plan $\gamma$.
- Then $\text{supp}(\gamma) \subset \partial^c \varphi$ for some $c$-concave function $\varphi$.
- Then $\text{supp}(\gamma) \subset \partial^- \bar{\varphi}$ for some convex function $\bar{\varphi}$.
- Thus for $\gamma$-a.e. $(x, y)$ it holds $y \in \partial^- \bar{\varphi}(x)$.
- Therefore for $\mu$-a.e. $x$ there is only one $y$ such that $(x, y) \in \text{supp}(\gamma)$, and this $y$ is given by $y := \nabla \bar{\varphi}(x)$.
Brenier’s theorem: proof

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

► Pick an optimal plan $\gamma$.
► Then $\text{supp}(\gamma) \subset \partial^c \varphi$ for some $c$-concave function $\varphi$.
► Then $\text{supp}(\gamma) \subset \partial^- \varphi$ for some convex function $\varphi$.
► Thus for $\gamma$-a.e. $(x, y)$ it holds $y \in \partial^- \varphi(x)$.
► Therefore for $\mu$-a.e. $x$ there is only one $y$ such that $(x, y) \in \text{supp}(\gamma)$, and this $y$ is given by $y := \nabla \varphi(x)$.
► This is the same as to say that $\gamma = (Id, \nabla \varphi)_* \mu$.
► Now assume that $\tilde{\gamma}$ is another optimal plan. Then the plan $(\gamma + \tilde{\gamma})/2$ would be optimal and not induced by a map.