Week 6

Brenier's theorem: Let $X = Y = \mathbb{R}^d$, $c(x,y) = \frac{1}{2}|x-y|^2$, let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$,

with $\int |x|^2 d\mu + \int |y|^2 d\nu < +\infty$, $\mu \ll \mathcal{L}^d$.

Then, there exists a unique optimizer $\bar{\gamma}$ in (KP).

In addition, $\exists T = \nabla\varphi$, φ convex, s.t. $\bar{\gamma} = (Id, T)_{\#}\mu$,

$T = \nabla \varphi$ is the unique optimizer in (MP).

Corollary: Let c be continuous, bounded below. Then,

$\bar{\gamma}$ is optimal in (KP) $\Leftrightarrow$ supp $\bar{\gamma}$ is c-cycl. monotone $\Leftrightarrow$ $\exists \varphi$ c-convex s.t. supp $\bar{\gamma} \subseteq \partial^c \varphi$

Prop:

- $\varphi(x) + \varphi^c(y) + c(x,y) \geq 0 \quad \forall x \in X, y \in Y$.
- $\varphi(x) + \varphi^c(y) + c(x,y) = 0 \Leftrightarrow y \in \partial^c \varphi(x)$.

Proof:

▷ Existence: By Cor., supp $\bar{\gamma}$ is c-cyclic. monotone, supp $\bar{\gamma} \subseteq \partial\varphi$, φ convex.

By Prop., $\varphi(x) + \varphi^*(y) = x \cdot y$ on supp $\bar{\gamma} \subseteq \partial\varphi$

$\Longrightarrow (\varphi(x), \varphi^*(y))$ are finite for $\bar{\gamma}$-a.e. $(x,y) \Longrightarrow \varphi(x)$ is finite for μ-a.e. $x \to$

$\to \varphi$ is μ-a.e. differentiable, $\mu \ll \mathcal{L}^d$. (convex functions are diff. a.e. where they are finite.)

Let A with $\mu(A) = 0$ s.t. $\nabla\varphi(x)$ exists $\forall x \in \mathbb{R}^d \setminus A$. We claim that

$y = \nabla\varphi(x)$ for $\bar{\gamma}$-a.e. (x,y) $(*)$.

Let us prove first that $y = \nabla\varphi(x)$ $\forall (x,y) \in ((\mathbb{R}^d \setminus A) \times \mathbb{R}^d) \cap$ supp $\bar{\gamma}$.

This will be enough to show $(*)$, since $\bar{\gamma}(A \times \mathbb{R}^d) = \mu(A) = 0$.

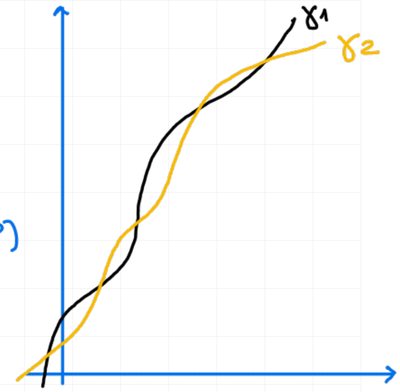
We have $(x,y) \in \partial\varphi$, and φ differentiable at x, so $y = \nabla\varphi(x)$,

and $(x,y) = (x, \nabla\varphi(x))$ $\bar{\gamma}$-a.e.

We now claim that $\bar{\gamma} = (Id, \nabla\varphi)_{\#}\mu$. Indeed, $\forall F$ measurable, $F \geq 0$,

$$\int F(x,y)\, d\bar{\gamma}(x,y) = \int F(x, \nabla\varphi(x))\, d\bar{\gamma} = \int F(x, \nabla\varphi(x))\, d\mu(x) = \int F(x,y)\, d[(Id, \nabla\varphi)_{\#}\mu] \checkmark$$

▷ <u>Uniqueness</u>: By contradiction, $\exists \gamma_1, \gamma_2$ optimal in (KP)



$\Rightarrow \dfrac{\gamma_1+\gamma_2}{2}$ is also optimal. Indeed:

$$\text{cost}\left(\frac{\gamma_1+\gamma_2}{2}\right) = \int c\, d\left(\frac{\gamma_1+\gamma_2}{2}\right) = \frac{1}{2}\int c\, d\gamma_1 + \frac{1}{2}\int c\, d\gamma_2 = (KP)$$

$\dfrac{\gamma_1+\gamma_2}{2}$ has the same marginals as γ_1 & γ_2.

By the first part of the proof, $\dfrac{\gamma_1+\gamma_2}{2} = (Id, \nabla\varphi)_{\#}\mu$, and even more precisely, $y = \nabla\varphi(x)$ $\forall (x,y) \in \underbrace{\text{supp}\,\dfrac{\gamma_1+\gamma_2}{2}}_{\text{supp}\,\gamma_1 \,\cup\, \text{supp}\,\gamma_2.} \cap ((\mathbb{R}^d \setminus A) \times \mathbb{R}^d)$.

In particular, $y = \nabla\varphi(x)$ $\forall (x,y) \in \text{supp}\,\gamma_i \cap [(\mathbb{R}^d \setminus A) \times \mathbb{R}^d]$ $i = 1, 2$.

As before:

$$\int F(x,y)\, d\gamma_1 = \int F(x, \nabla\varphi(x))\, d\gamma_1 = \int F(x,y)\, d((Id, \nabla\varphi)_{\#}\mu) \Rightarrow$$

$$\gamma_1 = (Id, \nabla\varphi)_{\#}\mu = \gamma_2.$$ Contradiction.

<u>Relation with Monge</u>: Let φ be the convex function constructed in the existence part. We know in general:

$$(KP) \le (MP) \le \int |\nabla\varphi(x) - x|^2\, d\mu(x)$$

$\nabla\varphi$ is a competitor in MP.

$\exists \rightarrow \|$

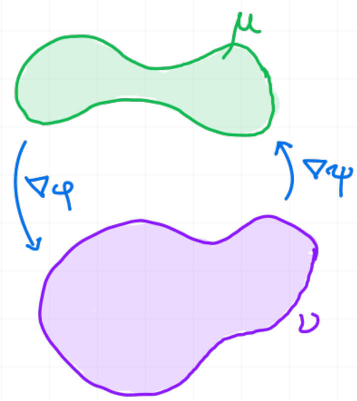
$$\int |x-y|^2\, d(Id, \nabla\varphi)_{\#}\mu = \int |x - \nabla\varphi(x)|^2\, d\mu(x)$$

Hence, they are all equal. Uniqueness of T in (MP) follows. □

<u>Corollary</u>: Let $\mu, \nu \ll \mathcal{L}^d$, $\int |x|^2 d\mu + \int |y|^2 d\nu < +\infty$.

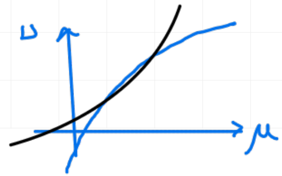
Let $\nabla\varphi$ optimal in (MP) with marginals μ and ν, and $\nabla\psi$ optimal in (MP) with marginals ν and μ.

Then, $\nabla\varphi$ is μ-a.e. invertible and its inverse is $\nabla\psi$:



$$\nabla\psi \circ \nabla\varphi = Id \quad \mu\text{-a.e.}, \qquad \nabla\varphi \circ \nabla\psi = Id \quad \nu\text{-a.e.}$$

<u>Idea</u>: Use uniqueness in (KP) and symmetry:



Proof: $\int |\nabla\varphi(x)-x|^2\,d\mu(x) = \int |x-y|^2\,d(Id,\nabla\varphi)_\#\mu = (KP) = \inf_{\gamma\in\Gamma(\mu,\nu)} \int |x-y|^2\,d\gamma$

$\int |\nabla\psi(x)-x|^2\,d\nu(x) = \int |x-y|^2\,d(\nabla\psi, Id)_\#\nu = (KP)$

By uniqueness of optimizer in (KP), $(Id,\nabla\varphi)_\#\mu = (\nabla\psi, Id)_\#\nu$ (*)

We claim that $\nabla\psi\circ\nabla\varphi - x = 0$ for μ-a.e. x. Indeed:

$$\int |\nabla\psi\circ\nabla\varphi(x) - x|^2\,d\mu = \int |\nabla\psi(y)-x|^2\,d[(Id,\nabla\varphi)_\#\mu] \overset{(*)}{=}$$

$$\int |\nabla\psi(y)-x|^2\,d(\nabla\psi, Id)_\#\nu = \int |\nabla\psi(y) - \nabla\psi(y)|^2\,d\nu = 0 \qquad \square$$

Theorem (Brenier's theorem for general costs):

Let $X = Y = \mathbb{R}^d$, (X Riem. manifold, Y metric space would be enough)

$\mu \ll \mathcal{L}^d$, supp ν compact. Let $c\in C^0(X\times Y)$ bounded below,

and assume $(KP) = \inf_{\gamma\in P(\mu,\nu)} \int c(x,y)\,d\gamma < +\infty$. Assume also:

- $\forall y\in \text{supp}\,\nu$, $\mathbb{R}^d \ni x \longmapsto c(x,y)$ is differentiable.
- $\forall y \in \text{supp}\,\nu$, $R>0$, $|\nabla_x c(x,y)| \le C_R$ for a.e. $x\in B_R$.
- $\forall x\in\mathbb{R}^d$, $\mathbb{R}^d \ni y \longmapsto \nabla_x c(x,y)\in\mathbb{R}^d$ is injective $\longleftarrow$ TWIST CONDITION

Then, there exists a unique optimizer $\bar\gamma$ in (KP), with $\bar\gamma = (Id, T)_\#\mu$,

and for some c-convex function φ, $\nabla_x c(x,T(x)) + \nabla\varphi = 0$.

Proof: Let $\bar\gamma$ optimal in (KP). By Cor., $\exists\,\varphi$ c-convex s.t. supp $\bar\gamma \subseteq \partial^c\varphi$.

By Prop.,

$$\underbrace{\varphi(x)+\varphi^c(y)+c(x,y)}_{\ge 0 \text{ always}} = 0 \quad \forall (x,y)\in \text{supp}\,\bar\gamma \subseteq \partial^c\varphi \quad (*).$$

We would like φ to be locally Lipschitz. We replace φ by $\tilde\varphi$ in the previous formula, with $\tilde\varphi$ locally Lipschitz:

$$\text{Let } \tilde\varphi(x) := \sup_{y\in \text{supp}\,\nu} \{\overbrace{-c(x,y)-\varphi^c(y)}^{\le \varphi(x)}\}$$ extended to $-\infty$ outside supp ν

Since supp $\bar\gamma \subseteq \mathbb{R}^d\times \text{supp}\,\nu$ and (*):

$$0 \le \tilde\varphi(x) + \varphi^c(y) + c(x,y) \le \varphi(x)+\varphi^c(y)+c(x,y) = 0 \quad \forall (x,y)\in\text{supp}\,\bar\gamma.$$

$\tilde\varphi$ is loc. Lipschitz because $x\mapsto -c(x,y)-\varphi^c(y)$ are C_R-Lipschitz in B_R, $R>0$, by 1st and 2nd assumptions.

Hence, $\sup_y \{-c(x,y) - \varphi^c(y)\}$ is C_R- Lipschitz. By Rademacher's theorem, $\tilde{\varphi}$ is μ-a.e. differentiable ($\mu \ll \mathcal{L}^d$)

Let A be the set of non-differentiability, $\mu(A) = 0$.

Let $(x,y) \in \operatorname{supp} \bar{\gamma}$, $x \notin A$. Then, $z \mapsto \tilde{\varphi}(z) + \varphi^c(y) + c(z,y)$ has a min at $z = x$, and it is differentiable there $\Longrightarrow \nabla \tilde{\varphi}(x) + \nabla_x c(x,y) = 0$.

Since, by assumption, $\nabla_x c(x,y)$ is injective in y, the equation has at most one solution, $y = T(x)$. As in the other proof (with quadratic cost) we showed: $\operatorname{supp} \bar{\gamma} \cap [(\mathbb{R}^d \setminus A) \times \operatorname{supp} \nu] \subseteq \operatorname{graph} T$, we conclude $\bar{\gamma} = (Id, T)_{\#}\mu$.

- The uniqueness works as in the other proof. $\square$

<u>Remark on the proof</u>: • Twist was used to invert the optimality condition.

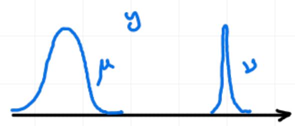
- $\|\nabla_x c\|_{L^\infty}$ was used to get $\tilde{\varphi}$ Lipschitz.
- if $\mu, \nu \ll \mathcal{L}^d$ and T, S are optimal maps, $\mu \underset{S}{\overset{T}{\rightleftarrows}} \nu$, then S and T are inverse of each other.

<u>Remark</u>: $c(x,y) = |x-y|^p$, $p > 1$, satisfies all the assumptions.

<u>Remark</u>: if $p = 1$, the twist condition fails.

The book shifting example had at least two optimizers in (KP)

<u>Remark</u> (extreme book shifting example)



μ, ν s.t. $x \le y$ $\forall (x,y) \in \operatorname{supp}\mu \times \operatorname{supp}\nu$. Then, <u>every</u> transport map (also every plan, why?) from μ to ν is optimal with linear cost, because

$$\int |T(x) - x| \, d\mu \overset{T(x) \in \operatorname{supp}\mu,\ x \in \operatorname{supp}\nu}{=} \int (T(x) - x) \, d\mu(x) = \underbrace{\int T(x) \, d\mu}_{} \overset{=\int x \, d\nu}{} - \int x \, d\mu$$

is <u>independent</u> of T.