

Week 7

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. $\bar{\gamma} = (\text{Id}, T)_\# \mu$,

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

Corollary: Let c be continuous, bounded below. Then, it is equivalent:

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

Stability of OT maps / plans

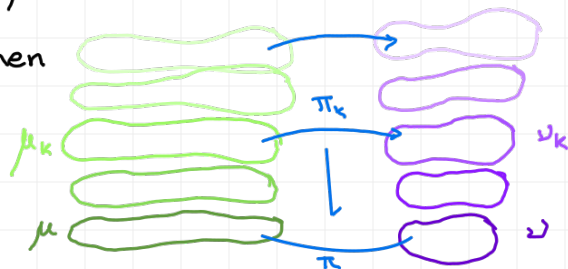
Theorem: Let $\{\mu_k\}_k, \{\nu_k\}_k \in \mathcal{P}(X)$, with $\text{supp } \mu_k, \text{supp } \nu_k \subseteq K$ compact,

$\mu_k \rightarrow \mu$, $\nu_k \rightarrow \nu$. Let $c: X \times X \rightarrow [0, \infty)$ continuous. Then:

① Any weak limit point π_k optimal in $\Gamma(\mu_k, \nu_k)$ ($\pi_k \rightarrow \pi$) is optimal in $\Gamma(\mu, \nu)$.

② If $X = \mathbb{R}^d$, $c(x,y) = \frac{|x-y|^2}{2}$, $\mu \ll \mathcal{L}^d$, then

$\pi_k \rightarrow (\text{Id}, \nabla \phi)_\# \mu$ as $k \rightarrow \infty$



If also $\mu_k \ll \mathcal{L}^d$, then it re-writes: $(\text{Id}, \nabla \phi_k)_\# \mu_k \rightarrow (\text{Id}, \nabla \phi)_\# \mu$;

as a consequence, if $\mu_k = \mu$, $\nabla \phi_k \rightarrow \nabla \phi$ in $L^p(\mu)$ $\forall p < +\infty$.

③ In addition to ②, if $\mu = \rho \mathcal{L}^d$ with $\rho > 0$ a.e. on an open bounded connected set A , then (for any $x_0 \in A$),

$\phi_k - \phi_k(x_0) \rightarrow \phi - \phi(x_0)$ uniformly on A .

Proof: ① Claim $\pi_k \rightarrow \pi \Rightarrow \pi \in \mathcal{P}(\mu, \nu)$ is optimal for (KP).

It is enough to show that $\text{supp } \pi$ is c-cycl. monotone.

General fact: $\{\pi_k\}_k \subseteq \mathcal{P}(X)$, $\pi_k \rightarrow \pi \Rightarrow \forall z \in \text{supp } \pi, \exists z_k \in \text{supp } \pi_k$
 s.t. $z_k \rightarrow z$. Indeed, $\forall r > 0$, if φ is a cut-off function
 between $B_{r/2}(z)$ and $B_r(z)$, $\underbrace{\int \varphi d\pi_k}_{> 0 \text{ for } k \text{ large enough}} \rightarrow \int \varphi d\pi > 0 \xrightarrow{z \in \text{supp } \pi}$
 $\Rightarrow \pi_k$ puts mass in B_r .

Now, given $(x_1, y_1), \dots, (x_N, y_N) \in \text{supp } \pi$, by the fact above
 $\exists (x_{i,k}, y_{i,k}) \in \text{supp } \pi_k$, with $(x_{i,k}, y_{i,k}) \rightarrow (x_i, y_i)$ as $k \rightarrow \infty$.

Since $\text{supp } \pi_k$ is c-cycl. monotone,

$$\sum_{i=1}^N c(x_{i+1,k}, y_{i,k}) \geq \sum_{i=1}^N c(x_{i,k}, y_{i,k}) \quad \forall k.$$

Letting $k \rightarrow \infty$, by continuity of $c \rightarrow \sum_{i=1}^N c(x_{i+1}, y_i) \geq \sum_{i=1}^N c(x_i, y_i)$.

② Claim: $(\text{Id}, \nabla \phi_k) \# \mu \rightarrow (\text{Id}, \nabla \phi) \# \mu \Rightarrow$

$$\Rightarrow \lim_{k \rightarrow \infty} \mu(\{|\nabla \phi_k - \nabla \phi| > \delta\}) = 0 \quad \forall \delta > 0 \quad \left(\begin{array}{l} \text{convergence in} \\ \text{measure or in} \\ \text{probability} \end{array} \right).$$

We conclude, since $|\nabla \phi_k|, |\nabla \phi| \leq M$ (target is compact in B_M)

$\Rightarrow \nabla \phi_k \rightarrow \nabla \phi$ in $L^p(\mu) \quad \forall p$. Indeed, fix $\delta > 0$,

$$\int |\nabla \phi_k - \nabla \phi|^p d\mu \leq \int_{|\cdot| \geq \delta} |\cdot|^p d\mu + \int_{|\cdot| \leq \delta} |\cdot|^p d\mu \leq \overbrace{\mu(\{|\nabla \phi_k - \nabla \phi| > \delta\})} \xrightarrow{\rightarrow 0} M^p + \delta^p$$

Hence, $\limsup_{k \rightarrow \infty} \int |\nabla \phi_k - \nabla \phi|^p d\mu \leq \delta^p$, δ arbitrary $\rightarrow$ lim is 0.

Proof of claim: Lusin's thm: $\mu \in \mathcal{P}(\mathbb{R}^d)$, $T: \mathbb{R}^d \rightarrow \mathbb{R}^d$ bounded,
 measurable, $\varepsilon > 0$. Then $\exists \tilde{T} \in C_b(\mathbb{R}^d)$, C closed s.t. $T = \tilde{T}$ on C ,
 $\mu(\mathbb{R}^d \setminus C) \leq \varepsilon$, $\sup_{\mathbb{R}^d} |\tilde{T}| \leq \sup |T|$
 $\hookrightarrow$ easy to enforce given the first two by truncation.

("measurable functions can be approximated by continuous functions")

So, by duis's theorem applied to $\nabla\phi$, $\exists \tilde{T}$, $\mu(\{|\nabla\phi - \tilde{T}|\} < \varepsilon)$,

$$\varphi(x,y) = \min\{2M, |y - \tilde{T}(x)|\} \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d) \quad \cdot \quad \int |\nabla\phi - \tilde{T}| d\mu \leq 2M\varepsilon$$

I test my hyp.: $\int \varphi(x,y) d(\text{Id}, \nabla\phi_k)_{\#} \mu \xrightarrow{\text{as } k \rightarrow \infty} \int \varphi(x,y) d(\text{Id}, \nabla\phi)_{\#} \mu$

$$\int |\nabla\phi_k - \tilde{T}| d\mu \geq \sqrt{\varepsilon} \mu(\{|\nabla\phi_k - \tilde{T}| > \sqrt{\varepsilon}\}) \quad \forall \varepsilon > 0.$$

Hence,

$$\lim_{k \rightarrow \infty} \mu(\{|\nabla\phi_k - \nabla\phi| > \sqrt{\varepsilon}\}) \leq \mu(\{|\nabla\phi_k - \tilde{T}| > \sqrt{\varepsilon}\}) + \mu(\{|\nabla\phi - \tilde{T}|\} < \varepsilon) \leq 2M\sqrt{\varepsilon} + \varepsilon.$$

Thus, $\lim_{k \rightarrow \infty} \mu(\{|\nabla\phi_k - \nabla\phi| > \delta\}) \leq 2M\sqrt{\varepsilon} + \varepsilon \quad \forall \varepsilon^2 < \delta.$

③ $\{\phi_k - \phi_k(x_0)\}$ are uniformly Lipschitz; $|\nabla\phi_k| \leq M$.

By Arzelà-Ascoli's theorem: $\phi_{k_i} - \phi_{k_i}(x_0) \rightarrow g$ uniformly in A

$$\nabla\phi_{k_i} \rightarrow \nabla\phi \quad \text{in } L^2(A)$$

$\Rightarrow g$ coincides, up to a constant, with ϕ , because $\forall \psi \in \mathcal{C}_c^\infty(A)$,

$$\int (\phi_{k_i} - \phi_{k_i}(x_0)) \cdot \nabla \cdot \psi = \int \nabla\phi_{k_i} \cdot \psi. \quad \text{Take } i \rightarrow \infty,$$

$$\int g \cdot \nabla \cdot \psi = \int \nabla\phi \cdot \psi \quad \forall \psi \Rightarrow \nabla g = \nabla\phi.$$

Regularity of optimal maps in Brenier's theorem

• Are they continuous? Not in general.

- What if the target is connected? Still no!

Counter-example: T_ε optimal between μ and ν_ε is discontinuous for ε sufficiently small.

If $\varepsilon = 0$, the balls are sent half to the left, half to the right. By stability:

$T_\varepsilon \rightarrow \nabla \phi$ in $L^p \Rightarrow$ up to a subsequence, pointwise.

$\exists \bar{x} \in B_\delta(x_2), \tilde{x} \in B_\delta(x_2)$ s.t. $T_\varepsilon(\bar{x})$ goes to $\bar{y}$ and $T_\varepsilon(\tilde{x})$ goes to $\tilde{y}$.

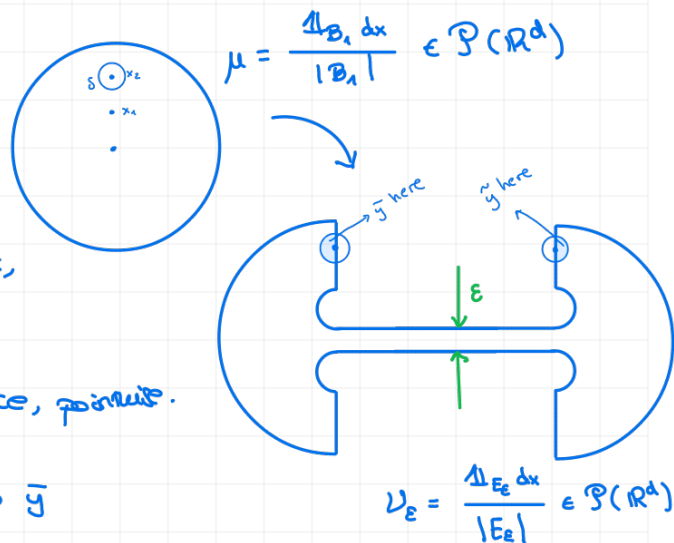
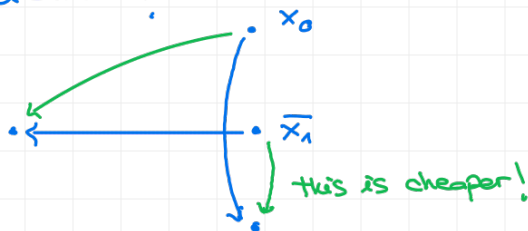
If T_ε continuous, $T_\varepsilon([\bar{x}, \tilde{x}])$ continuous, there is a point x_0 s.t. $T_\varepsilon(x_0) \cdot e_1 = 0$.

I also have $\bar{x}_1 \in B_\delta(x_1)$ s.t. $T_\varepsilon(\bar{x}_1) \in B_\delta(x_1) - (1, 0)$.

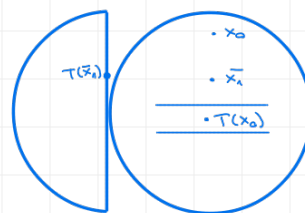
We look for a contradiction with c-cycl. monotonicity:

$\forall x_0, \bar{x}_1$ in the support, $(T(x_0) - T(\bar{x}_1)) \cdot (x_0 - \bar{x}_1) \geq 0$.

Contradiction!



$B_\delta(x_2)$



□ p-Wasserstein distance

(X, d) complete, separable metric space, $p \geq 0$

The p-Wasserstein distance is:

$$W_p(\mu, \nu) = \left[\min \left\{ \int_{X \times X} d^p(x, y) d\gamma(x, y) : \gamma \in \Gamma(\mu, \nu) \right\} \right]^{1/p}.$$

It is defined on (for some $x_0 \in X$), $\mathcal{P}_p(X) := \{ \mu \in \mathcal{P}(X) : \int d^p(x, x_0) d\mu < \infty \}$.

Question: does there exist μ s.t. $\int d^p(x, x_0) d\mu < \infty$ but $\int d^p(x, x_1) d\mu = +\infty$?

No: $\int d(x, x_1)^p d\mu \leq 2^p \int [d(x, x_0)^p + d(x_0, x_1)^p] d\mu < +\infty.$
 $\hookrightarrow d(x, x_1) \leq d(x, x_0) + d(x_0, x_1)$

Remark: The choice of the point x_0 does not play any role!

Remark: $\int d(x, x_0)^p d\mu = W_p^p(\mu, \delta_{x_0}).$

Theorem: W_p is a distance on $\mathcal{P}_p(X)$.

Proof: ① $W_p(\mu, \mu) = 0$ ($\gamma = (\text{Id}, \text{Id})_{\#} \mu$ is optimal).

② $W_p(\mu, \nu) = 0 \stackrel{?}{\Rightarrow} \mu = \nu$?

Let γ optimal in W_p , $\int d(x, y)^p d\gamma = 0 \Rightarrow x = y$ γ -a.e.

Let φ be a continuous function:

$$\int \varphi d\mu \stackrel{\gamma \in \Gamma(\mu, \nu)}{=} \int \varphi(x) d\gamma(x, y) = \int \varphi(y) d\gamma(x, y) = \int \varphi(y) d\nu(y)$$

$\Rightarrow \mu = \nu$ because φ is arbitrary.

③ $W_p(\mu, \nu) = W_p(\nu, \mu)$ by symmetry: let $S(x, y) = (y, x)$.

$\Gamma(\mu, \nu) \ni \gamma \mapsto S_{\#} \gamma \in \Gamma(\nu, \mu)$ and S does not change the cost of γ .

④ Triangle inequality and finiteness:

$$W_p(\mu, \sigma) \leq W_p(\mu, \nu) + W_p(\nu, \sigma) \quad (*)$$

Once this is proved, $W_p(\mu, \nu) \leq W_p(\mu, \delta_{x_0}) + W_p(\nu, \delta_{x_0}) \overset{\text{in } \mathcal{P}(X)}{\leq} +\infty$

Proof of (*) assuming there exist optimal maps:

$$\mu \xrightarrow{T} \nu \xrightarrow{S} \sigma \quad d(S \circ T(x), x) \quad d(S \circ T, \text{Id})$$

$$W_p(\mu, \sigma) \leq \left[\int |S \circ T(x) - x|^p d\mu(x) \right]^{1/p} = \|S \circ T - \text{Id}\|_{L^p(\mu)}$$

$$\overset{L^p \text{ is a distance}}{\leq} \underbrace{\|S \circ T - T\|_{L^p(\mu)}}_{T_{\#}\mu = \nu \rightarrow \|S - \text{Id}\|_{L^p(\nu)}} + \|T - \text{Id}\|_{L^p(\mu)} = W_p(\mu, \nu) + W_p(\nu, \sigma)$$

□