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

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

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

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

Stability of OT maps / plans

just a technicality

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

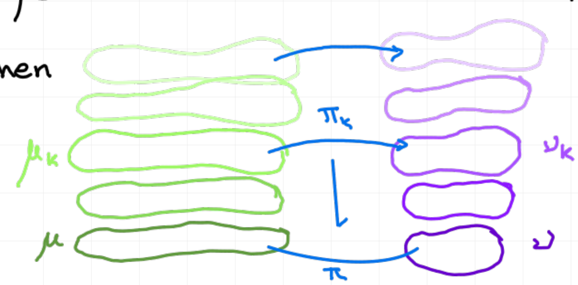
$\mu_k \rightharpoonup \mu$, $\nu_k \rightharpoonup \nu$. Let $c: X \times X \to [0,\infty)$ continuous. Then:

in the (KP), $\inf_{\gamma \in P(\mu,\nu)} \left\{ \int c(x,y)\, d\gamma \right\}$

① Any weak limit point π_k optimal in $\Gamma(\mu_k, \nu_k)$ $(\pi_k \rightharpoonup \pi)$ 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

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



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

as a consequence, if $\mu_k = \mu$, $\nabla\phi_k \to \nabla\phi$ 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 $x_0 \in A$),

$$\phi_k - \phi_k(x_0) \longrightarrow \phi - \phi(x_0) \text{ uniformly on } A.$$

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

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

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

$\implies \pi_k$ puts mass in B_r.

Now, given $(x_1,y_1),\dots,(x_N,y_N) \in \operatorname{supp}\pi$, by the fact above

$\exists (x_{i,k}, y_{i,k}) \in \operatorname{supp}\pi_k$, with $(x_{i,k},y_{i,k}) \to (x_i,y_i)$ as $k\to\infty$.

Since $\operatorname{supp}\pi_k$ 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 $k\to\infty$, by continuity of $c \to \sum_{i=1}^N c(x_{i+1},y_i) \geq \sum_{i=1}^N c(x_i,y_i)$.

(2) Claim: $(Id,\nabla\phi_k)_\#\mu \to (Id,\nabla\phi)_\#\mu \implies$

$\implies \lim_{k\to\infty} \mu(\{|\nabla\phi_k - \nabla\phi| > \delta\}) = 0 \quad \forall \delta>0$ (convergence in measure or in probability).

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

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

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

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

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

↳ easy to enforce given the first two by truncation.

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

So, by <u>Lusin's theorem</u> applied to $\nabla\phi$, $\exists \tilde{T}$, $\mu(\{\nabla\phi \neq \tilde{T}\}) < \varepsilon$,

$$\varphi(x,y) = \min\{2M, |y-\tilde{T}(x)|\} \in \mathcal{C}_b(\mathbb{R}^d \times \mathbb{R}^d).$$

I test my hyp.: $$\underbrace{\int \varphi(x,y)\, d(Id, \nabla\phi_k)_{\#}\mu}_{} \xrightarrow{\text{as } k\to\infty} \overbrace{\int \varphi(x,y)\, d(Id,\nabla\phi)_{\#}\mu}^{\leq \int |\nabla\phi - \tilde{T}|\, d\mu \,\leq\, 2M\varepsilon}$$

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

Hence,

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

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

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

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

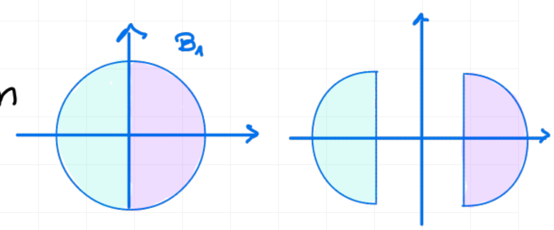
$\nabla\phi_{k_i} \to \nabla\phi$ in $L^2(A)$

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

$\int (\phi_{k_i} - \phi_{k_i}(x_0))\nabla\cdot\xi = \int \nabla\phi_{k_i}\cdot\xi$. Take $i \to \infty$,

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

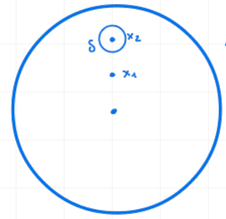
<u>Regularity of optimal maps in Brenier's theorem</u>



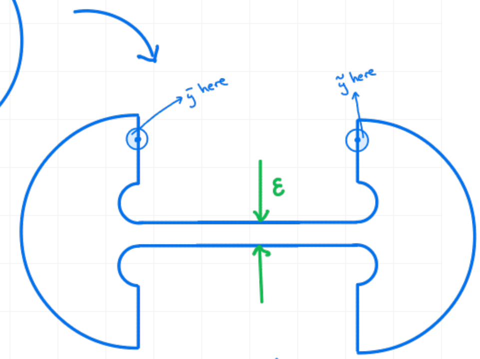
- Are they continuous? Not in general.

- What if the target is connected? Still no!

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



$$\mu = \frac{\mathbb{1}_{B_1}\,dx}{|B_1|} \in \mathcal{P}(\mathbb{R}^d)$$



$$\nu_\varepsilon = \frac{\mathbb{1}_{E_\varepsilon}\,dx}{|E_\varepsilon|} \in \mathcal{P}(\mathbb{R}^d)$$

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

$T_\varepsilon \longrightarrow \nabla\phi$ 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 $\tilde{y}$.

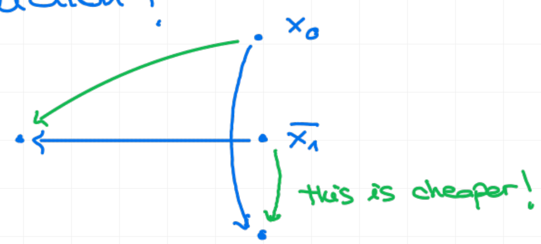
If T_ε continuous, $T_\varepsilon([\bar{x}, \tilde{x}])$ continuous, there is a point $\overset{B_\delta(x_2)}{\overset{\cup}{x_0}}$ s.t. $T_\varepsilon(x_0)\cdot e_1 = 0$.

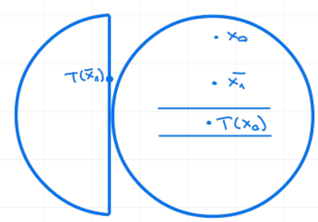
I also have $\bar{x}_1 \in B_\delta(x_1)$ s.t. $T_\varepsilon(\bar{x}_1) \in B_\delta(x_1) - (1,0)$.

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

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

Contradiction!





□ p-Wasserstein distance

(X,d) complete, separable metric space, $p \geq 1$.

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 $x_0\in X$), $\mathcal{P}_p(X) := \left\{\mu\in\mathcal{P}(X) : \int d^p(x,x_0)\,d\mu<\infty\right\}$.

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: $\int d(x,x_1)^p\,d\mu \leq 2^p\int\left[d(x,x_0)^p + d(x_0,x_1)^p\right]d\mu < +\infty.$

$\hookrightarrow d(x,x_1)\leq d(x,x_0)+d(x_0,x_1)$

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 $W_p(\mu,\mu)=0$ $\quad(\gamma=(Id,Id)_\#\mu$ is optimal$)$.

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

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:

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

$\Rightarrow \mu=\nu$ 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:

$$W_p(\mu,\sigma) \leq W_p(\mu,\nu) + W_p(\nu,\sigma) \qquad (*) \qquad \text{in } \mathcal{P}_p(X).$$

Once this is proved, $W_p(\mu,\nu) \leq W_p(\mu,\delta_{x_0}) + W_p(\nu,\delta_{x_0}) < +\infty$

Proof of $(*)$ assuming there exist optimal maps:

$$\mu \xrightarrow{T} \nu \xrightarrow{S} \sigma$$

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

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

$\square$