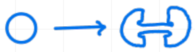
Week 8

Recall:

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

<u>Theorem</u>: W_p is a distance on $\mathcal{P}_p(X) := \{\mu \in \mathcal{P}(X) : \int d^p(x,x_0)\,d\mu < \infty\}$.

<u>Stability</u>: Any weak limit point of π_k optimal in $\Gamma(\mu_k,\nu_k)$ $(\pi_k \rightharpoonup \pi)$ is optimal in $\Gamma(\mu,\nu)$.

<u>Counter-example</u>: T_ε :  is discontinuous.

<u>Theorem</u> (Caffarelli, '90): $\mu_1 = \rho_1 \mathcal{L}^d|_U$, $\mu_2 = \rho_2 \mathcal{L}^d|_V$. U open bounded, V convex bounded, $0 < \lambda \le \rho_2 \le \Lambda < \infty$ in V. Then, the OT : $\nabla\varphi : \mu_1 \to \mu_2$ satisfies:

① $\varphi \in C^{1,\beta}$ for some $\beta = \beta(d,\lambda,\Lambda) > 0$.

② if $\rho_i \in C^{k,\alpha}$ for some $k \in \mathbb{N}$, $\alpha > 0 \Longrightarrow \varphi \in C^{k+2,\alpha}$

why do we gain two derivatives? $\det \nabla^2\varphi = \dfrac{\rho_1(x)}{\rho_2(\nabla\varphi(x))}$ Monge-Ampère equation.

<u>Disintegration theorem</u>: Z, X complete, separable metric spaces, $f : Z \to X$ Borel.

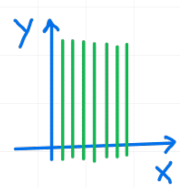
Then, $\gamma \in \mathcal{P}(Z)$, $\exists \{\gamma_x\}_{x\in X} \subseteq \mathcal{P}(Z)$ s.t.

- $\gamma = \int \gamma_x \, d\sigma$, where $\sigma = f\#\gamma$.
- γ_x concentrated on $f^{-1}(x)$.

$\int_Z \phi\, d\gamma = \int_X \int_Z \phi(y)\, d\gamma_x(y)\, d\sigma(x)$ by definition. $(*)$

<u>Special case</u>: $Z = X \times Y$, $f = \pi_1$, $\gamma \in \mathcal{P}(X\times Y)$, $\pi_1 \# \gamma = \mu$.

$\pi_1^{-1}(\{x\}) \approx Y$, γ_x can be thought of as a measure on Y:



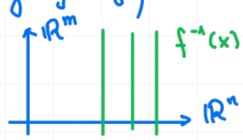
<u>Notation</u>: $\gamma = \mu \otimes \gamma_x$.

<u>Particular cases</u>:

- $\gamma \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^m)$, $f = \pi_1$, $\gamma = g(x,y)\,dx\,dy$. $\int \phi \, d(\pi_1\#\gamma)$

$\sigma = \pi_1 \# \gamma = \left(\int g(x,y)\,dy\right) dx$; indeed $\int \phi(x)\, g(x,y)\,dx\,dy = \int \phi(x) \cdot \left[\int g(x,y)\,dy\right] dx$

$\gamma_x = \dfrac{g(x,y)}{\sigma(x)}\, dy$



. One can check in $(*)$ that this is indeed the disintegration

- $\gamma = (Id, T)\#\mu$, $f = \pi_1$

$\pi_1 \# \gamma = (\pi_1 \circ (Id,T))\#\mu = \mu$, $\gamma_x = \delta_{T(x)}$

Indeed, in $(*)$: $\int \phi(x,y)\, d\gamma = \int \phi(x, T(x))\, d\mu = \iint \phi(x,y)\, d\delta_{\{y=T(x)\}}\, d\mu(x)$.

Lemma (Dudley): $\overset{\mu_1}{\downarrow}X_1, \overset{\mu_2}{\downarrow}X_2, \overset{\mu_3}{\downarrow}X_3$ metric spaces, $\gamma_{12} \in \Gamma(\mu_1, \mu_2)$, $\gamma_{23} \in \Gamma(\mu_2, \mu_3)$.

Then, there exists a "3-plan" $\Lambda \in \mathcal{P}(X_1 \times X_2 \times X_3)$ s.t.

$$\pi_{12} \# \Lambda = \gamma_{12}, \quad \pi_{23} \# \Lambda = \gamma_{23}.$$

Proof: Disintegrate wrt the marginal on which we see μ_2:

$$\gamma_{12} = \gamma_{12,x_2}(x_1) \otimes \mu_2(x_2) \qquad \gamma_{23} = \mu_2(x_2) \otimes \gamma_{23,x_2}(x_3)$$

Define: $\Lambda = \mu_2(x_2) \otimes \left[\gamma_{12,x_2}(x_1) \times \gamma_{23,x_2}(x_3)\right]$, namely:

$$\int \phi(x_1, x_2, x_3)\, d\Lambda = \int_{X_2} \int_{X_1 \times X_3} \phi(x_1, x_2, x_3)\, d\gamma_{12,x_2}(x_1)\, d\gamma_{23,x_2}(x_3)\, d\mu_2(x_2).$$

Λ has correct marginals $\pi_{12} \# \Lambda \overset{?}{=} \gamma_{12}$:

$$\int \phi(x_1, x_2)\, d\Lambda = \int_{X_2} \int_{X_1 \times X_3} \phi(x_1, x_2)\, d\gamma_{12,x_2}(x_1)\, \underbrace{d\gamma_{23,x_2}(x_3)}_{1}\, d\mu(x_2) =$$

by definition of disintegration

$$= \int_{X_2} \int_{X_1} \phi(x_1, x_2)\, d\gamma_{12,x_2}(x_1)\, d\mu(x_2) = \int \phi(x_1, x_2)\, d\gamma_{12}(x_1, x_2). \qquad \square$$

Proof of the triangle inequality for p-Wasserstein distance:

$$W_p(\mu_1, \mu_3) \overset{?}{\leq} W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3)\ ?$$

Let $\gamma_{12} \in \Gamma(\mu_1, \mu_2)$, $\gamma_{23} \in \Gamma(\mu_2, \mu_3)$, optimal.

By Dudley's lemma, $\exists\, \Lambda$ "composition of plans", with $\pi_{13} \# \Lambda \in \Gamma(\mu_1, \mu_3)$.

Indeed, $(\pi_1 \circ \pi_{13})_\# \Lambda = \pi_1 \# \Lambda = \pi_1 \# \pi_{12} \# \Lambda = \pi_1 \# \gamma_{12} = \mu_1$,

and same for the second marginal.

$$\int d(x_1, x_3)^p\, d\pi_{13}\#\Lambda = \int d(x_1, x_3)^p\, d\Lambda$$

$$W_p(\mu_1, \mu_3) \leq \| d(x_1, x_3) \|_{L^p(\pi_{13} \# \Lambda)} = \| d(x_1, x_3) \|_{L^p(\Lambda)}$$

$$\leq \| d(x_1, x_2) \|_{L^p(\Lambda)} + \| d(x_2, x_3) \|_{L^p(\Lambda)}$$

$$= \| d(x_1, x_2) \|_{L^p(\overbrace{\pi_{12} \# \Lambda}^{\gamma_{12}})} + \| d(x_2, x_3) \|_{L^p(\gamma_{23})}$$

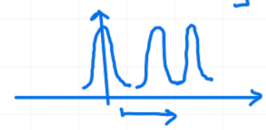
$$= W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3). \qquad \square$$

Remark: $(X, d) \hookrightarrow (\mathcal{P}(X), W_p)$ isometrically.

$$x \longmapsto \delta_x$$

Characterization of convergence in $\mathcal{P}_p(X)$ [Vertical L^2 — Horizontal W_2]

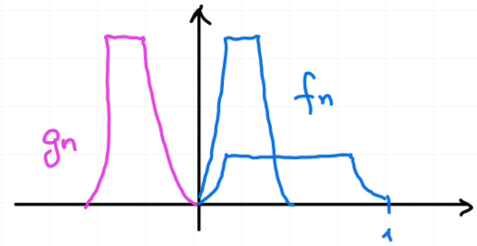
↳ see example



Remark: Let $f : \mathbb{R} \longrightarrow \mathbb{R}_{\geq 0}$ be supported in $[0,1]$, with $\int f = 1$.

Let $f_n(x) = n f(nx) \longrightarrow \|f_n\|_{L^1} = 1$

$g_n(x) = f_n(x - \frac{1}{n})$



- $\|f_n - g_n\|_{L^1} = ?$
- $W_2(f_n dx, g_n dx) = ?$

Theorem: Let $1 \leq p < +\infty$, $x_0 \in X$. Let $\{\mu_n\}_n \subseteq \mathcal{P}_p(X)$, $\mu \in \mathcal{P}_p(X)$.

Then, it is equivalent:

tested against $C_b(X)$

① $\mu_n \longrightarrow \mu$ narrowly and $\underbrace{\overbrace{\int_X d^p(x,x_0)\, d\mu_n}^{W_p^p(\mu_n,\delta_{x_0})} \longrightarrow \overbrace{\int_X d^p(x,x_0)\, d\mu}^{W_p^p(\mu,\delta_{x_0})}}_{(*)}$

② $W_p(\mu_n,\mu) \longrightarrow 0$

Corollary (W_p metrizes weak-* convergence) If $1 \leq p < \infty$, X compact metric space.

It is equivalent:

① $\mu_n \xrightarrow{*} \mu$ weakly-*

② $W_p(\mu_n,\mu) \longrightarrow 0$

Thm ⇒ Cor:

- narrow convergence = weak-* convergence when the space is compact
- Condition (*) is automatic because $d(x,x_0) \in C_b(X)$ in a compact space.

Proof of Theorem:

② ⇒ ① Assume $W_p(\mu_n,\mu) \to 0$.

By triangle inequality, $W_p(\mu_n, \delta_{x_0}) \to W_p(\mu, \delta_{x_0})$.

We want to show $\forall \varphi \in C_b(X)$, $\int \varphi\, d\mu_n \to \int \varphi\, d\mu$ (☆).

Take $\varphi \in Lip_b(X)$,

$$\left| \int \varphi(x)\, d\mu_n(x) - \int \varphi(y)\, d\mu(y) \right| \leq \left| \int \overbrace{(\varphi(x) - \varphi(y))}^{\leq L d(x,y)}\, d\overbrace{\gamma_n(x,y)}^{\in \Gamma(\mu_n,\mu)} \right|$$

$$\leq Lip\,\varphi \int d(x,y)\, d\gamma_n(x,y) \underset{\text{Hölder}}{\leq} Lip\,\varphi \left(\int d(x,y)^p\, d\gamma(x,y) \right)^{1/p}$$

$$\underset{\inf}{=} (Lip\,\varphi)\, W_p(\mu_n,\mu).$$

Now use a standard way to approach $\varphi \in C_b(X)$ with φ Lipschitz:

$$\mathrm{Lip}_m \ni \varphi^m(x) = \sup_y \big(\varphi(y) - m\, d(x,y)\big) \downarrow \varphi$$

$$\mathrm{Lip}_m \ni \varphi^m(x) = \inf_y \big(\varphi(y) + m\, d(x,y)\big) \uparrow \varphi \qquad \varphi_m \uparrow \varphi$$

For every m_0, $\displaystyle\int \varphi_{m_0}\, d\mu \overset{(\star)\text{ for Lipschitz}}{=} \lim_{n\to\infty} \int \varphi_{m_0}\, d\mu_n \le \liminf_{m\to\infty} \int \varphi_m\, d\mu$

Take sup in m_0: $\displaystyle\int \varphi\, d\mu \le \liminf_{m\to\infty} \int \varphi_m\, d\mu.$

The other inequality follows by the same argument approximating from above.

① $\Rightarrow$ ② <u>when X is compact</u>

Apply the stability result $\mu_n \to \mu$, $\overset{\mu}{\overset{\shortparallel}{\nu_n}} \to \overset{\mu}{\overset{\shortparallel}{\nu}}$. Optimal plan μ to μ is $(\mathrm{Id},\mathrm{Id})_\# \mu$,

by stability γ_n OT between μ_n and $\mu \overset{*}{\longrightarrow} (\mathrm{Id},\mathrm{Id})_\#\mu$. (same as narrow here)

Test the previous weak-$*$ convergence with $d^p(x,y) \in C_b(X)$:

$$\underbrace{\int d^p(x,y)\, d\gamma_n(x,y)}_{W_p(\mu_n,\mu)} \longrightarrow \int d^p(x,y)\, d\big((\mathrm{Id},\mathrm{Id})_\#\mu\big) = 0$$

<u>Remark</u>: Let $\{\mu_n\}_n \subseteq \mathcal{P}(X)$, $\mu \in \mathcal{P}(X)$. Then: $\mu_n \overset{*}{\rightharpoonup} \mu$ weakly-$*$ $\Leftrightarrow$ $\mu_n \to \mu$ narrowly.

<u>Proof in $\mathbb{R}^d$</u>: Let $\eta \in C_b(\mathbb{R}^d)$:

cut-off between B_{2R} and B_R.

$$\left|\int \eta\, d\mu_n - \int \eta\, d\mu\right| \le \underbrace{\left|\int \eta_R\, \eta\, d\mu_n - \int \eta_R\, \eta\, d\mu\right|}_{\to 0 \text{ as } n\to\infty \text{ for } R \text{ fixed}} + \underbrace{\left|\int (1-\eta_R)\eta\, d\mu_n\right| + \left|\int (1-\eta_R)\eta\, d\mu\right|}_{\le [\mu_n(B_R^c) + \mu(B_R^c)] \sup \eta}$$

We now make this small uniformly in n by taking R large:

$R > 0$ s.t. $\mu(B^c_{R/2}) \le \varepsilon$: $\qquad \int \eta_{R/2}\, d\mu \ge 1 - \varepsilon$

$$\mu_n(B_R^c) \le \int (1 - \eta_{R/2})\, d\mu_n = 1 - \int \eta_{R/2}\, d\mu_n \le (1 - (1-\varepsilon)). \qquad \square$$