

Week 8

Recall:

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

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

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

Counter-example: $T_\varepsilon : \mathbb{O} \rightarrow \mathbb{O}$ is discontinuous.

Theorem (Caffarelli, '90): $\mu_1 = \rho_1 x^d|_U$, $\mu_2 = \rho_2 x^d|_V$, U open bounded, V convex bounded, $0 < \lambda \leq \rho_2 \leq \Lambda < \infty$ in V . Then, the OT: $\nabla \varphi : \mu_1 \rightarrow \mu_2$ 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 $k \in \mathbb{N}$, $\alpha > 0 \Rightarrow \varphi \in C^{k+2, \alpha}$

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

Disintegration theorem: Z, X complete, separable metric spaces, $f: Z \rightarrow X$ Borel.

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

• $\gamma = \int \gamma_x dG$, where $G = f\# \gamma$.

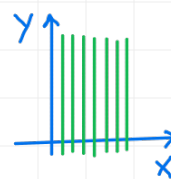
• γ_x concentrated on $f^{-1}(x)$.

$$\int_Z \phi d\gamma = \int_X \int_Z \phi(y) d\gamma_x(y) dG(x) \quad \text{by definition. } (*)$$

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

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

Notation: $\gamma = \mu \otimes \gamma_x$.

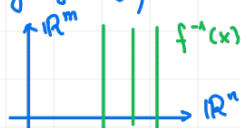


Particular cases:

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

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

$\gamma_x = \frac{g(x, y)}{G(x)} dx$



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

• $\gamma = (\text{Id}, T)\# \mu$, $f = \pi_1$

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

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 :

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

Define: $\Lambda = \mu_2(x_2) \otimes [\gamma_{12, x_2}(x_1) \times \gamma_{23, x_2}(x_3)]$, 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 $\pi_{12} \# \Lambda \stackrel{?}{=} \gamma_{12}$:

$$\begin{aligned} \int \phi(x_1, x_2) d\Lambda &= \int_{x_2} \int_{x_1 \times x_3} \phi(x_1, x_2) d\gamma_{12, x_2}(x_1) d\gamma_{23, x_2}(x_3) d\mu_2(x_2) = \\ &= \int_{x_2} \int_{x_1} \phi(x_1, x_2) d\gamma_{12, x_2}(x_1) d\mu_2(x_2) = \int \phi(x_1, x_2) d\gamma_{12}(x_1, x_2). \quad \square \end{aligned}$$

by definition of disintegration

Proof of the triangle inequality for p -Wasserstein distance:

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

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, $(\pi_{12} \circ \pi_{13}) \# \Lambda = \pi_{12} \# \Lambda = \pi_{12} \# \pi_{12, x_2} \# \Lambda = \pi_{12} \# \gamma_{12} = \mu_1$,

and same for the second marginal.

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

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

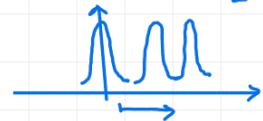
$$= \|d(x_1, x_2)\|_{L^p(\pi_{12} \# \Lambda)} + \|d(x_2, x_3)\|_{L^p(\gamma_{23})}$$

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

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



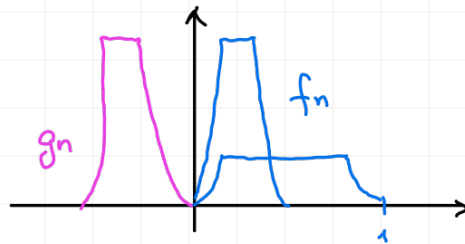
Remark: let $f: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ be supported in $[0,1]$, with $\int f = 1$.

$$\text{let } f_n(x) = n f(nx) \rightarrow \|f_n\|_{L^1} = 1$$

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

$$\bullet \|f_n - g_n\|_{L^1} =$$

$$\bullet W_2(f_n dx, g_n dx) =$$



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

Then, it is equivalent:
 tested against $C_b(X)$
 $W_p^p(\mu_n, \delta_{x_0})$
 $W_p^p(\mu, \delta_{x_0})$

$$\textcircled{1} \mu_n \rightarrow \mu \text{ narrowly and } \underbrace{\int_X d^p(x, x_0) d\mu_n \rightarrow \int_X d^p(x, x_0) d\mu}_{(*)}$$

$$\textcircled{2} W_p(\mu_n, \mu) \rightarrow 0$$

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

It is equivalent:

$$\textcircled{1} \mu_n \xrightarrow{*} \mu \text{ weakly-}^*$$

$$\textcircled{2} W_p(\mu_n, \mu) \rightarrow 0$$

Thm $\Rightarrow$ 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:

$$\textcircled{2} \Rightarrow \textcircled{1} \text{ Assume } W_p(\mu_n, \mu) \rightarrow 0.$$

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

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

Take $\varphi \in \text{Lip}_b(X)$,

$$\begin{aligned} \left| \int \varphi(x) d\mu_n(x) - \int \varphi(y) d\mu(y) \right| &\leq \left| \int \underbrace{(\varphi(x) - \varphi(y))}_{\leq L d(x,y)} \underbrace{d\mu_n(x,y)}_{\in \mathcal{P}(\mu_n, \mu)} \right| \\ &\leq \text{Lip } \varphi \int d(x,y) d\mu_n(x,y) \leq \text{Lip } \varphi \left(\int d(x,y)^p d\mu_n(x,y) \right)^{1/p} \\ &\stackrel{\text{H\"older}}{\leq} \text{Lip } \varphi \left(\int d(x,y)^p d\mu_n(x,y) \right)^{1/p} \\ &= \text{Lip } \varphi W_p(\mu_n, \mu). \end{aligned}$$

inf

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

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

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

(*) for Lipschitz

$$\text{For every } m_0, \int \varphi_{m_0} d\mu = \lim_{n \rightarrow \infty} \int \varphi_{m_0} d\mu_n \leq \liminf_{m \rightarrow \infty} \int \varphi_m d\mu$$

$$\text{Take sup in } m_0: \int \varphi d\mu \leq \liminf_{m \rightarrow \infty} \int \varphi_m d\mu.$$

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

① $\Rightarrow$ ② when X is compact

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

by stability γ_n or between μ_n and $\mu \xrightarrow{*} (\text{Id}, \text{Id})_{\#} \mu$. → same as narrow here

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

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

Remark: Let $\{\mu_n\}_n \in \mathcal{P}(X)$, $\mu \in \mathcal{P}(X)$. Then: $\mu_n \xrightarrow{*} \mu$ weakly* $\Leftrightarrow \mu_n \rightarrow \mu$ narrowly.

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

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

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

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

$$R > 0 \text{ s.t. } \mu(B_{R/2}^c) \leq \varepsilon:$$

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