Lecture 14 Announcements:

- Select Wiki Article for revision by Mon 2/21.

Recall:

Brenier's theorem

compact Polish space

Thm: For all $\mu,\nu \in \mathcal{P}(\mathcal{X})$, $c: \mathcal{X}\times\mathcal{X} \to [0,+\infty)$ cts,

$$\underbrace{\inf_{\gamma \in \Gamma(\mu,\nu)} \mathbb{K}(\gamma)}_{-D_0} = \underbrace{\sup_{\substack{(\varphi,\psi) \in C(\mathcal{X})\times C(\mathcal{X}) \\ \varphi \oplus \psi \leq c}} \int \varphi \, d\mu + \int \psi \, d\nu}_{-P_0}$$

Furthermore, the maximum is attained.

Thm: Given $\mathcal{X} \subset\subset \mathbb{R}^d$, $\mu,\nu \in \mathcal{P}(\mathcal{X})$,

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \mathbb{K}_1(\gamma) = \sup_{\varphi \in C(\mathcal{X}), \|\varphi\|_{Lip} \leq 1} \int \varphi \, d(\mu-\nu)$$

and $\exists\, \varphi^*$ that achieves maximum.

Pf: By our duality theorem, it suffices to show

$$\underbrace{\sup_{\substack{\varphi,\psi\in C(X)\\ \varphi(x^1)+\psi(x^2)\le|x^1-x^2|}} \int\varphi\,d\mu+\int\psi\,d\nu}_{LHS} = \underbrace{\sup_{\varphi\in C(X),\ \|\varphi\|_{Lip}\le 1} \int\varphi\,d(\mu-\nu)}_{RHS}$$

"$\geq$" ✓

Next "$\leq$"

Take (φ_*,ψ_*) that attain maximum on LHS.

Double convexification trick:

Define $\tilde{\psi}(x^2) = \inf_{x^1\in X} |x^1-x^2| - \varphi_*(x^1)$

- $\tilde{\psi} \geq \psi_*$
- $\forall\ \varepsilon>0, x^2, y^2, \exists\ y^1$ s.t

$$\tilde{\psi}(x^2)-\tilde{\psi}(y^2) \leq |y^1-x^2|-\varphi_*(y^1)-\left(|y^1-y^2|-\varphi_*(y^1)\right)+\varepsilon$$

$$\leq |x^2-y^2|+\varepsilon$$

$\|\tilde{\psi}\|_{Lip} \leq 1$

- $\varphi_*(x^1)+\tilde{\psi}(x^2) \leq |x^1-x^2|$

Thus $(\varphi_*, \tilde{\Psi})$ must also be optimal for original dual problem.

Define $\tilde{\varphi}(x^1) = \inf_{x^2 \in X} |x^1 - x^2| - \tilde{\Psi}(x^2)$

- $\tilde{\varphi} \geq \varphi^*$
- $\|\varphi\|_{Lip} \leq 1$
- $(\tilde{\varphi}, \tilde{\Psi})$ is optimal for original D.P.

Furthermore,

$$-\tilde{\Psi}(x^1) \geq \overbrace{\inf_{x^2 \in X} |x^1 - x^2| - \tilde{\Psi}(x^2)}^{\tilde{\varphi}(x^1)} \geq -\tilde{\Psi}(x^1)$$

$$\|\tilde{\Psi}\|_{Lip} \leq 1 \Rightarrow \tilde{\Psi}(x^2) - \tilde{\Psi}(x^1) \leq |x^1 - x^2|$$
$$-\tilde{\Psi}(x^1) \leq |x^1 - x^2| - \tilde{\Psi}(x^2)$$

Thus, $\tilde{\varphi} = -\tilde{\Psi}$.

$$\text{LHS} = \int \tilde{\varphi}\, d\mu + \int \tilde{\Psi}\, d\nu = \int \tilde{\varphi}\, d(\mu - \nu) \leq \text{RHS} \quad \square$$

We now have obtained characterizations of the dual problem for both W_1, W_2.

Exercise: Consider

$$W_p(\mu,\nu) := \min_{\gamma \in \Gamma(\mu,\nu)} \left(\int d(x^1,x^2)^p \, d\gamma(x^1,x^2) \right)^{1/p}.$$

Use Hölder's inequality to prove that

$$p \leq q \Rightarrow W_p(\mu,\nu) \leq W_q(\mu,\nu).$$

Furthermore, if $\mathrm{supp}(\mu), \mathrm{supp}(\nu)$ bounded, prove that $\exists$ C s.t.

$$p \leq q \Rightarrow W_q(\mu,\nu) \leq C W_p(\mu,\nu).$$

We will now prove our earlier claim that W_2 is a metric on $\mathcal{P}_2(\mathbb{R}^d)$.

Our proof will use Brenier's theorem... which requires $\mu \ll \mathcal{L}^d$.

We need a way to approximate $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ by μ_ε that are abs cts.

Polish space

<u>Prop</u>: (W_2 jointly lsc wrt narrow convergence): Suppose $\mu_n, \nu_n \in \mathcal{P}(X)$ satisfying $\mu_n \to \mu$ and $\nu_n \to \nu$ narrowly.

Then $\liminf_{n\to\infty} W_2(\mu_n, \nu_n) \geq W_2(\mu, \nu)$.

<u>Recall</u>:

• <u>Thm</u> (Portmanteau): For any $g: X \to \mathbb{R} \cup \{+\infty\}$ lsc and bounded below, $\mu_n \xrightarrow{\text{narrowly}} \mu$ implies $\liminf_{n\to\infty} \int g \, d\mu_n \geq \int g \, d\mu$.

Exercise: Suppose $\mu_n \to \mu$, $\nu_n \to \nu$ narrowly.
Then, for any $\gamma_n \in \Gamma(\mu_n, \nu_n)$, there exists a subsequence γ_{n_k} s.t. $\gamma_{n_k} \to \gamma$ narrowly, where $\gamma \in \Gamma(\mu, \nu)$.

Pf of Prop:

Choose a subsequence so that

$$\lim_{k\to+\infty} W_2(\mu_{n_k}, \nu_{n_k}) = \liminf_{n\to\infty} W_2(\mu_n, \nu_n).$$

Let γ_{n_k} be OT plans from μ_{n_k} to ν_{n_k}.

By exercise, there exists a further subsequence (denoted still by γ_{n_k}) s.t. $\gamma_{n_k} \to \gamma \in \Gamma(\mu, \nu)$.

$$\liminf_{n\to\infty} W_2(\mu_n, \nu_n) = \lim_{k\to+\infty} W_2(\mu_{n_k}, \nu_{n_k})$$
$$= \lim_{k\to+\infty} \Big(\int \underbrace{|x^1 - x^2|^2}_{g} d\gamma_{n_k}\Big)^{1/2}$$
$$\overset{\text{Portmanteau}}{\geq} \Big(\int |x^1 - x^2|^2 d\gamma\Big)^{1/2}$$

$$\geq W_2(\mu, \nu)$$

Approximation by Convolution

Def: (mollifier) $\varphi : \mathbb{R}^d \to [0, +\infty)$, bdd, meas
$\varphi(x) = \varphi(-x)$, $\int \varphi(x)\,dx = 1$
$\varphi_\varepsilon(x) = \frac{1}{\varepsilon^d}\varphi\left(\frac{x}{\varepsilon}\right)$ $\leftarrow \int \varphi_\varepsilon(x)\,dx = 1$

Def: Given $\mu \in \mathcal{P}(\mathbb{R}^d)$, define

$$\varphi_\varepsilon * \mu(x) = \int \varphi_\varepsilon(x-y)\,d\mu(y)$$

We will often abuse notation and write

$$\underbrace{d(\varphi_\varepsilon * \mu)(x)}_{\in \mathcal{P}(\mathbb{R}^d)} = \underbrace{\varphi_\varepsilon * \mu(x)}_{\text{density}}\,dx$$

Note: $\int \varphi_\varepsilon * \mu(x)\,dx$
$= \int\int \varphi_\varepsilon(x-y)\,d\mu(y)\,dx$
$= \int\int \varphi_\varepsilon(x-y)\,dx\,d\mu(y)$
$= \int 1\,d\mu(y)$
$= 1$

Lemma: Suppose $\mu \in \mathcal{P}(\mathbb{R}^d)$.
(i) For any $f : \mathbb{R}^d \to \mathbb{R}$ meas, bdd below,

$$\int f\,d(\underline{\varphi_\varepsilon * \mu}) = \int (\underline{\varphi_\varepsilon * f})\,d\mu$$

"associativity of convolution"

(ii) $\varphi_\varepsilon * \mu \to \mu$ narrowly, as $\varepsilon \to 0$.

Pf:

(i) $\exists$ $m \in \mathbb{R}$ s.t. $f \geq m$

$$\int f \, d(\varphi_\varepsilon * \mu) = \int (f-m) \, d(\varphi_\varepsilon * \mu) + \overbrace{\int m \, d(\varphi_\varepsilon * \mu)}^{m}$$

$$\int (\varphi_\varepsilon * f) \, d\mu = \int \varphi_\varepsilon * (f-m) \, d\mu + \underbrace{\int \varphi_\varepsilon * m \, d\mu}_{\underbrace{\int \varphi_\varepsilon(x-y) m \, dy = m}_{= m}}$$

By Tonelli,

$$\int (f-m) \, d(\varphi_\varepsilon * \mu) = \int\int (f(x)-m) \, \varphi_\varepsilon(x-y) \, d\mu(y) \, dx$$

$$= \int\int (f(x)-m) \, \varphi_\varepsilon(x-y) \, dx \, d\mu(y)$$

$$= \int \varphi_\varepsilon * (f-m) \, d\mu$$

(ii) Recall that for any $f \in C_b(\mathbb{R}^d)$, $\varphi_\varepsilon * f \to f$ uniformly on compact subsets of $\mathbb{R}^d$.

Fix $f \in C_b(\mathbb{R}^d)$.

By Prokhorov, since $\{\mu\}$ is tight, $\forall \delta > 0$, $\exists$ $K_\delta \subset\subset \mathbb{R}^d$ s.t.

$$\underline{2 \|f\|_\infty \, \mu(\mathbb{R}^d \setminus K_\delta) < \frac{\delta}{2}}.$$

Furthermore, $\exists\ \varepsilon_\delta > 0$ s.t. $0<\varepsilon<\varepsilon_\delta$,

$$\|\varphi_\varepsilon * f - f\|_{L^\infty(K_\delta)} < \frac{\delta}{2}$$

Thus,

$$\left|\int f\, d(\varphi_\varepsilon * \mu) - \int f d\mu\right|$$

$$= \left|\int (\varphi_\varepsilon * f - f)\, d\mu\right|$$

$$\leq \left|\int_{K_\delta} (\varphi_\varepsilon * f - f)\, d\mu\right| + \left|\int_{\mathbb{R}^d \setminus K_\delta} (\varphi_\varepsilon * f - f)\, d\mu\right|$$

$$\leq \|\varphi_\varepsilon * f - f\|_{L^\infty(K_\delta)} \int_{K_\delta} d\mu$$

$$+ \left(\|\varphi_\varepsilon * f\|_\infty + \|f\|_\infty\right) \mu(\mathbb{R}^d \setminus K_\delta)$$

$$\left|\int \varphi_\varepsilon(x-y) f(y)\, dy\right| \leq \|f\|_\infty \left|\int \varphi_\varepsilon(x-y)\, dy\right| = \|f\|_\infty$$

$$\leq \frac{\delta}{2} \cdot 1 + \frac{\delta}{2}$$

$$\leq \delta$$

Since f was arbitrary, $\varphi_\varepsilon * \mu \to \mu$ narrowly. $\square$

Now, we consider behavior of W_2 and convolution.

<u>Lemma</u>: Given $\mu \in \mathcal{P}(\mathbb{R}^d)$ and φ_ε as above,

$\int |x|^2 \varphi(x)\,dx$

(i) $W_2(\mu, \varphi_\varepsilon * \mu) \leq \varepsilon (M_2(\varphi))^{1/2}$

(ii) $W_2(\varphi_\varepsilon * \mu, \varphi_\varepsilon * \nu) \leq W_2(\mu, \nu)$

(iii) $\lim_{\varepsilon \to 0} W_2(\varphi_\varepsilon * \mu, \varphi_\varepsilon * \nu) = W_2(\mu, \nu)$

Pf: Omitted for lack of time :)

$M_2(\mu) := \int |x|^2 d\mu(x)$

Here is the proof from the last time I taught the course:

Pf: We begin with (i). Consider $\gamma_\varepsilon \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^d)$ defined by $d\gamma_\varepsilon(x_1, x_2) = \varphi_\varepsilon(x_2 - x_1)\,dx_2\,d\mu(x_1)$.

Then $\forall$ f bdd, meas,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} f(x_1)\,d\gamma_\varepsilon(x_1, x_2) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x_1)\varphi_\varepsilon(x_2 - x_1)\,dx_2\,d\mu(x_1)$$

$$= \int_{\mathbb{R}^d} f(x_1)\,d\mu(x_1)$$

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} f(x_2)\, d\gamma_\varepsilon(x_1,x_2) = \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x_2)\varphi_\varepsilon(x_2-x_1)\, dx_2\, d\mu(x_1)$$

$$= \int_{\mathbb{R}^d} f*\varphi_\varepsilon(x_1)\, d\mu(x_1)$$

$$= \int_{\mathbb{R}^d} f(x)\, d(\varphi_\varepsilon*\mu)(x)$$

Thus, $\gamma_\varepsilon \in \Gamma(\mu, \varphi_\varepsilon*\mu)$. So by defn of W_2,

$$W_2^2(\mu, \varphi_\varepsilon*\mu) \le \int |x_2-x_1|^2\, d\gamma_\varepsilon(x_1,x_2)$$

$$= \int |x_2-x_1|^2 \varphi_\varepsilon(x_2-x_1)\, dx_2\, d\mu(x_1)$$

$z = \frac{x_2-x_1}{\varepsilon}$

$dz = \frac{1}{\varepsilon^d} dx_2$

$$= \int |\varepsilon z|^2 \varphi(z)\, dz\, d\mu(x_1)$$

$$= \varepsilon^2 \int |z|^2 \varphi(z)\, dz$$

$$= \varepsilon^2 m_2(\varphi).$$

Now prove (ii). Define $\overline{\varphi_\varepsilon}(z_1,z_2) = \begin{cases} \varphi_\varepsilon(z_1) & \text{if } z_1 = z_2 \\ 0 & \text{otherwise} \end{cases}$

Consider $\gamma_\varepsilon \in \mathcal{M}(\mathbb{R}^d\times\mathbb{R}^d)$ defined by

$$\gamma_\varepsilon := \overline{\varphi}_\varepsilon * \gamma.$$

where γ is an optimal plan from μ to ν.

For any $f: \mathbb{R}^d \to \mathbb{R}$ bdd, meas, define
$\tilde{f}(x_1, x_2) = f(x_1)$.

Then,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} f(x_1) \, d\gamma_\varepsilon(x_1, x_2) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \tilde{f}(x_1, x_2) \, d(\overline{\varphi}_\varepsilon * \gamma)(x_1, x_2)$$

$$= \int_{\mathbb{R}^d \times \mathbb{R}^d} \tilde{f} * \overline{\varphi}_\varepsilon(x_1, x_2) \, d\gamma(x_1, x_2)$$

$$= \int_{\mathbb{R}^d \times \mathbb{R}^d} \int_{\mathbb{R}^d \times \mathbb{R}^d} \tilde{f}(x_1 - z_1, x_2 - z_2) \overline{\varphi}_\varepsilon(z_1, z_2) \, d\vec{z} \, d\gamma(x_1, x_2)$$

$$= \int_{\mathbb{R}^d \times \mathbb{R}^d} \int_{\mathbb{R}^d} f(x_1 - z) \varphi_\varepsilon(z) \, dz \, d\gamma(x_1, x_2)$$

$$= \int_{\mathbb{R}^d \times \mathbb{R}^d} f * \varphi_\varepsilon(x_1) \, d\gamma(x_1, x_2)$$

$$= \int_{\mathbb{R}^d} f * \varphi_\varepsilon(x_1) \, d\mu(x_1)$$

$$= \int f \, d(\varphi_\varepsilon * \mu)$$

Similarly, $\int f(x_2) \, d\gamma_\varepsilon(x_1, x_2) = \int f(x_2) \, d(\varphi_\varepsilon * \nu)(x_2)$

Thus, we conclude $\gamma_\varepsilon \in \Gamma(\varphi_\varepsilon * \mu, \varphi_\varepsilon * \nu)$.

Therefore,

$$
\begin{aligned}
W_2^2(\varphi_\varepsilon * \mu, \varphi_\varepsilon * \nu) &\le \int |x_1 - x_2|^2 \, d\bar{\varphi}_\varepsilon * \gamma(x_1, x_2) \\
&= \int \int |(x_1 - z_1) - (x_2 - z_2)|^2 \bar{\varphi}_\varepsilon(z_1, z_2) \, d\vec{z} \, d\gamma(x_1, x_2) \\
&= \int \int |(x_1 - z) - (x_2 - z)|^2 \varphi_\varepsilon(z) \, dz \, d\gamma(x_1, x_2) \\
&= \int \int |x_1 - x_2|^2 \varphi_\varepsilon(z) \, dz \, d\gamma(x_1, x_2) \\
&= W_2^2(\mu, \nu)
\end{aligned}
$$

Thus, $W_2(\varphi_\varepsilon * \mu, \varphi_\varepsilon * \nu) \le W_2(\mu, \nu)$.

Now part (iii).

By part (ii), $\limsup_{\varepsilon \to 0} W_2(\varphi_\varepsilon * \mu, \varphi_\varepsilon * \nu) \le W_2(\mu, \nu)$.

Since $\varphi_\varepsilon * \mu \to \mu$, $\varphi_\varepsilon * \nu \to \nu$ narrowly and W_2 is jointly lsc wrt narrow convergence,

$$\liminf_{\varepsilon \to 0} W_2(\varphi_\varepsilon * \mu, \varphi_\varepsilon * \nu) \geq W_2(\mu, \nu).$$

Combining these two limits gives the result.