# Lecture 13

Reminders
- Solutions for 2-3 exercises
- First wiki article due Fri Feb 14
- Select article to revise by Fri, Feb 21

## Recall:

**Thm (Brenier):** Given $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, suppose $\mu \ll \mathcal{L}^d$.

① If $\gamma_*$ is an OT plan, $\exists$ $t$ meas s.t. $\gamma_* = (id \times t)\#\mu$.

"Any OT plan is induced by a transport map"

② $t$ is an OT map $\Leftrightarrow$ $(id \times t)\#\mu$ is an OT plan

③ Given $t$ s.t. $t\#\mu = \nu$,
$t$ is optimal $\Leftrightarrow$ $t = \nabla\varphi$ ($\mu$-a.e.) for $\varphi \in L^1(\mu)$ convex, lsc

④ the OT map from $\mu$ to $\nu$ is unique ($\mu$-a.e.)

More generally...

<u>Thm</u>: Given $\mathcal{X} \subseteq \mathbb{R}^d$ compact, $\mu, \nu \in \mathcal{P}(\mathcal{X})$, $c(x,y) = h(x-y)$ for $h$ strictly convex...

- $\exists$ OT plan $\gamma_*$
- if $\mu \ll \mathcal{L}^d$, then $\gamma_*$ is unique and $\gamma_* = (id \times t)\#\mu$ for

OT map

$$t(x) = x - \nabla h^{-1}(\nabla \varphi(x))$$

Pf: Santambrogio, Thm 1.17

One last important application of Kantorovich duality...

$$\mathbb{K}_p(\gamma) := \int d(x,y)^p \, d\gamma(x,y)$$

<u>Thm</u>: Given $\mathcal{X}$ cpt Polish, $\mu, \nu \in \mathcal{P}(\mathcal{X})$, $c(x,y) = d(x,y)$

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

and $\exists\, \varphi^*$ that achieves max on RHS.

<u>Remark</u>: The first part of the theorem continues to hold on any Polish space $\mathcal{X}$, under the additional constraint $\varphi \in L^1(|\mu - \nu|)$

Pf: By Kantorovich duality, it suffices to show

$$\sup_{\substack{\varphi \in C(X) \\ \|\varphi\|_{Lip} \leq 1}} \int \varphi \, d(\mu - \nu) = \sup_{\substack{\varphi, \psi \in C(X) \\ \varphi(x) + \psi(y) \leq d(x,y)}} \int \varphi \, d\mu + \int \psi \, d\nu$$

Last time "$\leq$". Now show "$\geq$".

Take $(\varphi_*, \psi_*)$ that attain max on RHS.

Define $\tilde{\psi}(y) = \inf_x d(x,y) - \varphi_*(x)$.

- $\tilde{\psi} \geq \psi_*$
- $\varphi_*(x) + \tilde{\psi}(y) \leq d(x,y)$
- $\tilde{\psi}$ is Moreau-Yosida regularization of $-\varphi_*$, so Exercise 9 ensures $\|\tilde{\psi}\|_{Lip} \leq 1$.

Thus, $(\varphi_*, \tilde{\psi})$ is a maximizer.

Define $\tilde{\varphi}(x) = \inf_y d(x,y) - \tilde{\psi}(y)$.
Then, as before, $(\tilde{\varphi}, \tilde{\psi})$ is a maximizer.

In particular, $\tilde{\varphi}(x) + \tilde{\psi}(y) \leq d(x,y)$ $\forall x,y$,
so $\tilde{\varphi}(x) \leq -\tilde{\psi}(x)$.

Since $\|\tilde{\psi}\|_{Lip} \leq 1$, $d(x,y) - \tilde{\psi}(y) \geq -\tilde{\psi}(x)$,
so $\tilde{\varphi}(x) \geq -\tilde{\psi}(x)$. Thus $\tilde{\varphi} = -\tilde{\psi}$.

This shows "$\geq$" and that an
optimizer of LHS exist. $\square$

## p-Wasserstein metrics

Def: Given $(X,d)$ Polish, $\mu, \nu \in \mathcal{P}(X)$, $p \geq 1$,

$$W_p(\mu,\nu) := \min_{\gamma \in \Gamma(\mu,\nu)} \left(\mathbb{K}_p(\gamma)\right)^{1/p}.$$

Rmk:

- By Jensen's inequality, for $p \leq q$,

$$\mathbb{K}_p(\gamma)^{1/p} = \mathbb{K}_p(\gamma)^{\frac{1}{p}\frac{q}{q}} \leq \left(\int \left(d^p(x,y)\right)^{q/p} d\gamma\right)^{1/q} = \mathbb{K}_q(\gamma)^{1/q}$$

Thus, $W_p(\mu,\nu) \leq W_q(\mu,\nu)$.

- If $\operatorname{diam}(X) = \sup_{x,y} d(x,y) < +\infty$, then for all $p \leq q$,

$$\mathbb{K}_q(\gamma) = \int d^q(x,y)\,d\gamma$$
$$\leq \int d^p(x,y)\, \mathrm{diam}(\mathcal{X})^{q-p}\,d\gamma$$

Hence $W_q(\mu,\nu) \leq \mathrm{diam}(\mathcal{X})^{1-p/q}\, W_p(\mu,\nu)^{p/q}$

**Goal:** Prove $(W_p, \mathcal{P}_p(\mathcal{X}))$ is a metric space.

$\mathcal{P}_p(\mathcal{X}) := \{\mu \in \mathcal{P}(\mathcal{X}) : \overbrace{\int d^p(x,x_0)\,d\mu(x)}^{=:M_p(\mu)} < +\infty$ for some $x_0 \in \mathcal{X}\}$

To prove our goal, we recall...

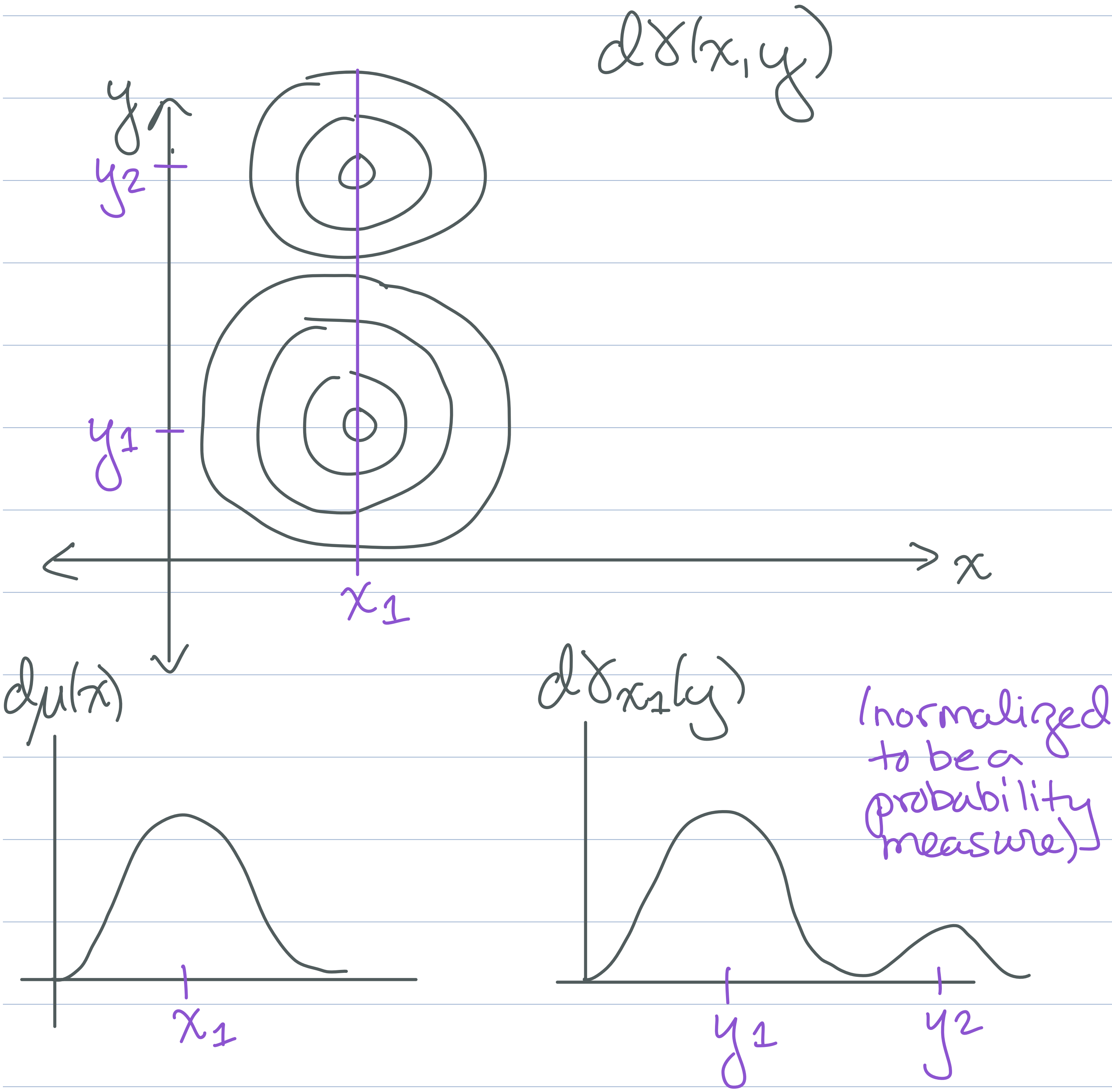
Thm (Disintegration): Given Polish spaces $\mathcal{X}, Y$ and $\gamma \in \mathcal{P}(\mathcal{X} \times Y)$, let $\mu := \pi_{\mathcal{X}}\#\gamma$. Then there exists unique ($\mu$-a.e.)

$$\{\gamma_x\}_{x\in\mathcal{X}} \subseteq \mathcal{P}(Y)$$

s.t., $\forall$ meas $f : \mathcal{X} \times Y \to [0,+\infty]$,

- $x \mapsto \int_Y f(x,y)\,d\gamma_x(y)$ is meas
- $\int_{\mathcal{X}\times Y} f(x,y)\,d\gamma(x,y) = \int_{\mathcal{X}}\int_Y f(x,y)\,d\gamma_x(y)\,d\mu(x)$

Pf: Bogachev, section 10.6.



Pf: Bogachev, section 10.6.

Ex: If $d\gamma(x,y) = f(x,y)\,d\mathcal{L}^m(x)\,d\mathcal{L}^n(y)$, then $d\mu(x) = g(x)\,d\mathcal{L}^m(x)$ for

$$g(x) := \int f(x,y)\,d\mathcal{L}^n(y)$$

and

$$d\gamma_x(y) = \frac{f(x,y)\,d\mathcal{L}^n(y)}{g(x)}$$

(well-defined $\mu$-a.e.)

Cor: Given Polish space $\mathcal{X}$, $\gamma^{12} \in \mathcal{P}(\mathcal{X}\times\mathcal{X})$, $\gamma^{23} \in \mathcal{P}(\mathcal{X}\times\mathcal{X})$ s.t.

$$\pi_2\#\gamma^{12} = \pi_1\#\gamma^{23} =: \mu.$$

Then $\exists\ \gamma \in \mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{X})$ s.t.

$$\pi_{1,2}\#\gamma = \gamma^{12}, \quad \pi_{2,3}\#\gamma = \gamma^{23}.$$

Pf: Let

$$d\gamma(x_1,x_2,x_3) := \left(d\gamma^{12}_{x_2}(x_1)\otimes d\gamma^{23}_{x_2}(x_3)\right) d\mu(x_2)$$

Then $\forall\, f(x_1,x_2)\geq 0$, meas,

$$\int_{X\times X\times X} f(x_1,x_2)\, d\gamma(x_1,x_2,x_3)$$

$$= \int_X \left(\int_{X\times X} f(x_1,x_2)\, d\gamma^{12}_{x_2}(x_1)\otimes d\gamma^{23}_{x_2}(x_3)\right) d\mu(x_2)$$

$$= \int_X \left(\int_X f(x_1,x_2)\, d\gamma^{12}_{x_2}(x_1)\cdot \overbrace{\int_X 1\, d\gamma^{23}_{x_2}(x_3)}^{1}\right) d\mu(x_2)$$

$$= \int_{X\times X} f(x_1,x_2)\, d\gamma^{12}(x_1,x_2)$$

Thus $\pi_{1,2\#}\gamma = \gamma^{12}$. Likewise $\pi_{2,3\#}\gamma = \gamma^{2,3}$. $\square$

Thm: Suppose $(X,d)$ is a Polish space. Then $\forall p \geq 1$, $(W_p, \mathcal{P}_p(X))$ is a metric space.

Fact: $d^p(x,y) \leq (d(x,x_0)+d(x_0,y))^p$
$\leq 2^{p-1}(d^p(x,x_0)+d^p(x_0,y))$

Pf:

for any $\gamma \in \Gamma(\mu,\nu)$

By Fact, $W_p^p(\mu,\nu) \leq \int d^p(x,y)\, d\gamma(x,y)$
$\leq 2^{p-1}\int d^p(x,x_0) + d^p(x_0,y)\, d\gamma$
$= 2^{p-1}(m_p(\mu)+m_p(\nu)) < +\infty$

Thus, $W_p : \mathcal{P}_p(X) \times \mathcal{P}_p(X) \to [0,+\infty)$.
By defn, it is symmetric.

To see $W_p$ is nondegenerate...

Suppose $\mu = \nu$. Then, $\gamma := (id \times id)_{\#}\mu$ is a plan from $\mu$ to $\nu$, and $\mathbb{K}_p(\gamma) = \int d^p(x,y)\, d\gamma = \int d^p(x,x)\, d\mu = 0$. Thus $W_p(\mu,\nu)^p = 0$.

For the converse, suppose $W_p(\mu,\nu) = 0$. Then, if $\gamma$ is the OT from $\mu$ to $\nu$, $0 = \mathbb{K}_p(\gamma) = \int d^p(x,y)\, d\gamma(x,y)$, so $x = y$ $\gamma$-a.e. Hence, $\forall f \geq 0$ meas,

$$\begin{array}{ccc} \int f(x)\, d\gamma(x,y) & = & \int f(y)\, d\gamma(x,y) \\ \shortparallel & & \shortparallel \\ \int f(x)\, d\mu(x) & & \int f(y)\, d\nu(y) \end{array}$$

Thus, $\mu = \nu$.

It remains to show triangle ineq.

Fix $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_p(\mathcal{X})$.

Suppose $\gamma_{12}$ is an OT plan from $\mu_1$ to $\mu_2$

$\gamma_{23}$ " " $\mu_2$ to $\mu_3$

By corollary, $\exists\, \gamma \in \mathcal{P}(\mathcal{X}\times\mathcal{X}\times\mathcal{X})$ s.t.

$\pi_{1,2}\#\gamma = \gamma_{12}$ and $\pi_{2,3}\#\gamma = \gamma_{2,3}$.

By def $\pi_1\#\gamma = \pi_1\#\gamma_{12} = \mu_1$

$\pi_3\#\gamma = \pi_3\#\gamma_{2,3} = \mu_3$,

so $\pi_{1,3}\#\gamma \in \Gamma(\mu_1, \mu_3)$.

$$W_p(\mu_1,\mu_3) \le \left(\int_{\mathcal{X}\times\mathcal{X}} d^p(x_1,x_3)\, d\pi_{1,3}\#\gamma\right)^{1/p}$$

$$= \left(\int_{\mathcal{X}\times\mathcal{X}\times\mathcal{X}} d^p(x_1,x_3)\, d\gamma\right)^{1/p}$$

$$= \| d(\pi_1, \pi_3) \|_{L^p(\gamma)}$$

$$\leq \| d(\pi_1,\pi_2) + d(\pi_2,\pi_3) \|_{L^p(\gamma)}$$
$$\leq \| d(\pi_1,\pi_2)\|_{L^p(\gamma)} + \| d(\pi_2,\pi_3)\|_{L^p(\gamma)}$$
$$= \left(\mathbb{K}_p(\pi_{1,2}\#\gamma)\right)^{1/p} + \left(\mathbb{K}_p(\pi_{2,3}\#\gamma)\right)^{1/p}$$
$$= W_p(\mu_1,\mu_2) + W_p(\mu_2,\mu_3) \qquad \square$$

Next goal: characterize topology of $W_p$

Recall from Lectures 4 and 5:

Thm (Prokhorov): Given $X$ Polish, $\mathcal{K} \subseteq \mathcal{P}(X)$, $\mathcal{K}$ is relatively narrowly cpt $\Leftrightarrow$ $\underbrace{\mathcal{K} \text{ tight}}$.

$\forall \varepsilon > 0$, $\exists K_\varepsilon \subset\subset X$ s.t. $\mu(X \setminus K_\varepsilon) \leq \varepsilon$ $\forall \mu \in \mathcal{K}$

Thm (Portmanteau): For $g : X \to \mathbb{R} \cup \{+\infty\}$ lsc and bdd below, $\mu \mapsto \int_X g \, d\mu$ is lsc wrt narrow convergence.

Exercise 29: Suppose $\mu_n \to \mu$ and $\nu_n \to \nu$ narrowly and $\gamma_n \in \Gamma(\mu_n, \nu_n)$. Then $\exists\ \gamma_{n_k}$ s.t. $\gamma_{n_k} \to \gamma \in \Gamma(\mu, \nu)$ narrowly.