Lecture 17, Wed., 5/4/2022

- An elementary proof of the triangle inequality for the W_2-metric.
- The metric $\beta(\mu,\nu)$ on $\mathcal{P}(X)$.

<u>Theorem</u> Let X, Y be Polish spaces. Then $(\mathcal{P}_2(X), W_2)$ is a metric space.

An elementary proof of the triangle ineq. (Clement & Desch 2008).

Ideas: First, consider X to be a countable set. The proof is the same as that for the discrete OT case. Next, for a separable X, consider approximations.

<u>Lemma 1</u> Let (X,d) be a countable metric space. Let $\mu^i \in \mathcal{P}_2(X)$ $(i=1,2,3)$, $\gamma^{1,2} \in \mathcal{A}(\mu^1,\mu^2)$, and $\gamma^{2,3} \in \mathcal{A}(\mu^2,\mu^3)$. Then $\exists\, \gamma \in \mathcal{P}(X\times X\times X)$ such that $\pi^{1,2}_{\#}\gamma = \gamma^{1,2}$ and $\pi^{2,3}_{\#}\gamma = \gamma^{2,3}$. Moreover, $\gamma^{1,3} = \pi^{1,3}_{\#}\gamma \in \mathcal{A}(\mu^1,\mu^3)$ and

$$\left[\int_{X\times X} d^2(x_1,x_3)\,d\gamma^{1,3}(x_1,x_3)\right]^{1/2} \le \left[\int_{X\times X} d^2(x_1,x_2)\,d\gamma^{1,2}(x_1,x_2)\right]^{1/2} + \left[\int_{X\times X} d^2(x_2,x_3)\,d\gamma(x_2,x_3)\right]^{1/2}.$$

<u>Proof</u> Let $X = \{v_1, v_2, \dots\}$. Denote $\mu^i_k = \mu^i(\{v_k\})$, $\gamma^{i,j}_{k,\ell} = \gamma^{i,j}(\{v_k, v_\ell\})$ $(i,j \in \{1,2,3\},\ k,\ell \in \mathbb{N})$. Define

$$\gamma = \sum_{k,m,n} \gamma_{k,m,n}\,(\delta_{v_k}\times\delta_{v_m}\times\delta_{v_n}), \qquad \gamma_{k,m,n} = \begin{cases} \dfrac{\gamma^{1,2}_{k,m}\,\gamma^{2,3}_{m,n}}{\mu^2_m} & \text{if } \mu^2_m \ne 0, \\ 0 & \text{if } \mu^2_m = 0. \end{cases}$$

Since $\pi^2_{\#}\gamma^{2,3} = \mu^2$,

$$\pi^{1,2}_{\#}\gamma(\{(v_k,v_m)\}) = \begin{cases} \sum_n \dfrac{\gamma^{1,2}_{k,m}\,\gamma^{2,3}_{m,n}}{\mu^2_m} = \gamma^{1,2}_{k,m} & \text{if } \mu^2_m \ne 0, \\ 0 = \gamma^{1,2}_{k,m} & \text{if } \mu^2_m \ne 0. \end{cases}$$

Hence, $\pi^{1,2}_{\#}\gamma = \gamma^{1,2}$. Similarly, $\pi^{2,3}_{\#}\gamma = \gamma^{2,3}$. Since both $\gamma^{1,2}$ and $\gamma^{2,3}$ are probability measures, γ is, too.

Similarly, $\gamma^{1,3} = \pi^{1,3}_{\#}\gamma \in \mathcal{P}(X_1 \times X_3)$ and further $\gamma^{1,3} \in \mathcal{A}(\mu^1, \mu^3)$. Finally, by Minkowski's inequality,

$$\left[\int_{X\times X} d(x_1,x_3)^2 \, d\gamma^{1,3}(x_1,x_3)\right]^{1/2} = \left[\int_{X\times X\times X} d^2(x_1,x_3)\, d\gamma(x_1,x_2,x_3)\right]^{1/2}$$

$$\leq \left[\int_{X\times X\times X} \left(d(x_1,x_2)+d(x_2,x_3)\right)^2 d\gamma(x_1,x_2,x_3)\right]^{1/2}$$

$$\leq \left[\int_{X\times X\times X} d^2(x_1,x_2)\, d\gamma(x_1,x_2,x_3)\right]^{1/2} + \left[\int_{X\times X\times X} d^2(x_2,x_3)\, d\gamma(x_1,x_2,x_3)\right]^{1/2}$$

$$= \left[\int_{X\times X} d^2(x_1,x_2)\, d\gamma^{1,2}(x_1,x_2)\right]^{1/2} + \left[\int_{X\times X} d^2(x_2,x_3)\, d\gamma^{2,3}(x_2,x_3)\right]^{1/2}.$$

QED

If X is separable and $\tilde{X} = \{v_1, v_2, \dots\}$ is a countable dense subset of X, then for any $\varepsilon > 0$, we set $S_1 = B(x_1;\varepsilon)$ and $S_i = B(v_i,\varepsilon) \setminus \bigcup_{j<i} S_j$ $(i=2,3,\cdots)$. Then $\{S_i\}_1^\infty$ is a partition of X. Define $f: X \to \tilde{X}$ by $f(x) = v_i$ if $x \in S_i$. Then f is continuous and surjective, $S_i = f^{-1}(\{v_i\})$, and $d(x, f(x)) < \varepsilon$ $\forall x \in X$.

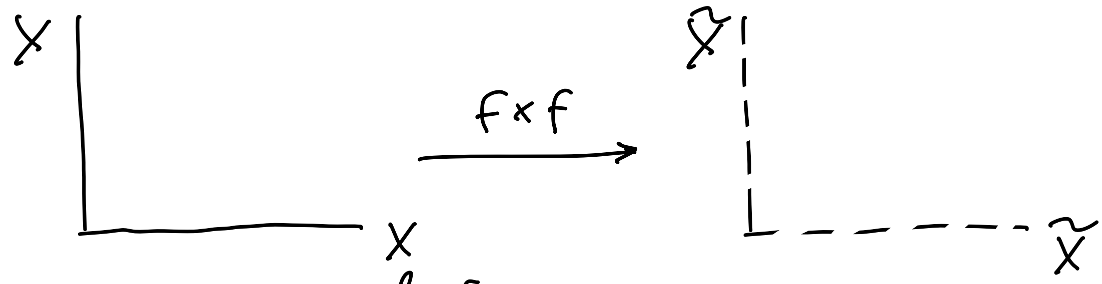
Lemma 2 Let (X,d) be a separable metric space and $\tilde{X} = \{v_1, v_2, \dots\}$ is a dense subset of X. Let $\varepsilon > 0$.

(1) There exists $f: X \to \tilde{X}$, Borel measurable, such that $d(x, f(x)) < \varepsilon$ $\forall x \in X$, and that $S_i = f^{-1}(v_i)$ $(i=1,2,\cdots)$ form a partition of X.

(2) If $\gamma \in \mathcal{P}(X\times X)$ and $\tilde{\gamma} := (f\times f)_{\#}\gamma \in \mathcal{P}(\tilde{X}\times\tilde{X})$ then

(2.1) $\pi^i_{\#}\tilde{\gamma} = f_{\#}(\pi^i_{\#}\gamma)$, $i=1,2$, and

(2.2) $$\left|\left[\int_{X\times X} d^2(x,y)\, d\gamma(x,y)\right]^{1/2} - \left[\int_{\tilde{X}\times\tilde{X}} d^2(x,y)\, d\tilde{\gamma}(x,y)\right]^{1/2}\right| \leq 2\varepsilon.$$



$$\begin{tikzcd}
\gamma \in \mathcal{P}(X\times X) \arrow[r, "f\times f"] \arrow[d, "\pi^i \text{ for } X"'] & \tilde\gamma = (f\times f)_\# \gamma \in \mathcal{P}(\tilde X\times \tilde X) \arrow[d, "\pi^i \text{ for } \tilde X"] \\
\pi^i_\# \gamma \in \mathcal{P}(X) \arrow[r, "f"] \arrow[d] & f_\#(\pi^i_\# \gamma) = \pi^i_\# \tilde\gamma \in \mathcal{P}(\tilde X) \arrow[d] \\
\text{marginals of } \gamma \arrow[r] & \text{marginals of } \tilde\gamma
\end{tikzcd}$$

<u>Proof</u> (1) Define $S_1 = B(v_1, \varepsilon)$, $S_i = B(v_i, \varepsilon) \setminus \bigcup_{j<i} S_j$ $(i = 2, 3, \cdots)$. Then, $\{S_i\}_1^\infty$ is a partition of X. Set $f(x) = v_i$ if $x \in S_i$.

(2) Let $\tilde U \in \tilde X$. Then

$$(\pi^1_\# \tilde\gamma)(\tilde U) = \tilde\gamma(\tilde U \times \tilde X) = \big((f\times f)_\# \gamma\big)(\tilde U \times \tilde X) = \gamma\big((f\times f)^{-1}(\tilde U \times \tilde X)\big)$$
$$= \gamma(f^{-1}(\tilde U) \times X) = (\pi^1_\# \gamma)(f^{-1}(\tilde U)) = [f_\#(\pi^1_\# \gamma)](\tilde U).$$

Hence, $\pi^1_\# \tilde\gamma = f_\#(\pi^1_\# \gamma)$. Similarly, $\pi^2_\# \tilde\gamma = f_\#(\pi^2_\# \gamma)$. (2.1) is proven. To prove (2.2), we note $\tilde\gamma = (f\times f)_\# \gamma$, and use the change of variable formula and the Minkowski's ineq.:

$$\left| \left[\int_{X\times X} d^2(x,y)\, d\gamma(x,y)\right]^{1/2} - \left[\int_{X\times X} d^2(x,y)\, d\tilde\gamma(x,y)\right]^{1/2} \right|$$
$$= \left| \left[\int_{X\times X} d^2(x,y)\, d\gamma(x,y)\right]^{1/2} - \left[\int_{X\times X} d^2(f(x), f(y))\, d\gamma(x,y)\right]^{1/2} \right|$$
$$\le \left[\int_{X\times X} |d(x,y) - d(f(x), f(y))|^2\, d\gamma(x,y)\right]^{1/2}$$
$$\le \left[\int_{X\times X} (2\varepsilon)^2\, d\gamma(x,y)\right]^{1/2} = 2\varepsilon. \quad \underline{QED}$$

<u>Theorem</u> Let X be a separable metric space and $\mu^i \in \mathcal{P}_2(X)$ $(i = 1, 2, 3)$. Then $W_2(\mu^1, \mu^3) \le W_2(\mu^1, \mu^2) + W_2(\mu^2, \mu^3)$.

Remark $W_2(\cdot,\cdot)$ is defined by inf instead of min, since X is not necessary complete.

Proof Let $\tilde{X}=\{v_1,v_2,\dots\}$ be a dense subset of X. Let $\varepsilon>0$ and let $f: X\to\tilde{X}$ be given as in Lemma 2. Let $S_i=f^{-1}(\{v_i\})$. Define $\tilde{\mu}^i=f_\#\mu^i$ $(i=1,2,3)$. For $(i,j)=(1,2)$ or $(2,3)$, let $\gamma^{i,j}\in\mathcal{A}(\mu^i,\mu^j)$ be such that

$$\left[\int_{X\times X} d^2(x_i,x_j)\,d\gamma^{i,j}(x_i,x_j)\right]^{1/2}<W_2(\mu^i,\mu^j)+\varepsilon.$$

Let $\tilde{\gamma}^{i,j}=(f\times f)_\#\gamma^{i,j}$. By Lemma 2, $\tilde{\gamma}^{i,j}\in\mathcal{A}(\mu^i,\mu^j)$, and

$$\left[\int_{\tilde{X}\times\tilde{X}} d^2(x_i,x_j)\,d\tilde{\gamma}^{i,j}(x_i,x_j)\right]^{1/2}\le\left[\int_{X\times X} d^2(x_i,x_j)\,d\gamma^{i,j}(x_i,x_j)\right]^{1/2}+2\varepsilon$$
$$\le W_2(\mu^i,\mu^j)+3\varepsilon.$$

By Lemma 1, $\exists\,\tilde{\gamma}^{1,3}\in\mathcal{A}(\tilde{\mu}^1,\tilde{\mu}^3)$ s.t.

$$\left[\int_{\tilde{X}\times\tilde{X}} d^2(x_1,x_3)\,d\tilde{\gamma}^{1,3}(x_1,x_3)\right]^{1/2}\le\left[\int_{\tilde{X}\times\tilde{X}} d^2(x_1,x_2)\,d\tilde{\gamma}^{1,2}(x_1,x_2)\right]^{1/2}+\left[\int_{\tilde{X}\times\tilde{X}} d^2(x_2,x_3)\,d\tilde{\gamma}^{2,3}(x_2,x_3)\right]^{1/2}$$
$$\le W_2(\mu^1,\mu^2)+W_2(\mu^2,\mu^3)+6\varepsilon.$$

Now, write $\tilde{\gamma}^{1,3}_{kn}=\tilde{\gamma}^{1,3}(\{(v_k,v_n)\})$ and $\tilde{\mu}^i_m=\tilde{\mu}^i(\{v_m\})=\mu^i(S_m)$.

Define the measure $\gamma^{1,3}$ on $X\times X$ by

$$\gamma^{1,3}(U)=\sum_{k,n\in\mathbb{N},\ \tilde{\mu}^1_k\neq0,\ \tilde{\mu}^3_n\neq0}\frac{\tilde{\gamma}^{1,3}_{k,n}}{\tilde{\mu}^1_k\,\tilde{\mu}^3_n}(\mu^1\times\mu^3)\big(U\cap(S_k\times S_n)\big).$$

It will be proved below $\gamma^{1,3}\in\mathcal{A}(\mu^1,\mu^3)$ and $(f\times f)_\#\gamma^{1,3}=\tilde{\gamma}^{1,3}$. By Lemma 2, we have

$$\left[\int_{X\times X} d^2(x_1,x_3)\,d\gamma(x_1,x_3)\right]^{1/2}\le\left[\int_{\tilde{X}\times\tilde{X}} d^2(x_1,x_3)\,d\tilde{\gamma}(x_1,x_3)\right]^{1/2}+2\varepsilon$$
$$\le W_2(\mu^1,\mu^2)+W_2(\mu^2,\mu^3)+8\varepsilon.$$

Hence, $W_2(\mu^1,\mu^3)\le W_2(\mu^1,\mu^2)+W_2(\mu^2,\mu^3)+6\varepsilon$.
Sending $\varepsilon\to0$, we get the desired ineq.

We show now $\gamma^{1,3} \in \mathcal{A}(\mu^1, \mu^3)$ by the definition of $\gamma^{1,3}$ and marginals of $\tilde{\gamma}^{1,3}$: (assume nonzero denomenaters, otherwise $\mu_k^1 = 0$ or $\mu_n^3 = 0 \Rightarrow \gamma_{k,n}^{1,3} = 0$)

$$\pi^1_{\#}\gamma^{1,3}(V) = \gamma^{1,3}(V \times X) = \sum_{k,n=1}^{\infty} \frac{\tilde{\gamma}^{1,3}_{k,n}}{\tilde{\mu}^1_k \tilde{\mu}^3_n} (\mu^1 \times \mu^3)\big((V \cap S_k) \times S_n\big)$$

$$= \sum_{k,n=1}^{\infty} \frac{\tilde{\gamma}^{1,3}_{k,n}}{\tilde{\mu}^1_k} \mu^1(V \cap S_k) = \sum_{k=1}^{\infty} \frac{1}{\tilde{\mu}^1_k} \mu^1(V \cap S_k) \sum_{n=1}^{\infty} \tilde{\gamma}^{1,3}_{k,n}$$

$$= \sum_{k=1}^{\infty} \mu^1(V \cap S_k) = \mu^1(V).$$

Let $\tilde{U} \subseteq \tilde{X} \times \tilde{X}$. Then $(f \times f)^{-1}(\tilde{U}) = \bigcup_{k,n:\,(v_k, v_n) \in \tilde{U}} S_k \times S_n$. So,

$$(f \times f)_{\#}\gamma^{1,3}(\tilde{U}) = \gamma^{1,3}\Big(\bigcup_{k,n:\,(v_k,v_n) \in \tilde{U}} S_k \times S_n\Big)$$

$$= \sum_{k,n:\,(v_k,v_n) \in \tilde{U}} \frac{\tilde{\gamma}^{1,3}_{k,n}}{\tilde{\mu}^1_k \tilde{\mu}^3_n} (\mu^1 \times \mu^3)(S_k \times S_n)$$

$$= \sum_{k,n:\,(v_k,v_n) \in \tilde{U}} \tilde{\gamma}^{1,3}_{k,n} = \tilde{\gamma}^{1,3}(\tilde{U}). \qquad \underline{\text{QED}}$$

We now introduce the β metric of $\mathcal{P}(X)$. This will be used for proving some of the main results about the space $(\mathcal{P}_2(X), W_2)$.

Let X be a metric space. Denote $BL(X) = \{$bounded, Lipschitz-continuous, and real-valued functions φ on X with $\|\varphi\|_{BL} < \infty\}$, where $\|\varphi\|_{BL} = \|\varphi\|_\infty + [\varphi]_L$ and $[\varphi]_L = \sup_{\substack{x,y \in X \\ x \neq y}} \frac{|\varphi(x) - \varphi(y)|}{d(x,y)}$.

Proposition For any metric space X, $(BL(X), \|\cdot\|_{BL})$ is a normed vector space. QED

Let X be a metric space. Let $\mu, \nu \in \mathcal{P}(X)$. Define

$$\beta(\mu,\nu) = \sup\left\{ \left| \int_X f\,d\mu - \int_X f\,d\nu \right| : f \in BL(X),\ \|f\|_{BL} \le 1 \right\}.$$

Theorem If X is a metric space, then β is a metric on $\mathcal{P}(X)$.

Proof Clearly, $\beta(\mu,\nu) \ge 0$ $\forall \mu,\nu \in \mathcal{P}(X)$ $\mu = \nu \Rightarrow \beta(\mu,\nu) = 0$. Suppose $\beta(\mu,\nu) = 0$. For any closed $F \subseteq X$, let $g(x) = \mathrm{dist}(x, F)$. Then $g \in BL(X)$ and $[g] \le 1$. Let $g_k = \min(kg, 1)$ $(k = 1, 2, \dots)$. Then $g_k \in BL(X)$ and $\|g_k\|_{BL} \le k+1$. Let $U = X \setminus F$. Since $g_k \uparrow \mathbb{1}_U$, and $\beta(\mu,\nu) = 0$, $\mu(U) = \nu(U)$ and $\mu(F) = \nu(F)$. Hence $\mu = \nu$. It is clear that $\beta(\mu,\nu) = \beta(\nu,\mu)$, and $\beta(\mu,\nu) \le \beta(\mu,\sigma) + \beta(\sigma,\nu)$ for any $\mu, \nu, \sigma \in \mathcal{P}(X)$. QED

Theorem If X is a separable metric space and $\mu_k, \mu \in \mathcal{P}(X)$ $(k = 1, 2, \dots)$, then the following are equivalent:

(1) $\mu_k \to \mu$ narrowly;

(2) $\int_X f\,d\mu_k \to \int_X f\,d\mu$ $\forall f \in BL(X)$;

(3) $\beta(\mu_k, \mu) \to 0$.

Proof See Dudley's book. QED

Theorem If X is a Polish space, then $(\mathcal{P}(X), \beta)$ is complete.

Proof See Dudley's book. QED

Proposition. If X is Polish, $\mu, \nu \in \mathcal{P}_2(X)$, then

$$\beta(\mu,\nu) \le W_2(\mu,\nu).$$

Proof Let $\mu, \nu \in \mathcal{P}(X)$ and $\gamma \in \mathcal{A}(\mu,\gamma)$. Then, for any $f \in BL(X)$ with $[f] \le 1$, we have

$$\left|\int_X f\,d\mu - \int_X f\,d\nu\right| = \left|\int_{X\times X} f(x)\,d\gamma(x,y) - \int_{X\times X} f(y)\,d\gamma(x,y)\right|$$
$$\le \int_{X\times X} |f(x)-f(y)|\,d\gamma(x,y) \le \int_{X\times X} d(x,y)\,d\gamma(x,y)$$
$$\le \left[\int_{X\times X} d^2(x,y)\,d\gamma(x,y)\right]^{1/2}.$$

Taking supremum over $f\in BL(X)$ with $\|f\|_{BL}\le 1$, we get
$\beta(\mu,\nu) \le \left[\int_{X\times X} d^2(x,y)\,d\gamma(x,y)\right]^{1/2}$. Hence, $\beta(\mu,\nu)\le W_2(\mu,\nu)$. QED

Proposition Let (X,d) be a Polish space and $\mu,\nu\in\mathcal{P}_2(X)$. Then, for any $x_0\in X$,
$$W_2^2(\mu,\nu) \le 2\int_X d^2(x,x_0)\,d|\mu-\nu|(x) = 2\,\|d^2(x_0,\cdot)(\mu-\nu)\|_{TV}.$$

Proof Define
$$\gamma = (Id\times Id)_\#(\mu\wedge\nu) + \frac{1}{(\mu-\nu)^+(X)}(\mu-\nu)^+\otimes(\mu-\nu)^-.$$
where $\mu\wedge\nu = \mu-(\mu-\nu)^+$. One can verify that $\gamma\in\Gamma(\mu,\nu)$. Moreover,
$$W_2^2(\mu,\nu) \le \int_{X\times X} d^2(x,y)\,d\gamma(x,y)$$
$$= \frac{1}{(\mu-\nu)^+(X)}\int_{X\times X} d^2(x,y)\,d(\mu-\nu)^+(x)\,d(\mu-\nu)^-(y)$$
$$\le \frac{2}{(\mu-\nu)^+(X)}\int_{X\times X}\left[d^2(x,x_0)+d^2(x_0,y)\right]d(\mu-\nu)^+(x)\,d(\mu-\nu)^-(y)$$
$$\le 2\left[\int_{X\times X} d^2(x,x_0)\,d(\mu-\nu)^+(x) + \int_{X\times X} d^2(x_0,y)\,d(\mu-\nu)^-(y)\right.$$
$$= 2\int_X d^2(x,x_0)\,d\left[(\mu-\nu)^+ + (\mu-\nu)^-\right](x)$$
$$= 2\int_X d^2(x,x_0)\,d|\mu-\nu|(x). \quad \text{QED}$$