Lecture #13: Optimal Transport

- So far, we have focused on the fundamental properties of solutions to transport problems: existence, uniqueness, and regularity of the solution to the Monge problem (i.e., Brenier's theorem).
- A major application of these results is to use the solution to

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^2 \, d\pi(x,y), \quad \mu,\nu \text{ measures, } \Pi(\mu,\nu) \text{ measures with marginals } \mu \text{ and } \nu$$

as a way to compare two probability measures.

Defn: Let X be any Polish space (e.g., $\mathbb{R}^d$), and let $p \geq 1$. Let d be a metric on the space X. We define the Wasserstein-p metric on X with respect to d as: for any μ,ν with finite p^{th} moments $\left(\int_X d(x_0,x)^p \, d\mu(x) < \infty \ \forall x_0, \text{ ditto } d\nu \right)$

$$W_p(\mu,\nu) = \left[\min_{\pi \in \Pi(\mu,\nu)} \iint_{X\,X} d^p(x,y) \, d\pi(x,y) \right]^{1/p}.$$

- The most important cases for us are $X = \mathbb{R}^d$, $d(x,y) = \|x-y\|$, so that the Wasserstein p-metric in this case is

$$W_p(\mu,\nu) = \Bigg[\underbrace{\min_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^d\,\mathbb{R}^d} \|x-y\|^p \, d\pi(x,y)}_{\text{just the solution to KP problem with } \|x-y\|^p \text{ cost}} \Bigg]^{1/p},$$

Note: Our analysis of existence and uniqueness was focused on squared Euclidean cost, which corresponds to Wasserstein-2: $W_2(\mu,\nu) = \left[\min_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^d\,\mathbb{R}^d} \|x-y\|^2 \, d\pi(x,y) \right]^{1/2}$.

. Let $\mathcal{P}_p^{p\geq 0}(X)$ be the measures on X with bounded p^{th} moment. We aim to show $W_p : \mathcal{P}_p(X) \times \mathcal{P}_p(X) \to \mathbb{R}_{\geq 0}$ is a metric. To show the requisite triangle inequality, we require a little buildup.

Lemma (Gluing Lemma): Let μ_1, μ_2, μ_3 be three probability measures, supported on Polish spaces X_1, X_2, X_3, respectively. Let $\pi_{12} \in \Pi(\mu_1, \mu_2)$ and $\pi_{23} \in \Pi(\mu_2, \mu_3)$. Then there exists π a measure on $X_1 \times X_2 \times X_3$ such that π has marginal π_{12} on $X_1 \times X_2$ and marginal π_{23} on $X_2 \times X_3$.

Proof: The key technical aspect of this proof is the notion of "disintegration of measure." This asserts that for any X, Y Polish, any measure on $X \times Y$ may be written as an "average" of measures on the fibers $\{x\} \times Y$. That is, if $\hat{\pi} \in \mathcal{P}(X \times Y)$ has marginal μ on X, then there exists a μ-a.e. unique map

$$x \mapsto \hat{\pi}_x$$

such that

$$\text{(☆)} \quad \hat{\pi} = \int_X (\delta_x \otimes \hat{\pi}_x)\, d\mu(x),$$

where as usual $\int_X \varphi\, \delta_x = \varphi(x)$ for any reasonable test function φ.

Note that apply (☆) to any test function $u \in C_b(X \times Y)$ yields

$$\int_{X \times Y} u(x,y)\, d\hat{\pi}(x,y) = \int_X \left[\int_Y u(x,y)\, d\hat{\pi}_x(y) \right] d\mu(x)$$

and for any $A \subset X \times Y$ measurable,

$$\hat{\pi}[A] = \int_A d\hat{\pi}(x,y) = \int_A (\delta_x \otimes \hat{\pi}_x)\, d\mu(x) = \int_X \hat{\pi}_x[A_x]\, d\mu(x),$$

where $A_x = \{ y \in Y \mid (x,y) \in A \}$ is the projection of A onto the second coordinate.

With this disintegration in mind, let π_{12}, π_{23} be as in the lemma statement. By D.M, we have decompositions with respect to μ_2 & call them $\pi_{12;2}$ and $\pi_{23;2}$ respectively, such that

$$\pi_{12} = \int_{X_2} \left(\pi_{12;2} \otimes \delta_{x_2} \right) d\mu_2(x_2)$$

$$\pi_{23} = \int_{X_2} \left(\underset{\in P(X_2)}{\pi_{23;2}} \otimes \delta_{x_2} \right) d\mu_2(x_2)$$

Define $\pi = \displaystyle\int_{X_2} \left(\underset{\in P(X_1)}{\pi_{12;2}} \otimes \delta_{x_2} \otimes \underset{\in P(X_3)}{\pi_{23;2}} \right) d\mu_2(x_2)$. Then for any test function $u(x_1, x_2)$,

$$\int_{X_1 \times X_2 \times X_3} u(x_1,x_2)\, d\pi(x_1,x_2,x_3) = \int_{X_1 \times X_2 \times X_3} u(x_1,x_2) \int_{X_2} \pi_{12;2} \otimes \delta_{x_2} \otimes \pi_{23;3}\, d\mu_2(x_2)$$

$$= \int_{X_1 \times X_2 \times X_3} u(x_1,x_2)\, d\pi_{12;2}(x_1)\, d\underbrace{\pi_{23;3}(x_3)}_{1}\, d\mu_2(x_2)$$

$$= \int_{X_1 \times X_2} u(x_1,x_2)\, d\pi_{12;2}(x_1)\, d\mu_2(x_2)$$

$$= \int_{X_1 \times X_2} u(x_1,x_2)\, d\pi_{12}(x_1,x_2),$$

so that $\hat{\pi}$ has $X_1 \times X_2$ marginal π_{12}. Same argument shows that it has $X_2 \times X_3$ marginal π_{23}. ∎

- We now establish that W_p is a metric.

Theorem: (i) For all $p \in [1,\infty)$, W_p is a metric on $\mathcal{P}_p(X)$.

(ii) For all $p \in [0,1)$, $W_p := \min\limits_{\pi \in \Pi(\mu,\nu)} \int_{X \times X} d^p(x,y)\, d\pi(x,y)$

is a metric on $P_p(X)$.

Remark: We are mainly interested in $p \geq 1$, and in fact $p \in (0,1)$ have a slightly different definition.

Proof: We show (i) since (ii) follows from the $p=1$ case (d^p and d are topologically equivalent in this case). We must establish:

(1.) $W_p(\mu,\nu) = W_p(\nu,\mu)$ "Obvious, just by swapping x,y in the optimal plans in $\Pi(\mu,\nu)$ and $\Pi(\nu,\mu)$, respectively."

(2.) $W_p(\mu,\nu) = 0 \iff \mu = \nu$ a.e. "$\Leftarrow$ is obvious. To see $\Rightarrow$, suppose $W_p(\mu,\nu)=0$ and let π be a plan realizing $0 = \int_{X \times X} d^p(x,y)\, d\pi(x,y)$. Then since d is a metric, $d\pi$ is supported on $\{(x,x) \mid x \in X\} \subset X \times X$. Then for all test functions $\varphi \in \mathcal{C}_b(X)$, marginal constraints

$$\int_X \varphi(x)\, d\mu(x) \overset{\downarrow}{=} \int_{X \times Y} \varphi(x)\, d\pi(x,y) = \int_Y \varphi(y)\, d\pi(x,y) \overset{\downarrow}{=} \int_Y \varphi(y)\, d\nu(y)$$

$\Rightarrow \mu = \nu$ a.e.

(3) <u>For all $\mu_1, \mu_2, \mu_3 \in \mathcal{P}_p(X)$, $W_p(\mu_1, \mu_3) \leq W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3)$</u> : Let π_{ij} be optimal for μ_i, μ_j. Let π glue π_{12}, π_{23}. Then for $X_i := \operatorname{supp}(\mu_i)$,

$$W_p(\mu_1, \mu_3) \leq \left(\iint_{X_1 \times X_3} d(x_1, x_3)^p \, d\pi_{13}(x_1, x_3) \right)^{1/p}$$

where π_{13} is the marginal of π onto $X_1 \times X_3$,

$$\overset{\text{DoM}}{=} \left(\iint_{X_1 \times X_2 \times X_3} d(x_1, x_3)^p \, d\pi(x_1, x_2, x_3) \right)^{1/p}$$

$$\overset{d \text{ metric}}{\leq} \left(\iint_{X_1 \times X_2 \times X_3} \left[d(x_1, x_2) + d(x_2, x_3) \right]^p d\pi(x_1, x_2, x_3) \right)^{1/p}$$

$$\overset{\text{Minkowski}}{\leq} \left(\iint_{X_1 \times X_2 \times X_3} d(x_1, x_2)^p \, d\pi(x_1, x_2, x_3) \right)^{1/p} + \left(\iint_{X_1 \times X_2 \times X_3} d(x_2, x_3)^p \, d\pi(x_1, x_2, x_3) \right)^{1/p}$$

$$= \left(\iint_{X_1 \times X_2} d(x_1, x_2)^p \, d\pi(x_1, x_2) \right)^{1/p} + \left(\iint_{X_2 \times X_3} d(x_2, x_3)^p \, d\pi(x_2, x_3) \right)^{1/p}$$

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

<u>Remark</u> : Using Hölder's inequality for L^p spaces ($\|fg\|_{L^1} \leq \|f\|_{L^p} \cdot \|f\|_{L^q}$ where $\frac{1}{p} + \frac{1}{q} = 1$), we get that for $p_1 \geq p_2 \geq 1$, $W_{p_1}(\mu, \nu) \geq W_{p_2}(\mu, \nu)$ $\forall \mu, \nu$.