Lecture 12, 4/22/2022

- (Continuous) OT: Monge's formulation
- (Continuous) OT: Kantorovich formulation

Monge's formulation

Assume:

- $(X, \mathcal{F})$, $(Y, \mathcal{G})$: measurable spaces

  $(X \times Y, \mathcal{F} \otimes \mathcal{G})$: the product measure space
- $\mathcal{P}(X) = \mathcal{P}(X, \mathcal{F})$, $\mathcal{P}(Y) = \mathcal{P}(Y, \mathcal{G})$: spaces of probability measures.
- $c : X \times Y \to [0, \infty]$ measurable
- $\forall \mu \in \mathcal{P}(X)$ $\forall \nu \in \mathcal{P}(Y)$:

  $$\mathcal{T}(\mu, \nu) = \{ T : X \to Y : \text{measurable}, T \# \mu = \nu \}$$

Define the Monge cost functional $E_M : \mathcal{T}(\mu, \nu) \to [0, \infty]$

$$E_M[T] = \int_X c(x, T(x))\, d\mu(x) \quad \forall T \in \mathcal{T}(\mu, \nu).$$

Monge's OT problem $\quad \min_{T \in \mathcal{T}(\mu, \nu)} E_M[T]$.

If $\hat{T} = \arg\min_{\mathcal{T}(\mu, \nu)} E_M[\cdot]$ exists, then $\hat{T}$ is called an optimal (transport) map.

Recall (Pushforward measure): $T \# \mu = \mu \circ T^{-1} : \mathcal{G} \to \mathcal{F}$.

$$(T \# \mu)(A) = \mu(T^{-1}(A)), \quad \forall A \in \mathcal{F}.$$

T is a random variable $\Rightarrow$ $\nu = T \# \mu$ is the distribution of T, the law of T, w.r.t. the prob. space $(X, \mathcal{F}, \mu)$.

Kantorovich's formulation Denote

$\mathcal{P}(X\times Y) = \{$ probability measures on $(X\times Y, \mathcal{F}\otimes\mathcal{G})\}$.

X-projection: $\pi^X : X\times Y \to X$, $\pi^X(x,y) = x$,

Y-projection: $\pi^Y : X\times Y \to Y$, $\pi^Y(x,y) = y$, $\forall (x,y)\in X\times Y$.

If $\gamma\in\mathcal{P}(X\times Y)$ then $\pi^X_\#\gamma\in\mathcal{P}(X)$ and $\pi^Y_\#\gamma\in\mathcal{P}(Y)$.
They are marginal probability measures.

Let $\mu\in\mathcal{P}(X)$, $\nu\in\mathcal{P}(Y)$. Define

$$\mathcal{A}(\mu,\nu) = \{\gamma\in\mathcal{P}(X\times Y) : \pi^X_\#\gamma = \mu,\ \pi^Y_\#\gamma = \nu\}.$$

Define Kantorovich's cost functional $E_K : \mathcal{A}(\mu,\nu)\to[0,\infty]$

$$E_K[\gamma] = \int_{X\times Y} c(x,y)\, d\gamma(x,y), \quad \gamma\in\mathcal{A}(\mu,\nu).$$

Kantorovich's OT problem: $\min\limits_{\gamma\in\mathcal{A}(\mu,\nu)} E_K[\gamma]$.

Call $\hat\gamma = \arg\min\limits_{\gamma\in\mathcal{A}(\mu,\nu)} E_K[\gamma]$, if exists, an optimal (transport) plan.

Basic questions structure/properties of $\mathcal{T}(\mu,\nu)$ and $\mathcal{A}(\mu,\nu)$; existence, uniqueness, and characterization of optimal maps/plans; duality; Wasserstein metric and its convergence properties, related gradient flows; etc.

Assumptions Unless otherwise stated, we alw assume: both X, Y: Polish spaces: complete + separable metric spaces (e.g. $\mathbb{R}^d$, L^p, $W^{k,p}$, $1 \le p < \infty$), $\mathcal{F} = \mathcal{B}_X$, $\mathcal{G} = \mathcal{B}_Y$: Borel σ-algebras.

- In the case $X = Y$, we often consider $c(x,y) = d(x,y)$ or $d(x,y)^p$ ($1 \le p < \infty$), where $d(\cdot,\cdot)$ is the metric of d.
- Sometimes, we shall assume that X and Y are locally compact.
- Even more often, we assume that $X = Y = \mathbb{R}^d$, $\mathcal{F} = \mathcal{G} =$ Lebesgue σ-algebra.

Examples
- The discrete OT (M or K version) is a special case of (general, continuous) OT.

$X = \{x_1, \cdots, x_m\}$ and $Y = \{y_1, \cdots, y_n\}$.

$\mathcal{F} = \{$all subsets of $X\}$, $\mathcal{G} = \{$all subsets of $Y\}$

$\mu \in \mathcal{P}(X) \leftrightarrow a \in \mathcal{P}_m$, $\nu \in \mathcal{P}(Y) \leftrightarrow b \in \mathcal{P}_n$. $a_i = \mu(\{x_i\})$, $b_j = \mu(\{y_j\})$

$c : X \times Y \to [0, \infty) \iff C = [C_{ij}] \in \mathbb{R}^{m \times n}$, $C \ge 0$, $C_{ij} = c(x_i, y_j)$.

$\mathcal{J}(\mu, \nu) \leftrightarrow \mathcal{J}(a, b)$, $T : X \to Y$.

$$T \in \mathcal{J}(a,b) \iff b_j = \sum_{i : T(x_i) = y_j} a_i \quad \forall j \iff \nu(\{y_j\}) = \mu(T^{-1}(\{y_j\})) \quad \forall j$$

$$\iff \nu = T\#\mu \iff T \in \mathcal{J}(\mu, \nu)$$

$$E_M[T] = \int_X c(x, T(x))\, d\mu(x) = \int_{\bigcup_{i=1}^m \{x_i\}} c(x, T(x))\, d\mu(x)$$

$$= \sum_{i=1}^m \int_{\{x_i\}} c(x_i, T(x_i))\, d\mu = \sum_{i=1}^m c(x_i, T(x_i))\, \mu(\{x_i\})$$

$$= \sum_{i=1}^m a_i C(x_i, T(x_i)), \quad \text{where } a_i = \mu(\{x_i\}) \ \forall i.$$

$$= E_M^{disc}[T].$$

Now, $P \in \mathcal{A}(a,b) \longleftrightarrow \gamma \in \mathcal{A}(\mu,\nu)$. $\gamma(\{(x_i, y_j)\}) = P_{ij}$.

$$E_K[\gamma] = \int_{X\times Y} c(x,y)\,d\gamma(x,y) = \sum_{i,j} \int_{\{(x_i,y_j)\}} c(x,y)\,d\gamma(x,y)$$

$$= \sum_{i,j} c(x_i, y_j)\,\gamma(\{(x_i,y_j)\}) = \sum_{i,j} C_{ij} P_{ij} = \langle C, P\rangle = E_K^{disc}[P].$$

$\odot$ $(X, \mathcal{F}), (Y, \mathcal{G})$: measurable spaces, $\mu \in \mathcal{P}(X)$, $y_0 \in Y$.
$\nu = \delta_{y_0}$ $T \in \mathcal{J}(\mu,\nu) \Leftrightarrow \mu(T^{-1}(B)) = \nu(B) = \begin{cases} 1 & \text{if } y_0 \in B, \\ 0 & \text{if } y_0 \notin B. \end{cases}$
clearly, $y_0 \in \text{Range}(T)$ i.e., $\exists x_0 \in X$ s.t. $T(x_0) = y_0$.
Otherwise, $y_0 \notin \text{Range}(T) =: B \subseteq Y$. $T\#\mu(B) = \mu(T^{-1}(B))$
$= \mu(X) = 1$, and $\nu(B) = \delta_{y_0}(B) = 0$, contradicting $T\#\mu = \nu$.
Now, assume $y_0 \in \text{Range}(T)$. Denote $A_{y_0} = \{x \in X : T(x) \neq y_0\}$. Then, $T\#\mu = \delta_{y_0} \Leftrightarrow \mu(T^{-1}(A_{y_0})) = 0$. Thus

$$E_M[T] = \int_X c(x, T(x))\,d\mu(x) = \int_{A_{y_0}} + \int_{X\setminus A_{y_0}} = \int_{X \setminus A_{y_0}}$$

$$= \int_{\{x\in X : T(x) = y_0\}} c(x, T(x))\,d\mu(x) = \int_{T^{-1}(\{y_0\})} c(x, y_0)\,d\mu(x) = \int_X c(x,y_0)\,d\mu(x).$$

Let $\mu \in \mathcal{P}(X)$ and $\nu = \delta_{y_0}$ with $y_0 \in Y$. If $\gamma \in \mathcal{A}(\mu,\nu)$, then $\gamma = \mu \times \delta_{y_0}$, i.e., $\gamma(A\times B) = \mu(A)\delta_{y_0}(B)$ $\forall A \in \mathcal{B}(X), B \in \mathcal{B}(Y)$.
In fact, if $y_0 \notin B$, then $0 \le \gamma(A\times B) \le \gamma(X \times B) = \delta_{y_0}(B) = 0$, hence $\gamma(A\times B) = \mu(A)\nu(B) = 0$. If $y_0 \in B$ then $y_0 \notin Y\setminus B$.
Hence $\gamma(A\times Y) - \gamma(A\times B) = \gamma(A\times(Y\setminus B)) = 0$, i.e., $\gamma(A\times B)$
$= \gamma(A\times Y) = \mu(A) = \mu(A)\nu(B)$ as $\nu(B) = \delta_{y_0}(B) = 1$. Now,

$$E_K[\gamma] = \int_{X\times Y} c(x,y)\,d\gamma(x,y) = \int_{X\times Y} c(x,y)\,d(\mu\times\delta_{y_0})(x,y)$$

$$= \int_X \left(\int_Y c(x,y)\,d\delta_{y_0}(y)\right) d\mu(x) = \int_X c(x,y_0)\,d\mu(x) = E_M[T].$$

So, if $\nu = \delta_{y_0}$ for some $y_0 \in Y$, and $\mu \in \mathcal{P}(X)$, then $E_K[\gamma] = E_M[T]$ for any $\gamma \in \mathcal{A}(\mu,\nu)$ and any $T \in \mathcal{T}(\mu,\nu)$.

<u>Monge vs. Kantorovich</u>

- $\mathcal{T}(\mu,\nu) = \phi$: possible, e.g., discrete OT.

In general, let $\mu = \delta_{x_0}$ for some $x_0 \in X$. Suppose $T: X \to Y$ is measurable. Then $T\#\mu = \delta_{T(x_0)}$. In fact, let $B \in \mathcal{G}$. $T(x_0) \in B \iff x_0 \in T^{-1}(B)$. Hence, $(T\#\mu)(B)$
$= \mu(T^{-1}(B)) = \begin{cases} 1 & \text{if } x_0 \in T^{-1}(B) \\ 0 & \text{if } x_0 \notin T^{-1}(B) \end{cases} = \begin{cases} 1 & T(x_0) \in B \\ 0 & T(x_0) \notin B \end{cases} = \delta_{T(x_0)}(B)$.

Thus, if $\mu = \delta_{x_0}$, unless $\nu = \delta_{y_0}$ for some $y \in X$, the set $\mathcal{T}(\mu,\nu) = \phi$.

$\mathcal{A}(\mu,\nu) \neq \phi$ as $\mu \times \nu \in \mathcal{A}(\mu,\nu)$.

Both $\mathcal{T}(\mu,\nu)$ and $\mathcal{A}(\mu,\nu)$ are convex.

- $\gamma \longmapsto E_K[\gamma]$ is linear.

  $T \longmapsto E_M[T]$ is (in general) nonlinear.

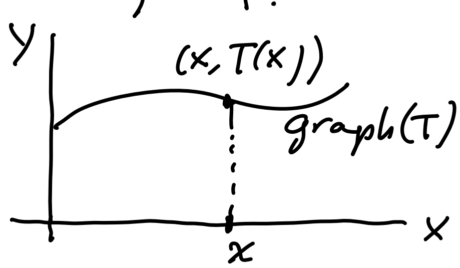
- K-OT is a relaxation of M-OT.

Let $T \in \mathcal{T}(\mu,\nu)$. Let $\gamma_T = (Id \times T)\#\mu$. Here $Id \times T$: $X \to X \times Y$ is defined by $(Id \times T)(x) = (x, T(x))$ $\forall x$.

$Id \times T$ is measurable. Moreover,

$$\gamma_T(X \times Y) = \mu((Id \times T)^{-1}(X \times Y)) = \mu(X) = 1.$$

Hence $\gamma_T \in \mathcal{P}(X \times Y)$. Now,



$\pi^X_\# \gamma_T = \pi^X_\#((Id \times T)_\# \mu)$
$= ((Id \times T)\#\mu) \circ (\pi^X)^{-1} = \mu \circ (Id \times T)^{-1} \circ (\pi^X)^{-1}$

$= \mu \circ (\pi^X \circ Id \times T)^{-1} = \mu \circ (Id)^{-1} = \mu.$

$\pi^Y_{\#} \gamma_T = \pi^Y_{\#}((Id \times T)_{\#}\mu) = ((Id \times T)_{\#}\mu) \circ (\pi^Y)^{-1}$

$= \mu \circ (Id \times T)^{-1} \circ (\pi^Y)^{-1} = \mu \circ (\pi^Y \circ (Id \times T))^{-1} = \mu \circ T^{-1} = T_{\#}\mu = \nu.$

Hence, $\gamma_T \in \mathcal{A}(\mu, \nu)$. Now, by the change of variable formula (see next lecture),

$$E_K[\gamma_T] = \int_{X \times Y} c(x,y)\, d\gamma_T(x,y) = \int_X c \circ (Id \times T)\, d\mu$$

$$= \int_X c(x, T(x))\, d\mu(x) = E_M[T]$$

So, in general, $\inf_{\mathcal{T}(\mu,\nu)} E_M[\cdot] \geq \inf_{\mathcal{A}(\mu,\nu)} E_K[\cdot]$.

⊙ Note γ_T is the measure on $(X \times Y, \mathcal{F} \otimes \mathcal{G})$ defined by

$$\int_{X \times Y} \xi(x,y)\, d\gamma_T(x,y) = \int_X \xi(x, T(x))\, d\mu(x),$$

for any $\xi : X \times Y \to [0, \infty]$ measurable. [Need proper structural assumptions on $X, Y, \mathcal{F}, \mathcal{G}$.]

⊙ $X = Y$: Polish space, $x_0, y_0 \in X$, $\mu = \delta_{x_0}$, $\nu = \delta_{y_0}$, $c(x,y) = d(x,y)$.

$T \in \mathcal{T}(\mu, \nu) \Leftrightarrow \mu \circ T^{-1} = \nu \Leftrightarrow \forall B \subseteq X = Y: (\mu \circ T^{-1})(B) = \nu(B)$

$\Leftrightarrow \delta_{x_0}(T^{-1}(B)) = \delta_{y_0}(B) = \begin{cases} 1 & \text{if } y_0 \in B \\ 0 & \text{if } y_0 \notin B \end{cases} \Leftrightarrow T(x_0) = y_0$

So, $\mathcal{T}(\mu,\nu) = \{T : X \to Y : \text{measurable}, T(x_0) = y_0\}$.

Note that $T_0(x) = y_0$ $\forall x \in X$ defines $T_0 \in \mathcal{T}(\mu,\nu)$. So, $\mathcal{T}(\mu,\nu) \neq \emptyset$. Moreover,

$$E_M[T] = \int_X d(x, T(x))\, d\mu(x) = d(x_0, T(x_0)) = d(x_0, y_0).$$

In particular, $\min_{T \in \mathcal{T}(\mu,\nu)} E_M[T] = d(x_0, y_0)$.

Now, $\gamma \in \mathcal{A}(\mu,\nu) = \mathcal{A}(\delta_{x_0}, \delta_{y_0}) \iff \gamma = \delta_{(x_0,y_0)}$.

$$E_K[\gamma] = \int_{X\times Y} d(x,y)\, d\delta_{x_0}(x)\, d\delta_{y_0}(y) = d(x_0,y_0).$$

So, $\min_{\gamma \in \mathcal{A}(\mu,\nu)} E_K[\gamma] = d(x_0,y_0)$. $K = M$ in this case.

⊙ A related result (A Polish space is a complete and separable metric space.)

<u>Theorem</u> Let X, Y be Polish spaces, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Suppose $\gamma \in \mathcal{A}(\mu,\nu)$, $\Gamma \subseteq X \times Y$ is a γ-measurable graph and γ is concentrated on Γ. Then there exists a Borel map $T \in \mathcal{T}(\mu,\nu)$ such that $\gamma = (Id \times T)_{\#}\mu$. <u>QED</u>

⊙ The following result shows the relation between the M. and K. OT problems in the continuum setting, quite different from the discrete setting. A key assumption is that $\mu \in \mathcal{P}(X)$ is μ contains no atom: $\mu(\{x\}) = 0 \ \forall x \in X$.

<u>Theorem</u> (Pratelli) Let X, Y be Polish, and $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Assume $c : X \times Y \to [0,\infty]$ is continuous. Assume also that μ contains no atoms. Then

$$\inf_{\mathcal{T}(\mu,\nu)} E_M[\cdot] = \inf_{\mathcal{A}(\mu,\nu)} E_K[\cdot].$$ <u>QED</u>