Lecture 13, Monday, 4/25/2022

Given: X, Y: Polish spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$.
$c: X \times Y \to [0,\infty]$ measurable.

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

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

Monge's OT: $\inf\limits_{T\in\mathcal{T}(\mu,\nu)} E_M[T]$,

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

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

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

$T \in \mathcal{T}(\mu,\nu)$: transport map.

$\gamma \in \mathcal{A}(\mu,\nu)$: transport plan.

- If $\mathcal{T}(\mu,\nu) \neq \emptyset$ and $T \in \mathcal{T}(\mu,\nu)$ then $\gamma_T = (Id \times T)_{\#}\mu \in \mathcal{A}(\mu,\nu)$, where $Id\times T(x) = (x, T(x))$, defines $Id\times T: X \to X\times Y$, and $E_M[T] = E_K[\gamma_T]$.

<u>Def</u>. A measure μ on X is non atomic or contains no atoms if $\mu(\{x\}) = 0$ $\forall x \in X$.

<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 is non-atomic. Then

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

Today $\odot$ More about transport maps and transport plans.

$\odot$ Existence of minimizers for the K-OT problem.

$\odot$ Direct methods in the calculus of variations

<u>Proposition</u> Let X, Y be Polish, $\mu \in \mathcal{P}(X)$, and $T: X \to Y$ Borel measurable.

(1) $T\#\mu \in \mathcal{P}(Y)$.

(2) Let $\nu \in \mathcal{P}(Y)$. Then $T\#\mu = \nu$ if and only if for any bounded and measurable $\varphi: Y \to \mathbb{R}$,

$$\int_Y \varphi(y)\, d\nu(y) = \int_X \varphi(T(x))\, d\mu(x). \quad (*)$$

<u>Corollary</u> (change of variables) Let X, Y be Polish, $\mu \in \mathcal{P}(X)$, and $T: X \to Y$ Borel measurable. Then, for any bounded and measurable $\varphi: Y \to \mathbb{R}$,

$$\int_Y \phi \, d(T\#\mu) = \int_X \phi \circ T \, d\mu. \qquad \underline{QED}$$

<u>Proof of Proposition</u>

(1) By definition, $T\#\mu$ is a measure, and a probability measure on Y.

(2) $\forall B \in \mathcal{B}(Y)$. Set $\varphi = \chi_B$ ($\chi_B = 1$ on B and 0 on $B^c = Y \setminus B$). Note that $\chi_B \circ T = \chi_{T^{-1}(B)}$.

Now $(*) \Rightarrow \nu(B) = \int_X (\chi_B \circ T)(x)\, d\mu(x) = \mu(T^{-1}(B))$
$= (T_\# \mu)(B)$. Hence, $T_\# \mu = \nu$.

Conversely, assume $T_\# \mu = \nu$. If $\varphi = \chi_B$ for $B \in \mathcal{B}(Y)$ then

$$\int_Y \varphi\, d\nu = \nu(B) = \mu(T^{-1}(B)) = \int_X \chi_{T^{-1}(B)}\, d\mu = \int_X (\chi_B \circ T)\, d\mu.$$

Hence $(*)$ is true for simple functions φ.
Now, for any bounded and Borel (i.e., Borel measurable) $\varphi : Y \to \mathbb{R}$, there simple functions $\varphi_k : Y \to \mathbb{R}$ s.t. $\|\varphi_k - \varphi\|_\infty \to 0$. Hence,

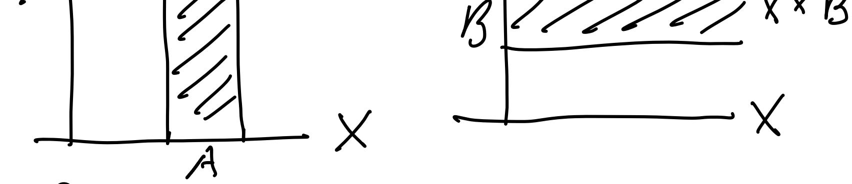
$$\int_Y \varphi\, d\nu = \lim_{k\to\infty} \int_Y \varphi_k\, d\nu = \lim_{k\to\infty} \int_X \varphi_k \circ T\, d\mu = \int_X \varphi \circ T\, d\mu.$$ QED

Now, study $\mathcal{A}(\mu, \nu) \subseteq \mathcal{P}(X \times Y)$. Given $\gamma \in \mathcal{P}(X\times Y)$.
What are the conditions that $\gamma \in \mathcal{A}(\mu, \nu)$?

Proposition Let X and Y be Polish. Let $\gamma \in \mathcal{P}(X \times Y)$.
Then the following are equivalent:

(1) $\gamma \in \mathcal{A}(\mu, \nu)$;

(2) $\gamma(A \times Y) = \mu(A)\ \forall A \in \mathcal{B}(X)$, $\gamma(X \times B) = \nu(B)\ \forall B \in \mathcal{B}(Y)$;



(3) $\int_{X\times Y} \varphi\, d\gamma = \int_X \varphi\, d\mu\ \ \forall \varphi : X \to \mathbb{R}$: Borel measurable,

$$\int_{X\times Y}\psi\,d\gamma=\int_Y \psi\,d\nu,\ \forall \psi: X\to\mathbb{R}: \text{Borel measurable};$$

(4) $\int_{X\times Y}\varphi\,d\gamma=\int_X \varphi\,d\mu\ \ \forall\varphi: X\to\mathbb{R}$: bounded and Borel measurable,

$\int_{X\times Y}\psi\,d\gamma=\int_Y \psi\,d\nu\ \ \forall\psi: Y\to\mathbb{R}$: bounded and Borel measurable;

(5) $\int_{X\times Y}\varphi\,d\gamma=\int_X \varphi\,d\mu\ \ \forall\varphi\in C_b(X)$, $\int_{X\times Y}\psi\,d\gamma=\int_Y \psi\,d\nu\ \ \forall\psi\in C_b(Y)$.

Note:

$C_b(Z)\triangleq\{\text{continuous and bounded functions } f: Z\to\mathbb{R}\}$.

$C_b(Z)$ is a normed vector space with $\|\varphi\|=\sup_{x\in Z}|\varphi(x)|$.

<u>Proof</u> We show (1) $\Leftrightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5) $\Rightarrow$ (2).

(1) $\Rightarrow$ (2). $\forall A\in\mathcal{B}(X)$.

$$\mu(A)=(\pi^X_\#\gamma)(A)=\gamma\left((\pi^X)^{-1}(A)\right)=\gamma(A\times Y).$$

Similarly, $\nu(B)=\gamma(X\times B)$ for any $B\in\mathcal{B}(Y)$.

(2) $\Rightarrow$ (1) $\forall A\in\mathcal{B}(X)$.

$$(\pi^X_\#\gamma)(A)=\gamma((\pi^X)^{-1}(A))=\gamma(A\times Y)=\mu(A)$$

Hence, $\pi^X_\#\gamma=\mu$. Similarly, $\pi^Y_\#\gamma=\nu$.

(2) $\Rightarrow$ (3) If $\varphi=\mathbb{1}_A$ for some $A\in\mathcal{B}(X)$ then by (2),

$$\int_X\varphi\,d\mu=\mu(A)=\gamma(A\times Y)=\int_{A\times Y}d\gamma=\int_{X\times Y}\varphi\,d\gamma.$$

We have then

$$\int_X\varphi\,d\mu=\int_{X\times Y}\varphi\,d\gamma \qquad (*)$$

if φ is a simple Borel function on X. But a nonnegative Borel function can be approximated by a sequence of nonnegative increasing simple functions. So, the monotone convergence theorem, and the decomposition $\varphi = \varphi^+ - \varphi^-$, imply that (*) is true for any Borel function φ.

$(3) \Rightarrow (4)$ This is obvious.

$(4) \Rightarrow (5)$ This is obvious.

$(5) \Rightarrow (2)$ Let $A \in \mathcal{B}(X)$. Since X, Y and $X \times Y$ are Polish, probability measures on these spaces are regular (see next lecture). Thus, $\forall \varepsilon > 0$, $\exists$ open $U \subseteq X$ and compact $K \subseteq X$ such that $K \subseteq A \subseteq U$ and $\mu(U \setminus K) < \varepsilon$. Similarly, $\exists$ open $\hat{U} \subseteq X \times Y$ and compact $\hat{K} \subseteq X \times Y$ such that $\hat{K} \subseteq A \times Y \subseteq \hat{U}$ and $\gamma(\hat{U} \setminus \hat{K}) < \varepsilon$. Let $K_0 = \pi^X(\hat{K}) \cup K \subseteq X$ and $U_0 = \pi^X(\hat{U}) \cap U \subseteq X$. $\pi^X(\hat{K})$ is compact in X. So, K_0 is compact in X. Also, $\pi^X(\hat{U}) \subseteq X$ is open. So, U_0 is open in X. Moreover, $K_0 \subseteq A \subseteq U_0$ and $\mu(U_0 \setminus K_0) \le \mu(U \setminus K) < \varepsilon$.

Define $\varphi : X \to \mathbb{R}$ by $\varphi(x) = \dfrac{d_X(x, U_0^c)}{d_X(x, K_0) + d_X(x, U_0^c)}$ for any $x \in X$, where d_X is the metric of X and $U_0^c = X \setminus U_0$. Clearly, φ is continuous: the denominator $\ne 0$, since $x \in X$ and $d_X(x, K_0) + d_X(x, U_0^c) = 0$ $\Rightarrow d_X(x, K_0) = 0 \Rightarrow x \in K_0 \Rightarrow d_X(x, U_0^c) > 0$, contradiction. Clearly $0 \le \varphi \le 1$. So, $\varphi \in C_b(X)$. Note that $\varphi = 0$ on U_0^c and $\varphi = 1$ on K_0. We have now

$$\int_X \varphi \, d\mu \ge \int_{K_0} d\mu = \mu(K_0) \ge \mu(A) - \varepsilon.$$

$$\int_X \varphi \, d\mu \le \int_{U_0} d\mu = \mu(U_0) \le \mu(A) + \varepsilon.$$

Since $K_0 \times Y = (\pi^X(\hat{K}) \cup K) \times Y \supseteq \pi^X(\hat{K}) \times Y \supseteq \hat{K}$,

$$\int_{X\times Y}\varphi\,d\gamma \geq \int_{K_0\times Y}\varphi\,d\gamma = \int_{K_0\times Y} d\gamma = \gamma(K_0\times Y) \geq \gamma(\hat{K}) \geq \gamma(A\times Y)-\varepsilon.$$

Note that for $(x,y)\in X\times Y$, $(x,y)\notin \hat{U} \Rightarrow x\notin \pi^X(\hat{U}) \Rightarrow x\notin U_0$. So, $\varphi(x)=0$ if $(x,y)\notin\hat{U}$. Hence

$$\int_{X\times Y}\varphi\,d\gamma = \int_{\hat{U}}\varphi\,d\gamma \leq \int_{\hat{U}} d\gamma = \gamma(\hat{U}) \leq \gamma(A\times Y)+\varepsilon.$$

Therefore, since $\int_X \varphi\,d\mu = \int_{X\times Y}\varphi\,d\gamma$ by (4),

$$|\gamma(A\times Y)-\mu(A)| \leq \left|\gamma(A\times Y)-\int_{X\times Y}\varphi\,d\gamma\right| + \left|\int_X\varphi\,d\mu-\mu(A)\right| \leq 2\varepsilon.$$

Thus $\gamma(A\times Y)=\mu(A)$. Similarly, $\gamma(X\times B)=\nu(B)$ $\forall B\in\mathcal{B}(Y)$. QED

Theorem Let X and Y be Polish spaces, $\mu\in\mathcal{P}(X)$ and $\nu\in\mathcal{P}(Y)$, and $c: X\times Y\to[0,\infty]$ be lower-semicontinuous. Then there exists $\hat{\gamma}\in\mathcal{A}(\mu,\nu)$ such that $E_K[\hat{\gamma}] = \min_{\gamma\in\mathcal{A}(\mu,\nu)} E_K[\gamma]$.

Def. If (Z,d) is a metric space, then that $f: Z\to\mathbb{R}\cup\{+\infty\}$ is weak-lower semi-continuous means that

$$\liminf_{k\to\infty} f(x_k) \geq f(x) \quad \text{if } x_k\to x.$$

Note continuity $\Rightarrow$ lower semi-continuity.

To prove the theorem, we need to prove the following lemma which is itself important:

Lemma Let X be a metric space, $G\subseteq X$ an open subset, and $f: X\to[0,\infty]$ a lower semi-continuous function. Suppose $\mu_n\to\mu$ narrowly

in $\mathcal{P}(X)$, then
$$\liminf_{n\to\infty} \int_G f\,d\mu_n \geq \int_G f\,d\mu.$$
In particular $\liminf_{n\to\infty} \int_X f\,d\mu_n \geq \int_X f\,d\mu.$

Proof Let $g(x) = \mathbb{1}_G(x) f(x)$ $(x \in X)$. Let $x_k \to x$ in X. Then $\liminf_{k\to\infty} f(x_k) \geq f(x)$. If $x \in G$, then for k large, $x_k \in G$. Hence $\liminf_{k\to\infty} g(x_k) = \liminf_{k\to\infty} f(x_k) \geq f(x) = g(x)$. If $x \notin G$ then $g(x) = 0$ but $g(x_k) \geq 0$. So $\liminf_{k\to\infty} g(x_k) \geq g(x)$. Hence, g is lower semi continuous. So, we can assume $G = X$.

Define for each $k \in \mathbb{N}$
$$f_k(x) = \inf_{y\in X} \{f(y) \wedge k + k\,d(x,y)\}, \quad x \in X.$$
Then $0 \leq f_k \leq f_{k+1} \leq \cdots \leq f \wedge k$. (The last inequality results from setting $y = x$ in $\inf_{y\in X}$.) Each f_k is Lipschitz-continuous. In fact, if $x, x' \in X$ then
$$f_k(x') \leq f(y) \wedge k + k\,d(x',y) \leq f(y) \wedge k + k\,d(x,y) + k\,d(x',x) \quad \forall y \in X.$$
Hence, taking $\inf_{y\in X}$, we get $f_k(x') \leq f_k(x) + k\,d(x',x)$, and switching x and x', we have $|f_k(x') - f_k(x)| \leq k\,d(x',x)$.

Claim: $f_k \uparrow f$, i.e. $f = \sup_k f_k$. Fix $x \in X$. Assume without loss of generality that $\sup_k f_k(x) < \infty$. $\forall k$. $\exists x_k \in X$ s.t.
$$f(x_k) \wedge k + k\,d(x, x_k) \leq f_k(x) + \frac{1}{k}.$$
Since $f \geq 0$, $d(x, x_k) \leq \frac{1}{k} f_k(x) + \frac{1}{k^2} \to 0$, and $f(x_k) \wedge k \leq f_k(x) + \frac{1}{k}$.
So, by the lower semi continuity of f,
$$\sup_{k\geq 1} f_k(x) \geq \liminf_{k\to\infty} \left(f(x_k) \wedge k - \frac{1}{k}\right) = \liminf_{k\to\infty} f(x_k) \wedge k \geq f(x).$$
Now, for each k, since $\mu_n \to \mu$ narrowly,

$$\liminf_{n\to\infty} \int_X f\,d\mu_n \geq \liminf_{n\to\infty} \int_X f_k\,d\mu_n = \int_X f_k\,d\mu.$$

Hence, by ~~the~~ monotone convergence theorem,

$$\liminf_{n\to\infty} \int_X f\,d\mu_n \geq \lim_{k\to\infty} \int_X f_k\,d\mu = \int_X f\,d\mu. \quad \underline{\text{QED}}$$

How to prove the existence of a minimizer of $f\in C(\mathbb{R}^d)$ with $f(\infty)=\infty$?

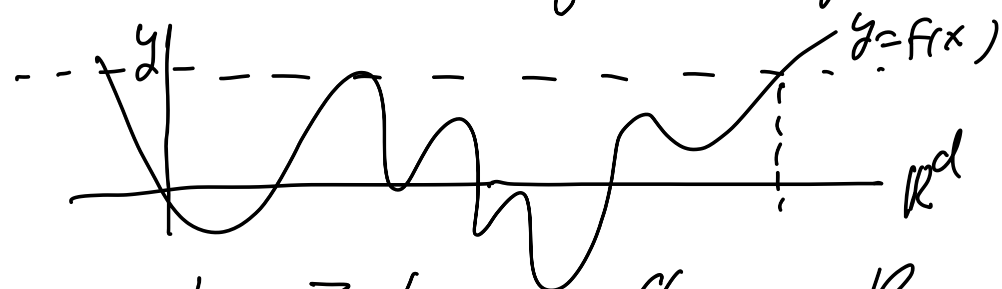
<u>Direct Methods in the calculus of variations</u>

<u>Step 1</u> f is bounded below. So $\alpha := \inf_{\mathbb{R}^d} f > -\infty$

$\exists\, x_k \in \mathbb{R}^d$ s.t. $f(x_k) \to \alpha$.

<u>Step 2</u> $f(\infty)=\infty$ So, $\exists\, M>0$ s.t. $\|x_k\| \leq M$ $\forall k\geq 1$.

<u>Step 3</u> Compactness of $\overline{B}(0,M) \Longrightarrow \exists$ subseq. $x_{k_j} \to x_\infty \in \mathbb{R}^d$. Continuity $\Longrightarrow f(x_{k_j}) \to f(x_\infty)=\alpha$.



<u>Theorem</u> Let Z be a reflexive Banach space and K a nonempty, convex, and (strongly) closed subset of Z. Assume $f: K \to \mathbb{R}\cup\{+\infty\}$ satisfies the following:

- $\inf_K f > -\infty$;
- $\exists\, c_1>0,\ c_2\in\mathbb{R}$, s.t. $f(x) \geq c_1\|x\| + c_2$ $\forall x\in K$;
- $f: K\to\mathbb{R}$ is sequentially weakly lower semicontinuous.

Then $\exists\, \hat{z} \in K$ s.t. $f(\hat{z}) = \min_{z\in K} f(z)$.

<u>Proof</u> Let $\alpha := \inf_K f > -\infty$. So, $\exists\, x_k \in K$ s.t. $f(x_k) \to \alpha$. Since $f(x_k) \geq C_1 \|x_k\| + C_2$ $(\forall k)$, $\{x_k\}$ is bounded. Thus, it has a subsequence, not relabeled, such that $x_k \to \hat{z}$ weakly for some $\hat{z} \in K$. Since K is convex and (strongly) closed, it is weakly closed. Hence, $\hat{z} \in K$. Since f is weakly lower semicontinuous, $\liminf_{k\to\infty} f(x_k) \geq f(\hat{z})$. Hence $\alpha \geq f(\hat{z}) \geq \alpha$, and $-(\hat{z}) = \alpha$. <u>QED</u>

<u>Proof of the existence theorem</u> for the K-OT problem.

<u>Proof</u> Note $\mathcal{A}(\mu,\nu) \neq \phi$. Since $c \geq 0$, $\alpha := \inf_{\gamma \in \mathcal{A}(\mu,\nu)} E_K[\gamma] \geq 0$. If $\alpha = +\infty$ then any $\gamma \in \mathcal{A}(\mu,\nu) \neq \phi$ is a minimizer. Assume $0 \leq \alpha < \infty$. Then, since $\mathcal{A}(\mu,\nu)$ is narrowly compact (cf. next lectures), $\exists\, \gamma_k \in \mathcal{A}(\mu,\nu)$ s.t. $E_K[\gamma_k] \to \alpha$.

Since $E_K : \mathcal{A}(\mu,\nu) \to [0,\infty]$ is lower semicontinuous w.r.t. narrow convergence (cf. next lectures),

$$\alpha = \liminf_{j\to\infty} E_K[\gamma_{k_j}] \geq E_K[\hat{\gamma}] \geq \alpha. \quad \underline{\text{QED}}$$

<u>Questions</u>
- Narrow convergence / topology?
- $\mathcal{A}(\mu,\nu)$ is narrowly compact?
- E_K is lower semi-cont. w.r.t. the narrow topology?