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 \(T(\mu, \nu) = \{ T: X \to Y: \text{measurable}: T#\mu = \nu \}\)
\(A(\mu, \nu) = \{ \gamma \in \mathcal{P}(X \times Y): \pi_X^* \gamma = \mu, \pi_Y^* \gamma = \nu \}\)

Monge's OT: \(\inf_{T \in T(\mu, \nu)} E_{\mu}[T], \quad E_{\mu}[T] = \int_X c(x, T(x)) \, d\mu(x)\)

Kantorovich's OT: \(\inf_{\gamma \in A(\mu, \nu)} E_{\gamma}[\gamma], \quad E_{\gamma}[\gamma] = \int_{X \times Y} c(x, y) \, d\gamma(x, y)\)

\(T \in T(\mu, \nu): \) transport map,
\(\gamma \in A(\mu, \nu): \) transport plan,

\(\bullet\) If \(T(\mu, \nu) \neq \emptyset\) and \(T \in T(\mu, \nu)\) then
\(Y_T = (\text{Id}_X \times T)#\mu \in A(\mu, \nu),\) where \(\text{Id}_X \times T(x) = (x, T(x))\),
defines \(\text{Id}_X \times T: X \to X \times Y,\) and \(E_{\mu}[T] = E_{\gamma}[\gamma_T].\)

Def. A measure \(\mu\) on \(X\) is non-atomic or contains no atoms if \(\mu(\{x\}) = 0\) \(\forall x \in X.\)

Theorem (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 \(\mu\) is non-atomic. Then
\(\inf_{T(\mu, \nu)} E_{\mu}[\cdot] = \inf_{A(\mu, \nu)} E_{\gamma}[\cdot]. \quad \square \)
Today 1. More about transport maps and transport plans.
   (2) Existence of minimizers for the K-OT problem.
   (3) Direct methods in the calculus of variations.

Proposition Let \( X, Y \) be Polish, \( \mu \in P(X) \), and \( T : X \to Y \) Borel measurable.
\begin{enumerate}
\item \( T^\# \mu \in P(Y) \).
\item Let \( \nu \in P(Y) \). Then \( T^\# \mu = \nu \) if and only if for any bounded and measurable \( \varphi : Y \to \mathbb{R} \),
   \[ \int_Y \varphi(y) \, d\nu(y) = \int_X \varphi(T(x)) \, d\mu(x). \tag{X} \]
\end{enumerate}

Corollary (Change of variables) Let \( X, Y \) be Polish, \( \mu \in P(X) \), and \( T : X \to Y \) Borel measurable. Then for any bounded and measurable \( \varphi : Y \to \mathbb{R} \),
\[ \int_Y \varphi \, d(T^\# \mu) = \int_X \varphi \circ T \, d\mu. \tag{QED} \]

Proof of Proposition
\begin{enumerate}
\item By definition, \( T^\# \mu \) is a measure, and a probability measure on \( Y \).
\item \( \forall B \in \mathcal{B}(Y) \). Set \( \varphi = X_B \) (\( X_B \) = 1 on \( B \) and 0 on \( B^c = X \setminus B \)). Note that \( X_B \circ T = X_{T^{-1}(B)} \).
\end{enumerate}
Now \( (\mathcal{X} \times Y) \Rightarrow \mathcal{V}(B) = \int (X_\beta \circ T)(x) \, d\mu(x) = \mu(T^{-1}(B)) = T_\# \mu(B). \) Hence, \( T_\# \mu = \nu. \)

Conversely, assume \( T_\# \mu = \nu. \) If \( \varphi = X_\beta \) for \( \beta \in \mathcal{B}(Y) \) then
\[
\int_Y \varphi \, d\nu = \nu(B) = \mu(T^{-1}(B)) = \int_X X_\beta^{-1} \, d\mu = \int_X (X_\beta \circ T) \, d\mu.
\]
Hence \((\ast)\) is true for simple functions \( \varphi. \)

Now, for any bounded and Borel (i.e., Borel measurable) \( \varphi : X \to \mathbb{R}, \) there simple functions \( \varphi_k : X \to \mathbb{R} \) s.t. \( \| \varphi_k - \varphi \|_\infty \to 0. \) Hence,
\[
\int_Y \varphi \, d\nu = \lim_{k \to \infty} \int_Y \varphi_k \, d\nu = \lim_{k \to \infty} \int_X \varphi_k \circ T \, d\mu = \int_X \varphi \circ T \, d\mu. \quad \text{QED}
\]

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. \( \int_{X \times Y} \varphi d\gamma = \int_X \varphi d\mu \forall \varphi : X \to \mathbb{R} : \text{Borel measurable}, \)
\[
\int \int_{x \times y} \varphi \, dy \cdot dx = \int \int_{x \times y} \varphi \, dx \cdot dy.
\]

\text{Proof: We show (1) } \iff \text{ (2) } \implies \text{ (3) } \implies \text{ (4) } \implies \text{ (5) } \implies \text{ (2).}

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

\[
\mu(A) = (\mathcal{T}^X \times \gamma)(A) = \gamma(\mathcal{T}^X(A)) = \gamma(A \times Y).
\]

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

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

\[
(\mathcal{T}^X \times \gamma)(A) = \gamma(\mathcal{T}^X(A)) = \gamma(A \times Y) = \mu(A).
\]

Hence, \( \mathcal{T}^X \times \gamma = \mu \). Similarly, \( \mathcal{T}^X \gamma = \nu \).

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

\[
\int \int_{x \times y} \varphi \, dx \cdot dy = \int_{A \times Y} dy = \int_{A \times Y} \varphi \, dy.
\]

We have then

\[
\int \int_{x \times y} \varphi \, dx \cdot dy = \int \int_{x \times y} \varphi \, dy \cdot dx \quad (\ast)
\]

If \( \varphi \) 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 \((\ast)\) is true for any Borel function \( \varphi \).

\((3) \Rightarrow (4)\) 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 \( U \in Z \) and compact \( K \subseteq Z \) such that \( K \subseteq A \subseteq U \) and \( \mu(U \setminus K) < \varepsilon \).

Similarly, \( \exists \) open \( \widetilde{U} \subseteq X \times Y \) and compact \( \widetilde{K} \subseteq X \times Y \) such that \( \widetilde{K} \subseteq A \times Y \subseteq \widetilde{U} \) and \( \mu(\widetilde{U} \setminus \widetilde{K}) < \varepsilon \). Let \( K_0 = \pi^X(\widetilde{K}) \cup K \subseteq X \) and \( U_0 = \pi^X(\widetilde{U}) \cap U \subseteq X \). \( \pi^X(\widetilde{U}) \) is compact in \( X \). So, \( K_0 \) is compact in \( X \). Also, \( \pi^X(\widetilde{U}) \subseteq X \) is open. So, \( U_0 \) is open in \( X \). Moreover, \( U_0 \subseteq A \subseteq U_0 \) and \( \mu(U_0 \setminus K_0) < \varepsilon \).

Define \( \varphi : X \to \mathbb{R} \) by \( \varphi(x) = \frac{d_X(x, U_0^c)}{d_X(x, K_0) + d_X(x, U_0)} \) for any \( x \in X \), where \( d_X \) is the metric of \( X \) and \( U_0^c = X \setminus U_0 \). Clearly, \( \varphi \) is continuous, the denominator \( \neq 0 \), 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 \( 0 \leq \varphi \leq 1 \). So, \( \varphi \in C_b(X) \). Note that \( \varphi = 0 \) on \( U_0^c \) and \( \varphi = 1 \) on \( K_0 \). We have now

\[
\int_X \varphi \, d\mu = \int_{K_0} \varphi \, d\mu = \mu(K_0) \geq \mu(A) - \varepsilon.
\]

\[
\int_X \varphi \, d\mu = \int_{U_0} \varphi \, d\mu = \mu(U_0) \leq \mu(A) + \varepsilon.
\]

Since \( K_0 \times Y = \left( \pi^X(\widetilde{K}) \cup K \right) \times Y \subseteq \pi^X(\widetilde{U}) \times Y \subseteq \widetilde{U}, \)
\[ \sum \varphi d\gamma = \sum_{x \in X} \varphi d\gamma = \sum_{y \in Y} \varphi d\gamma = \gamma(K \times Y) \geq \gamma(\hat{K}) \geq \gamma(A \times Y) + \varepsilon. \]

Note that for \((x, y) \in X \times Y\), \((x, y) \in \hat{U} \Rightarrow x \notin \pi_X(\hat{U}) \Rightarrow x \notin U\). So \(\varphi(x) = 0\) if \((x, y) \in \hat{U}\). Hence
\[ \int_X \varphi d\mu = \int_{X \times Y} \varphi d\gamma \text{ by (1)}, \]
\[ |\gamma(A \times Y) - \mu(A)| \leq |\gamma(A \times Y) - \int_X \varphi d\gamma| + |\int_X \varphi d\mu - \mu(A)| \leq 2 \varepsilon. \]
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 \(\varphi : X \times Y \to [0, \infty]\) be lower-semicontinuous. Then there exists \(\hat{\varphi} \in \mathcal{A}(\mu, \nu)\) such that \(E_{\mu}[\hat{\varphi}] = \min_{\gamma \in \mathcal{A}(\mu, \nu)} E_{\mu}[\gamma]\).

**Def.** If \((Z, d)\) is a metric space, then \(f : Z \to \mathbb{R} \cup \{+\infty\}\) is weak-lower semi-continuous means that
\[ \liminf_{k \to \infty} f(x_k) \geq f(x) \text{ if } x_k \to x. \]

**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 \(\mu \to \mu\) narrowly
in \( P(X) \), then
\[
\liminf_{n \to \infty} \int_I f \, d\mu_n \geq \int_I f \, d\mu.
\]
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 \( x \notin G \), 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 \( 0 \leq f_k \leq f_{k+1} \leq \ldots \leq f \wedge k \). Moreover, each \( f_k \) is
Lipschitz continuous, as it is the infimum of a family
(parameterized by \( y \)) of equi-Lipschitz functions on \( X \).

Claim: \( f_k \uparrow f \), i.e. \( f = \sup_k f_k \).Fix \( x \in X \). Assume without loss of generality that \( \sup_k f_k(x) < \infty \). Fix \( x \in X \).s.t.
\[
f(x_k) \wedge k + k \, d(x, x_k) \leq f_k(x) + \frac{1}{k}
\]
This implies that \( d(x, x_k) \to 0 \). Moreover, \( f(x_k) \wedge k \leq f_k(x) + \frac{1}{k} \).
So, by the lower semi-continuity of \( f \),
\[
\sup_{k \geq 1} f_k(x) \geq \liminf_{k \to \infty} (f_k(x) + \frac{1}{k}) \geq \liminf_{k \to \infty} f(x_k) \wedge k \geq f(x).
\]
Now, for each \( k \), since \( d \) is \( \mu \)-narrowly
\[
\liminf_{n \to \infty} \int_X f \, d\mu_n \geq \liminf_{k \to \infty} \int_X f_k \, d\mu_n = \int_X f_k \, d\mu.
\]
Hence, by the monotone convergence theorem,
\[
\liminf_{n \to \infty} \int_X f \, d\mu_n \geq \lim_{k \to \infty} \int_X f_k \, d\mu_n = \int_X f \, d\mu. \quad \Box \]
How to prove the existence of a minimizer of $f \in C(\mathbb{R}^d)$ with $f(\infty) = \infty$?

Direct Method in the Calculus of Variations

**Step 1** $f$ is bounded below. So $\alpha := \inf_{x \in \mathbb{R}^d} f > -\infty$.

**Step 2** $f(\infty) = \infty$. So, $\exists M > 0 \text{ s.t. } ||x|| \leq M \forall x \in \mathbb{R}^d$.

**Step 3** Compactness of $B(0,M) \Rightarrow \exists$ sub-reg. $x_{k_j}$

$\rightarrow x_{\infty} \in \mathbb{R}^d$. Continuity $\Rightarrow f(x_{\infty}) = f(\infty) = \alpha$.

---

**Theorem** Let $Z$ be a reflexive Banach space and $K$ a non-empty, convex, and (strongly) closed subset of $Z$. Assume $f : K \rightarrow (\mathbb{R} \cup \{+\infty\})$ satisfies the following:

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

Then $\exists \hat{x} \in K$ s.t. $f(\hat{x}) = \inf_{x \in K} f(x)$.

**Proof** Let $\alpha := \inf_{x \in K} f > -\infty$. So, $\exists x_k \in K$ s.t. $f(x_k) \rightarrow \alpha$. Since $f(x_k) \geq c_1 \cdot ||x_k|| + c_2 (\forall k)$,
\{x_k\} is bounded. Thus, it has a subsequence not relabeled, such that \(x_k \to \tilde{x}\) weakly for some \(\tilde{x} \in K\). Since \(K\) is convex and (strongly) closed, it is weakly closed. Hence, \(\tilde{x} \in K\). Since \(f\) is weakly lower semi-continuous, \(\liminf_{k \to \infty} f(x_k) \geq f(\tilde{x})\). Hence \(\alpha \geq f(\tilde{x}) \geq \alpha\) and \(-f(\tilde{x}) = \alpha\). \(\boxempty\)

Proof of the existence theorem for the K-OT problem.

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

Since \(E_k : A(\mu, \nu) \to [0, \infty]\) is lower semi-continuous w.r.t. narrow convergence (cf. next lectures),

\[\alpha = \liminf_{j \to \infty} E_k[\gamma_j] \geq E_k[\tilde{\gamma}] \geq \alpha.\] \(\boxempty\)

Questions:

- Narrow convergence/topology?
- \(A(\mu, \nu)\) is narrowly compact?
- \(E_k\) is lower semi-cont. w.r.t. the narrow topology?