Lecture 12 \textbf{Announcements:}

- First wiki articles due this Friday 2/11
- No class Tuesday 2/15, rescheduled to Friday 2/18, 1:30-2:45pm, SH 6635

Recall:

- No class Tuesday 2/15, rescheduled to Friday 2/18, 1:30-2:45pm, SH 6635

Suppose $X$ cpt Polish space.

\textbf{Thm:} For all $\mu, \nu \in \mathcal{P}(X)$, $c : X \times X \rightarrow [0, \infty)$ cts,

$$\inf \{K(\phi) = \sup \int \phi d\mu + \int \psi d\nu \}$$

\text{where} \( (\phi, \psi) \in \mathcal{C}(X) \times \mathcal{C}(X) \)

From Kantorovich back to Monge

Furthermore, the maximum is attained.

Questions:

1. When does an OT map $t(x)$ exist?
2. When do the optima of Monge and Kantorovich's problems coincide?
3. When does $t(x) = \nabla \phi(x)$ for $\phi$ convex?
Thm. (Knott-Smith Optimality Criterion): Fix $X \subseteq \mathbb{R}^d$, $\mu, \nu \in \mathcal{P}(\chi)$. Let $C(x_1, x_2) = |x_1 - x_2|^2$.

(i) There exists $f_\ast \in L^1(\mu)$ proper, lsc, convex s.t.

\[
\inf \left\{ \int f \, d\mu + \int (-f) \, d\nu : f \in C_c(\mathbb{R}^d), f \leq C \right\} = -P_0 = \int |x_1|^2 - 2f_\ast(x) \, d\mu(x) + \int |x_1|^2 - 2f_\ast(x) \, d\nu(x).
\]

(ii) Conversely, if $\gamma \in \mathcal{P}(\mu, \nu)$ and $f \in L^1(\mu)$ proper, lsc, convex for which $x_1 \leq \int f(x) \, d\gamma(x)$ for $\mu$-a.e. $(x_1, x_2)$ then...

\[\text{(ii.a) } \gamma \text{ is optimal} \]
\[\text{(ii.b) } -P_0 = \int |x_1|^2 - f(x) \, d\mu(x) + \int |x_1|^2 - 2f_\ast(x) \, d\nu(x)\]

Remark: More generally, the result continues to hold for $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, i.e.,

\[\mathcal{P}_2(\mathbb{R}^d) = \{ \mu \in \mathcal{P}_2(\mathbb{R}^d) : \int |x|^2 \, d\mu(x) < \infty \}.\]
**Lemma (Double convexification for quadratic cost on \( \mathbb{R}^n \))**

Given \( \{(\varphi, \psi)\} \subseteq C(\mathbb{R}^n) \times C(\mathbb{R}^n) \), \( F((\varphi, \psi), 0) < +\infty \)

Define

\[
\tilde{\varphi}(x) = \inf_{x^2 \in \mathbb{R}^n} |x^1 - x^2|^2 - \psi(x^2),
\]

\[
\tilde{\psi}(x^2) = \inf_{x^1 \in \mathbb{R}^n} |x^1 - x^2|^2 - \tilde{\varphi}(x^2).
\]

Then,

1. \( f(x^1) = \frac{1}{2}(|x^1|^2 - \tilde{\varphi}(x^4)) \in L^1(\mu) \) is proper, lsc, and convex
2. \( f^*(x^2) = \frac{1}{2}(|x^2|^2 - \tilde{\psi}(x^2)) \)
3. \( F((\tilde{\varphi}, \tilde{\psi}), 0) \leq F((\varphi, \psi), 0) \)

**Remark:** \( \tilde{\varphi}_i, \tilde{\psi}_i \) are the Moreau-Yosida regularizations of \( -\psi_i, -\varphi_i \) with respect to the square distance

**Proof:**

First, note that
\[ \Phi(x^2) + \Psi(x^2) \leq |x^1 - x^2|^2 \]
\[ \implies \Phi(x^2) \leq |x^1 - x^2|^2 - \Psi(x^2) \]
\[ \implies \Phi(x^2) \leq \Phi(x^2) \]

Likewise, note that
\[ \Phi(x^2) + \Psi(x^2) \leq |x^1 - x^2|^2 \]
\[ \implies \Psi(x^2) \leq |x^1 - x^2|^2 - \Phi(x^2) \]
\[ \implies \Psi(x^2) \leq \Psi(x^2) \]

Furthermore,
\[ \Phi(x^2) = \inf_{x^2 \in \mathbb{R}^n} \left| x^1 - x^2 \right|^2 - \Psi(x^2) \]

\[ \frac{1}{2} \left( |x^2|^2 - \Phi(x^2) \right) = \sup_{x^2 \in \mathbb{R}^n} \left< x_1, x^2 \right> - \frac{1}{2} \left( |x^2|^2 - \Psi(x^2) \right) \]

By defn of convex conjugate,
\[ f(x^1) = g^*(x^1) \] is proper, lsc, convex.

Likewise, \( g^*(x^2) = f(x^2) \).

\[ \frac{1}{2} \left( 1x^2 \right) = \sup \langle x^1, x^2 \rangle = \frac{1}{2} \left( 1x^1 \right) \ . \]

To see \( f \in L^2(\mu) \), note that Young's inequality implies

\[ g^*(x^2) + g(x^2) \leq \langle x^2, x^2 \rangle \]

Hence, \( g^*(x^2) \leq \langle x^2, x^2 \rangle - \frac{1}{2} (1x^2) \ . \]

OTOH, \( g^*(x^1) = \frac{1}{2} \left( 1x^1 \right) \leq \frac{1}{2} (1x^1) \). 

Thus, \( f = g^* \in L^2(\mu) \).

Now, we have everything we need to prove the Knott-Smith optimality criterion.
Part (i)

By Kantorovich Duality Thm, \( \exists \phi, \psi_0 \in C(X) \) s.t. \( \phi(x^1) + \psi_0(x^2) \leq |x^1 - x^2|^2 \) with

\[
-P_0 = \int \phi \, d\mu + \int \psi_0 \, d\mu = -F((\phi_0, \psi_0), 0).
\]

By Double Convexification Lemma, \( \exists f \in L^1(\mu) \) proper, lsc, convex where

\[
F((\phi_0, \psi_0), 0) = F((\bar{\phi}, \bar{\psi}), 0)
\]

for

\[
f(x^2) = \frac{1}{2} (|x^2|^2 - \bar{\phi}(x^2))
\]

\[
f^*(x^1) = \frac{1}{2} (|x^1|^2 - \bar{\psi}(x^2)).
\]

Thus if \( \delta^* \) is an OT plan,

\[
P_0 = F((\phi_0, \psi_0), 0)
\]

\[
= F((\bar{\phi}, \bar{\psi}), 0)
\]

\[
= F((|x^1|^2 - 2f(x^1), |x^2|^2 - 2f^*(x^2)), 0)
\]
\[ \begin{align*}
&= -\int |x^2| - 2f(x^2) \, d\mu(x^2) - \int |x^2| - 2f^*(x^2) \, d\nu(x^2) \\
&= -\int |x^2| - 2f(x^2) + |x^2| - 2f^*(x^2) \, d\mathcal{E}_\text{y}(x^2, x^2) \\
&\begin{aligned}
\text{Young's} \\
&\geq -\int |x^2| + 1\cdot |x^2| - 2\langle x^2, x^2 \rangle \, d\mathcal{E}_\text{y}(x^2, x^2) \\
&= -\int |x^2 - x^2| \, d\mathcal{E}_\text{y}(x^2, x^2) \\
&= D_0 \\
&= P_0
\end{aligned}
\end{align*} \]

Thus, equality must hold throughout.

Ensures (i.a).

Subtracting eqn below from above,

\[ \int f(x^2) + f(x^2) - \langle x^2, x^2 \rangle \, d\mathcal{E}_\text{y}(x^2, x^2) = 0 \]

Since Young's inequality guarantees the integrand is nonnegative, it must vanish \( \mathcal{E}_\text{y} \)-a.e.
Thus \( x^2 \in \partial f(x^4) \) \( \delta^*\)-a.e.

**Part (ii)**

Suppose
- \( \delta \in \mathcal{M}(\mu, \nu) \)
- \( f \in L^2(\mu) \) proper, lsc, convex
- \( x^2 \in \partial f(x^4) \) \( \delta^*\)-a.e.

**WTS**

(ii.a) \( \delta \) is optimal

(ii.b) \( P_0 = \int |x|^2 - f(x) \, d\mu(x) + \int |x|^2 - 2f^*(x^2) \, d\nu(x^2) \)

Since equality holds in Young's inequality \( \delta^*\)-a.e.

\[
-\int |x|^2 - 2f(x^2) \, d\mu(x^2) - \int |x|^2 - 2f^*(x^2) \, d\nu(x^2)
\]

\[
= \int |x|^2 - 2f(x^4) + |x^4|^2 - 2f^*(x^2) \, d\delta(x^4, x^2)
\]

\[
= -\int |x|^2 - 2\langle x^4, x^2 \rangle + |x^2|^2 \, d\delta(x^4, x^2)
\]

\[
= -\int |x|^2 - |x^2|^2 \, d\delta(x^4, x^2)
\]
For arbitrary $\gamma \in \Gamma(\mu, \nu)$ we have $\varepsilon^* \gamma$. Thus

\[-\int x_1^2 - x_2^2 \Gamma(x_1, x_2)\]

\[= -\int x_1^2 - 2f(x_1) \mu(x_1) - \int x_2^2 - 2f^*(x_2) \nu(x_2)\]

\[\geq -\int x_1^2 - x_2^2 \Gamma(x_1, x_2)\]

Thus $\gamma$ is optimal.

Finally, since $\gamma$ is optimal, which shows (ii.a).

Applying $\varepsilon^*$ again shows (ii.a). $\square$
Now, we have what we need to not only solve the Monge problem, but also to characterize its unique solution.

\[ m = \delta_x, \quad v = 1_{[0,1]} \]

**Thm (Brenier):** Given \( \mu, v \in \mathcal{P}_2(\mathbb{R}^d) \), \( \mu \ll \lambda^d 

1. For any optimal transport plan \( \pi \), \( \exists \) an optimal transport map \( t^* \) s.t. \( \pi = (\text{id}_{t^*})^\# \mu \)
   "Any OT plan is induced by an OT map"

2. Given \( t \) s.t. \( t^\# \mu = v \), \( v \) defined everywhere on \( \mathbb{R}^d \)
   \( t \) is optimal \( \iff t = \nabla \phi \) for \( \phi \in L^1(\mu) \) convex, lsc
   \( \phi \) defined \( \mu \)-a.e., differentiable \( \mu \)-a.e.

3. The OT map is unique, up to \( \mu \)-a.e. equi-
   The proof relies strongly on following theorem:

**Thm:** Given \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \), \( \mu \ll \lambda^d \), \( \phi \in L^1(\mu) \) convex, then
   * \( \phi \) is differentiable \( \mu \)-a.e.
   * where it is differentiable, \( \partial \phi = \{ \nabla \phi \} \)
\[ \nabla \Phi \text{ coincides with the distributional gradient.} \]

Sketch of Proof:
- Any convex function is locally Lip on \( \text{Int}(D(\Phi)) \).
- \( D(\Phi) \) is convex, so \( \partial D(\Phi) \) has Lebesgue meas 0.
- Thus \( \Phi \) is differentiable a.e. on \( D(\Phi) \).
- \( \Phi \in L^1(\mu) \implies \mu(D(\Phi)) = 1 \implies \Phi \text{ is differentiable} \) \( \mu \text{-a.e.} \).

Now solve Monge’s problem!