Lecture 12

Announcements:

- First wiki articles due this Friday 2/11
- No class Tuesday 2/15, rescheduled to Friday 2/18, 1:30-2:45pm, SH6635

Recall:

Suppose $\mathcal{X}$ cpt Polish space.

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

$$\underbrace{\inf_{\gamma \in \Gamma(\mu,\nu)} \mathbb{K}(\gamma)}_{-D_0 \text{ ``Primal''}} = \underbrace{\sup_{\substack{(\varphi,\psi) \in C(\mathcal{X}) \times C(\mathcal{X}) \\ \varphi \oplus \psi \leq c}} \int \varphi \, d\mu + \int \psi \, d\nu}_{-P_0 \text{ ``Dual''}}$$

Furthermore, the maximum is attained.

From Kantorovich back to Monge

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 $t(x) = \nabla \varphi(x)$ for φ convex?

Thm (Knott-Smith Optimality Criterion): Fix $X \subset\subset \mathbb{R}^d$, $\mu, \nu \in \mathcal{P}(X)$. Let $c(x^1, x^2) = |x^1 - x^2|^2$.

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

(i.a) $$\sup_{\substack{\varphi, \psi \in C_b(\mathbb{R}^d) \\ \varphi \oplus \psi \leq c}} \int \varphi \, d\mu + \int \psi \, d\nu = -P_0 = \int |x|^2 - 2 f_*(x) \, d\mu(x) + \int |x|^2 - 2 f_*^*(x) \, d\nu(x)$$

(i.b) For any optimal transport plan γ_*, we have $x^2 \in \partial f_*(x^1)$ for γ_*-a.e. (x^1, x^2)

(ii) Conversely, if $\gamma \in \Gamma(\mu, \nu)$ and $f \in L^1(\mu)$ proper, lsc, convex for which $x^2 \in \partial f(x^1)$ for γ-a.e. (x^1, x^2) then...

(ii.a) γ is optimal

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

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 $\mathbb{R}^d$)

$$-\int_X \varphi \, d\mu - \int_X \psi \, d\nu + \chi_{\{\varphi(x^1)+\psi(x^2) \le |x^1-x^2|^2\}}$$

Given

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

define

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

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

Then,

(i) $f(x^1) = \frac{1}{2}\left(|x^1|^2 - \tilde{\varphi}(x^1)\right) \in L^1(\mu)$ is proper, lsc, and convex

(ii) $f^*(x^2) = \frac{1}{2}\left(|x^2|^2 - \tilde{\psi}(x^2)\right)$

(iii) $F((\tilde{\varphi},\tilde{\psi}),0) \le F((\varphi,\psi),0)$

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

Pf:

First, note that

$$\varphi(x^1) + \psi(x^2) \le |x^1 - x^2|^2$$

$$\Updownarrow$$

$$\varphi(x^1) \le |x^1 - x^2|^2 - \psi(x^2)$$

$$\Updownarrow$$

$$\varphi(x^1) \le \tilde{\varphi}(x^1)$$

Likewise, note that

$$\tilde{\varphi}(x^1) + \psi(x^2) \le |x^1 - x^2|^2$$

$$\Updownarrow$$

$$\psi(x^2) \le |x^1 - x^2|^2 - \tilde{\varphi}(x^1)$$

$$\Updownarrow$$

$$\psi(x^2) \le \tilde{\psi}(x^2)$$

Since, by definition $\tilde{\varphi}(x^1) + \tilde{\psi}(x^2) \le |x^1 - x^2|^2$, we have,

$$\begin{aligned} F((\varphi,\psi),0) &= -\int \varphi \, d\mu - \int \psi \, d\nu \\ &\ge -\int \tilde{\varphi} \, d\mu - \int \tilde{\psi} \, d\nu \\ &= F((\tilde{\varphi},\tilde{\psi}),0) \end{aligned}$$

Furthermore,

$$\tilde{\varphi}(x^1) = \inf_{x^2 \in X} |x^1 - x^2|^2 - \psi(x^2)$$

$$\Updownarrow$$

$$\underbrace{\tfrac{1}{2}\left(|x^1|^2 - \tilde{\varphi}(x^1)\right)}_{g^*(x^1)} = \sup_{x^2 \in X} \langle x^1, x^2 \rangle - \overbrace{\tfrac{1}{2}\left(|x^2|^2 - \psi(x^2)\right)}^{g(x^2)}$$

By defn of convex conjugate,

$f(x^1) = g^*(x^1)$ is proper, lsc, convex.

Likewise,

$$\underbrace{\tfrac{1}{2}\left(|x^2|^2 - \tilde{\psi}(x^2)\right)}_{= g^{**}(x^2) = f^*(x^1)} = \sup_{x^1 \in X} \langle x^1, x^2 \rangle - \overbrace{\tfrac{1}{2}\left(|x^1|^2 - \tilde{\varphi}(x^1)\right)}^{= g^*(x^1) = f(x^1)}$$

To see $f \in L^1(\mu)$, note that Young's inequality implies

$$g^*(x^1) + g(x^2) \geq \langle x^1, x^2 \rangle$$

Hence, $g^*(x^1) \geq \langle x^1, x^2 \rangle - \frac{1}{2}\left(|x^2|^2 - \psi(x^2)\right)$

OTOH, $g^*(x^1) = \frac{1}{2}\left(|x^1|^2 - \tilde{\varphi}(x^1)\right) \leq \frac{1}{2}\left(|x^1|^2 - \varphi(x^1)\right)$

Thus, $f = g^* \in L^1(\mu)$. $\square$

Now, we have everything we need to prove the Knott-Smith optimality criterion.

Pf:

Part (i)

By Kantorovich Duality Thm, $\exists$ $\varphi_0, \psi_0 \in C(X)$ s.t. $\varphi_0(x^1) + \psi_0(x^2) \leq |x^1 - x^2|^2$ with

$$-P_0 = \int \varphi_0 \, d\mu + \int \psi_0 \, d\mu = -F((\varphi_0, \psi_0), 0).$$

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

$$F((\varphi_0, \psi_0), 0) \geq F((\tilde{\varphi}, \tilde{\psi}), 0)$$

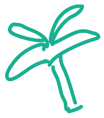
for $f(x^1) = \frac{1}{2}(|x^1|^2 - \tilde{\varphi}(x^1))$
$f^*(x^2) = \frac{1}{2}(|x^2|^2 - \tilde{\psi}(x^2))$.

Thus if γ_* is an OT plan,

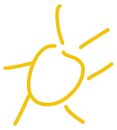
$$P_0 = F((\varphi_0, \psi_0), 0)$$

$$\geq F((\tilde{\varphi}, \tilde{\psi}), 0)$$

$$= F((|x^1|^2 - 2f(x^1), |x^2|^2 - 2f^*(x^2)), 0)$$



$$= -\int |x^1|^2 - 2f(x^1)\, d\mu(x^1) - \int |x^2|^2 - 2f^*(x^2)\, d\nu(x^2)$$



$$= -\int |x^1|^2 - 2f(x^1) + |x^2|^2 - 2f^*(x^2)\, d\gamma_*(x^1, x^2)$$

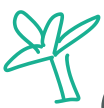
$$\overset{\text{Young}}{\geq} -\int |x^1|^2 + |x^2|^2 - 2\langle x^1, x^2\rangle\, d\gamma_*(x^1, x^2)$$

$$= -\int |x^1 - x^2|^2\, d\gamma_*(x^1, x^2)$$

$$= D_0$$

$$= P_0$$

Thus, equality must hold throughout.

 ensures (i.a).

Subtracting eqn below  from ,

$$\int f(x^1) + f^*(x^2) - \langle x^1, x^2\rangle\, d\gamma_*(x^1, x^2) = 0$$

Since Young's inequality guarantees the integrand is nonnegative, it must vanish γ_*-a.e.

Thus $x^2 \in \partial f(x^1)$ γ_*-a.e.

Part (ii)

Suppose
- $\gamma \in \Gamma(\mu, \nu)$
- $f \in L^1(\mu)$ proper, lsc, convex
- $x^2 \in \partial f(x^1)$ γ-a.e.

WTS
(ii.a) γ is optimal
(ii.b) $-P_0 = \int |x|^2 - f(x) d\mu(x) + \int |x|^2 - 2f^*(x) d\nu(x)$

Since equality holds in Young's inequality γ-a.e.

$$-\int |x^1|^2 - 2f(x^1) d\mu(x^1) - \int |x^2|^2 - 2f^*(x^2) d\nu(x^2)$$

$$= -\int |x^1|^2 - 2f(x^1) + |x^2|^2 - 2f^*(x^2) d\gamma(x^1, x^2)$$

$\downarrow$ Young equality

$$= -\int |x^1|^2 - 2\langle x^1, x^2 \rangle + |x^2|^2 d\gamma(x^1, x^2)$$

$$= -\int |x^1 - x^2|^2 d\gamma(x^1, x^2)$$

For arbitrary $\gamma' \in \Gamma(\mu,\nu)$ we have "$\geq$" at ~.

Thus

$$-\int |x^1 - x^2|^2 d\gamma(x^1, x^2)$$

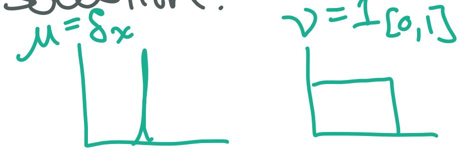
$$= -\int |x^1|^2 - 2f(x^1) d\mu(x^1) - \int |x^2|^2 - 2f^*(x^2) d\nu(x^2)$$

$$\geq -\int |x^1 - x^2|^2 d\gamma'(x^1, x^2)$$

Thus γ is optimal.
Finally, since γ is optimal, which shows (ii.b)

Applying ~ 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.



Thm (Brenier): Given $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, $\mu \ll \mathcal{L}^d$

(1) For any optimal transport plan γ_*, $\exists$ an optimal transport map t_* s.t. $\gamma_* = (id \times t_*)\#\mu$

"Any OT plan is induced by an OT map"

(2) Given t s.t. $t\#\mu = \nu$,

defined everywhere on $\mathbb{R}^d$

t is optimal $\iff$ $t = \nabla\varphi$ for $\varphi \in L^1(\mu)$ convex, lsc

defined on $\mathbb{R}^d$ μ-a.e. differentiable μ-a.e.

(3) the OT map is unique, up to μ-a.e. equiv

The proof relies strongly on following theorem:

Thm: Given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, $\mu \ll \mathcal{L}^d$, $\varphi \in L^1(\mu)$ convex, then

- φ is differentiable μ-a.e.
- where it is differentiable, $\partial\varphi = \{\nabla\varphi\}$

- $\nabla\varphi$ coincides with the distributional gradient.

Sketch of Proof:

- Any convex function is locally Lip on $\mathrm{Int}(D(\varphi))$.
- $D(\varphi)$ is convex, so $\partial D(\varphi)$ has Lebesgue meas 0.
- Thus φ is differentiable a.e. on $D(\varphi)$.
- $\varphi \in L^1(\mu) \Rightarrow \mu(D(\varphi)) = 1 \Rightarrow \varphi$ is differentiable μ-a.e.

Now solve Monge's problem!