Lecture 11

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

Recall:

Suppose X cpt Polish space.

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

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

Furthermore, the maximum is attained.

$$F((\varphi,\psi),u) = -\int\varphi\, d\mu - \int\psi\, d\nu + \chi_{\{u+\varphi\oplus\psi\le c\}}((\varphi,\psi),u)$$

"Double convexification trick"

Prop: Suppose $c: X\times X\to[0,+\infty)$ is continuous.
Given

$$\{(\varphi_i,\psi_i)\}_{i\in I}\subseteq C(X)\times C(X),\quad \psi_i\ge 0,$$
$$\{u_i\}_{i\in I}\subseteq C(X\times X) \text{ unif bdd, e-cts,}$$
$$F((\varphi_i,\psi_i),u_i)<+\infty \quad \forall i\in I.$$

define

$\circ$ $\tilde{\varphi}_i(x^1) = \inf_{x^2 \in X} c(x^1,x^2) - u_i(x^1,x^2) - \psi_i(x^2)$,

$\tilde{\psi}_i(x^2) = \inf_{x^1 \in X} c(x^1,x^2) - u_i(x^1,x^2) - \tilde{\varphi}_i(x^1)$.

Then $\{\tilde{\varphi}\}_{i\in I}, \{\tilde{\psi}\}_{i\in I}$ are unif. bdd, e-cts, and

$$F((\tilde{\varphi}_i, \tilde{\psi}_i), u_i) \leq F((\varphi_i, \psi_i), u_i).$$

Recall: By Arzelá-Ascoli, for $\mathcal{F} \subseteq C(\mathcal{X})$ compact,
$\mathcal{F}$ unif bdd, equicontinuous $\iff$ $\overline{\mathcal{F}}$ compact.

Proof (Kantorovich Duality):

$$F((\varphi,\psi),u) = -\int \varphi \, d\mu - \int \psi \, d\nu + \chi_{\{u+\varphi\oplus\psi\leq c\}}((\varphi,\psi),u)$$

$$\mathcal{P}(u) := \inf_{(\varphi,\psi)\in C(X)\times C(X)} F((\varphi,\psi),u)$$

By theorem on equivalence of primal and dual problems, it suffices to show that...

Step 1: F is convex
Step 2: $\mathcal{P}(0) < +\infty$
Step 3: $\mathcal{P}$ is lsc at 0

... to conclude $P_0 = D_0$.

Step 4: Prove the primal problem has a solution.

<u>Step 1</u>

Since the first two terms in F are linear, hence convex, it suffices to show $\chi_C((\varphi,\psi),u)$ is convex, where

$u(x^1,x^2) + \varphi(x^1) + \psi(x^2) \le c(x^1,x^2)$

$$C = \{((\varphi,\psi),u) : u + \varphi \oplus \psi \le c\}.$$

Thus, it suffices to show C is convex. Fix $((\varphi_0,\psi_0),u_0)$, $((\varphi_1,\psi_1),u_1) \in C$. Their convex combination is

$$(1-\alpha)((\varphi_0,\psi_0),u_0) + \alpha((\varphi_1,\psi_1),u_1)$$
$$= ((\underbrace{(1-\alpha)\varphi_0 + \alpha\varphi_1}_{\varphi_\alpha}, \underbrace{(1-\alpha)\psi_0 + \alpha\psi_1}_{\psi_\alpha}), \underbrace{(1-\alpha)u_0 + \alpha u_1}_{u_0}).$$

Then, for all $\alpha \in [0,1]$,

$$\begin{aligned} u_\alpha + \varphi_\alpha \oplus \psi_\alpha &= (1-\alpha)[u_0 + \varphi_0 \oplus \psi_0] + \alpha[u_1 + \varphi_1 \oplus \psi_1] \\ &\le (1-\alpha)c + \alpha c \\ &= c \end{aligned}$$

So C is convex.

Step 2:

$$\mathcal{P}(0) = \inf_{(\varphi,\psi)\in C(X)\times C(X)} F((\varphi,\psi),0) \leq F((0,0),0) = 0 < +\infty.$$

$\in C(X\times X)$

Step 3: Suppose $u_n \to 0$ uniformly on $X\times X$.
We must show $\liminf_{n\to\infty} \mathcal{P}(u_n) \geq \mathcal{P}(0)$.

Case #1: $\liminf_{n\to\infty} \mathcal{P}(u_n) = +\infty$. Then the inequality automatically holds.

Case #2: $\liminf_{n\to\infty} \mathcal{P}(u_n) < +\infty$. Choose a subsequence u_{n_k} s.t. $\lim_{k\to\infty} \mathcal{P}(u_{n_k}) = \liminf_{n\to\infty} \mathcal{P}(u_n)$.
It suffices to show $\lim_{k\to\infty} \mathcal{P}(u_{n_k}) \geq \mathcal{P}(0)$.
For simplicity of notation, denote the subsequence by $\mathcal{P}(u_n)$.

By defn of infimum, $\forall n\in\mathbb{N}$, $\exists (\varphi_n,\psi_n)\in C(X)\times C(X)$ s.t.

$$+\infty > \mathcal{P}(u_n) \geq F((\varphi_n,\psi_n),u_n) - \frac{1}{n}.$$

Note that, $\forall\ c\in\mathbb{R}$, defining

$\bar{\varphi}_n = \varphi_n + c$, $\bar{\psi}_n = \psi_n - c$, we have

$$F((\varphi_n, \psi_n), u_n) = F((\bar{\varphi}_n, \bar{\psi}_n), u_n).$$

So we may assume, WLOG, $\psi_n \geq 0$ $\forall n \in \mathbb{N}$.

Furthermore, since $u_n \to 0$, $\overline{\{u_n\}} \subseteq C(X \times X)$ is cpt, so $\{u_n\}$ bdd, e-cts.

Thus, by Double Convexification Prop, $\exists$ $\{\tilde{\varphi}_n\}, \{\tilde{\psi}_n\}$ unif bdd e-cts so that

$$+\infty > P(u_n) \geq F((\varphi_n, \psi_n), u_n) - \frac{1}{n}.$$
$$\geq F((\tilde{\varphi}_n, \tilde{\psi}_n), u_n) - \frac{1}{n}.$$

Arzelá-Ascoli guarantees $\exists$ subsequences $\tilde{\varphi}_{n_k}, \tilde{\psi}_{n_k}$ s.t. $\tilde{\varphi}_{n_k} \to \varphi_* \in C(X)$, $\tilde{\psi}_{n_k} \to \psi_* \in C(X)$.

Furthermore, since by defn

$$u_{n_k} + \tilde{\varphi}_{n_k} \oplus \tilde{\psi}_{n_k} \leq c$$

we have

$$\varphi_* \oplus \psi_* \leq c.$$

$\lim_{n\to\infty} P(u_n)$
$\shortparallel$

$$\lim_{k\to\infty} \mathcal{P}(u_{n_k}) \geq \liminf_{k\to\infty} F((\tilde{\varphi}_n, \tilde{\psi}_n), u_n) - \frac{1}{n}.$$

$$= \liminf_{k\to\infty} -\int \varphi_{n_k} d\mu - \int \psi_{n_k} d\nu - \frac{1}{n_k}$$

$$\overset{\text{DCT}}{=} -\int \varphi_* d\mu - \int \psi_* d\nu$$

$$= F((\varphi_*, \psi_*), 0)$$

$$\geq \inf_{(\varphi_*, \psi_*) \in C(X) \times C(X)} F((\varphi_*, \psi_*), 0)$$

$$= \mathcal{P}(0).$$

Thus $\mathcal{P}$ is lsc at zero, so $P_0 = D_0$.

<u>Step 4</u>: It remains to show $\exists$ optimizer for primal problem.

Consider $u_n \equiv 0$ in the previous argument. Taking φ_*, ψ_* as above,

$$\mathcal{P}(0) = \lim_{k\to\infty} \mathcal{P}(u_{n_k}) = F((\varphi_*, \psi_*), 0) \geq \mathcal{P}(0).$$

Thus $P_0 = F((\varphi_*, \psi_*), 0)$, so (φ_*, ψ_*) attains the optimum for the primal problem. $\square$

What about $\mathcal{X}$ noncompact?
Key difficulty: no more Arzelà-Ascoli.

- For any Polish space $(\mathcal{X}, d)$, we still have $P_0 = D_0$.

- In general, need to enlarge the space $C_b(\mathcal{X}) \times C_b(\mathcal{X})$ to get existence of optimizers (φ^*, ψ^*).

Alert: common terminology abuse

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \mathbb{K}(\gamma) = \sup_{\substack{(\varphi,\psi) \in C(\mathcal{X}) \times C(\mathcal{X}) \\ \varphi \oplus \psi \le c}} \int \varphi \, d\mu + \int \psi \, d\nu$$

"Primal" "Dual"

From Kantorovich back to Monge

Ex:

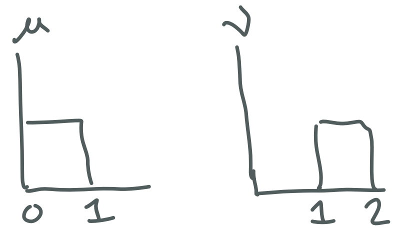
$(\mathcal{X}, d) = (\mathbb{R}, |\cdot|)$, $c(x^1, x^2) = |x^1 - x^2|^2$

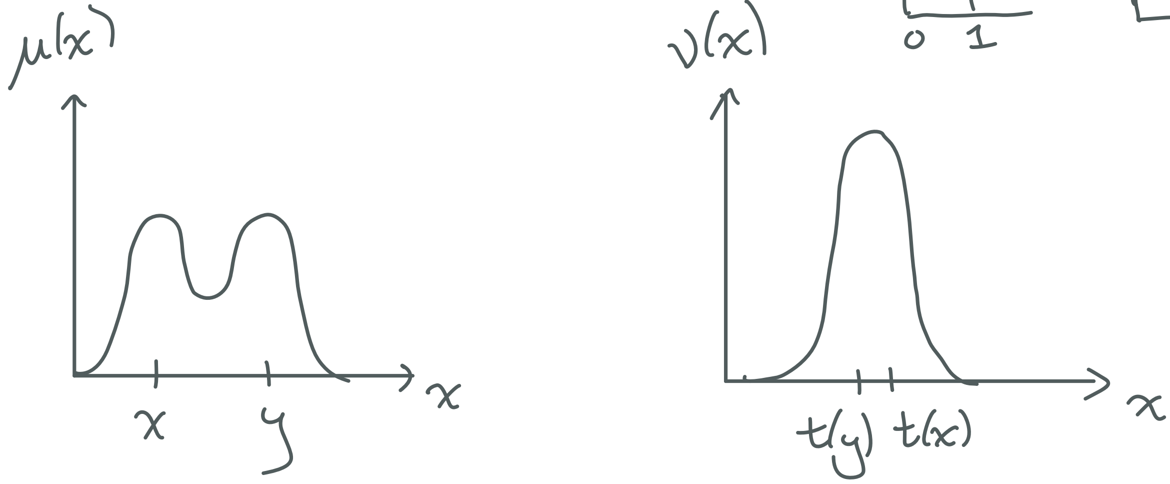
Kantorovich:

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} |x^1 - x^2|^2 \, d\gamma(x^1, x^2)$$

Monge:

$$\inf_{t : t_\# \mu = \nu} \int_{\mathbb{R}} |t(x) - x|^2 \, d\mu(x)$$





Intuitively, if t is an optimal transport map from μ to ν, then "mass doesn't cross", that is

$$x \le y \Rightarrow t(x) \le t(y).$$

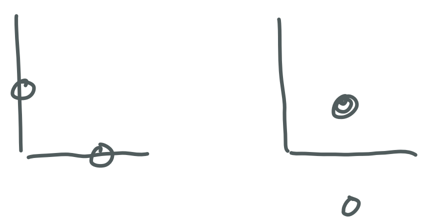
In other words, t is increasing.

What is the appropriate generalization of this property in higher dimensions?

Note: define $\varphi(x) = \int_{-\infty}^{x} t(z)\,dz$.
Then (for t sufficiently smooth, decay at $+\infty$)

- $t(x) = \varphi'(x)$
- $t'(x) = \varphi''(x) \geq 0 \Rightarrow \varphi$ convex

In higher dimensions, we will see that an OT map t satisfies $t =$ "$\nabla\varphi$" for φ convex.



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 $\mathcal{X} \subset\subset \mathbb{R}^d$, $\mu, \nu \in \mathcal{P}(\mathcal{X})$. 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 <u>any</u> 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: Suprisingly, (i.b) does not imply uniqueness of OT plans.

Exercise: Consider $(\mathcal{X}, d) = (\mathbb{R}^2, |\cdot|)$

$$\mu = \tfrac{1}{2}(\delta_{(-1,-1)} + \delta_{(1,1)}), \quad \nu = \tfrac{1}{2}(\delta_{(-1,1)} + \delta_{(1,-1)})$$

Show that $f_*(\underline{x}) = f_*(x_1, x_2) = |x_1 - x_2|$ satisfies (i) above and find two distinct OT plans satisfying (ii).