Optimal Transport: Lecture 3

· Last time, we assumed the existence of a minimax principle to sketch a proof of Kantorovich duality. Today, we'll establish this result more rigorously.

Recall:

Theorem (Kantorovich Duality): Let X, Y be Polish spaces and $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. Let $c: X \times Y \to \mathbb{R}_{\geq 0} \cup \{+\infty\}$ be l.s.c. For any $\pi \in \mathcal{P}(X \times Y)$, $(\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu)$, let us define

$$I[\pi] := \int_{X \times Y} c(x,y)\, d\pi(x,y)$$

$$J(\varphi, \psi) := \int_X \varphi\, d\mu + \int_Y \psi\, d\nu$$

$$\Pi(\mu,\nu) := \left\{ \pi \in \mathcal{P}(X,Y) \;\middle|\; \forall \text{ measurable } A \subset X,\ B \subset Y,\ \begin{array}{l} \pi(A \times X) = \mu(A) \\ \pi(X \times B) = \nu(B) \end{array} \right\},$$

$$\Phi_c := \left\{ (\varphi, \psi) \in L^1(d\mu) \times L^1(d\nu) \;\middle|\; \underset{\mu\text{-a.e. in } x,\ \nu\text{-a.e. in } y}{\varphi(x) + \psi(y) \leq c(x,y)} \right\}$$

Then $\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = \sup_{(\varphi,\psi) \in \Phi_c} J(\varphi, \psi)$ and the infimum on the left is actually achieved. Moreover, Φ_c may, wlog, be restricted to continuous and bounded (φ, ψ).

· To rigorously prove this, we will use [redacted] Fenchel-Rockafellar duality.

Defn: Let E be a normed vector space, $\Theta : E \to \mathbb{R} \cup \{+\infty\}$ a convex function on E, i.e. $\forall (z_1, z_2, \lambda) \in E \times E \times \mathbb{R}$,

$$\Theta(\lambda z_1 + (1-\lambda) z_2) \leq \lambda \Theta(z_1) + (1-\lambda)\Theta(z_2).$$

The Legendre-Fenchel transform of Θ is the function Θ^* defined on the topological dual E^* as

$$\Theta^*(z^*) := \sup_{z \in E} \{ \langle z^*, z \rangle - \Theta(z) \}.$$

ex: Let $E = \mathbb{R}$, $\Theta(z) = e^z$. It is easily seen Θ is convex (exercise) and we see

$$\Theta^*(z^*) = \sup_{z \in \mathbb{R}} \langle z^*, z \rangle - \Theta(z)$$

$$= \sup_{z \in \mathbb{R}} z^* z - e^z,$$ (the dual-pairing is simply the inner product)

which can be solved by calculus: $f(z) := z^* z - e^z$

$$\Rightarrow f'(z) = z^* - e^z = 0 \iff z = \ln(z^*)$$

$$\Rightarrow \Theta^*(z^*) = z^* \ln(z^*) - z^*.$$

Exercise. Prove Θ convex $\Rightarrow \Theta^*$ convex.

Theorem (Fenchel-Rockafellar Duality): Let E be a (separable) normed vector space, E^* its topological dual space, and Θ, Ξ two convex functions on E taking values in $\mathbb{R} \cup \{+\infty\}$. Let Θ^*, Ξ^* be their LF-transforms. Suppose $\exists z_0 \in E$ s.t.

(i) $\Theta(z_0) < +\infty$, $\Xi^*(z_0) < +\infty$

(ii) $\Theta(z)$ is continuous at z_0.

Then $$\inf_{z \in E} \{\Theta(z) + \Xi(z)\} = \max_{z^* \in E^*} [-\Theta^*(-z^*) - \Xi^*(z^*)]$$

<u>Proof</u>: Recalling the definition of the LF-transform, we must show

$$\inf_{z\in E}\left\{\Theta(z)+\Xi(z)\right\} \quad \text{(I)}$$

$$= \sup_{z^*\in E^*}\left[-\Theta^*(-z^*)-\Xi^*(z^*)\right]$$

$$= \sup_{z^*\in E^*}\left[-\sup_{x\in E}\left[\langle -z^*,x\rangle-\Theta(x)\right] - \sup_{y\in E}\left[\langle z^*,y\rangle-\Xi(y)\right]\right]$$

$$= \sup_{z^*\in E^*}\left[\inf_{x\in E}\left[\langle z^*,x\rangle+\Theta(x)\right] + \inf_{y\in E}\left[\langle z^*,-y\rangle+\Xi(y)\right]\right]$$

$$= \sup_{z^*\in E^*}\left\{\inf_{x,y\in E}\left\{\langle z^*,x-y\rangle+\Theta(x)+\Xi(y)\right\}\right\} \quad \text{(II)}$$

If we take $x=y$, then $\sup_{z^*\in E^*}\inf_{x\in E}\left\{\langle z^*,x-x\rangle+\Theta(x)+\Xi(x)\right\}$

$$= \sup_{z^*\in E^*}\inf_{x\in E}\left\{0+\Theta(x)+\Xi(x)\right\}$$

$$= \inf_{x\in E}\left\{\Theta(x)+\Xi(x)\right\},$$

so that (II) $\leq$ (I). To establish (II) $\geq$ (I), we must find a linear form $z^*\in E^*$ s.t.

$$< \infty \text{ by hypoth.}$$

$$(\star)\quad \forall\, x,y\in E,\quad \Theta(x)+\Xi(y)+\langle z^*,x-y\rangle \geq m := \inf_{z\in E}\left[\Theta(z)+\Xi(z)\right]$$

We now show $(\star)$. Let $C := \left\{(x,\lambda)\in E\times\mathbb{R} \mid \lambda > \Theta(x)\right\}$

$$C' := \left\{(y,\mu)\in E\times\mathbb{R} \mid \mu \leq m-\Xi(y)\right\}$$

Since Θ, Ξ are convex, so are $\mathcal{C}$, $\mathcal{C}'$ (exercise). By continuity of Θ at z_0,

$(z_0, \Theta(z_0)+1) \in \operatorname{Int}(\mathcal{C}) \Rightarrow \mathcal{C}$ has non-$\emptyset$ interior $\Rightarrow \overline{\mathcal{C}} = \overline{\operatorname{Int}(\mathcal{C})}$.

By definition of m, $\mathcal{C} \cap \mathcal{C}' = \emptyset$.

Then by Hahn-Banach (see Section 14.2 of Royden), we can find an element of $(E\times\mathbb{R})^*$ that separates $\mathcal{C}$, $\mathcal{C}'$, i.e.

$$\exists\, \ell \in (E\times\mathbb{R})^* \text{ s.t. } \inf_{c\in\mathcal{C}} \langle \ell, c\rangle = \inf_{c\in \operatorname{Int}(\mathcal{C})} \langle \ell, c\rangle$$

$$\geq \sup_{c'\in\mathcal{C}'} \langle \ell, c'\rangle$$

Writing $\ell = (w^*, \alpha)$ for $w^* \in E^*$, $\alpha \in \mathbb{R}^* = \mathbb{R}$, we get that $(w^*, \alpha) \neq (w, \alpha)$ and

$$\left.\begin{array}{l} \lambda > \Theta(x) \\ \mu \leq m - \Xi(y) \end{array}\right\} \Rightarrow \langle w^*, x\rangle + \alpha\lambda \geq \langle w^*, y\rangle + \alpha\mu$$

This requires $\alpha > 0$ (exercise), so defining $z^* := \frac{w^*}{\alpha}$, this yields

$$\langle z^*, x\rangle + \lambda \geq \langle z^*, y\rangle + \mu,$$

and in particular $\langle z^*, x\rangle + \Theta(x) \geq \langle z^*, y\rangle - \Xi(y) + m$, which holds $\forall x, y \in E$. ∎

<u>Proof of Kantorovich Duality</u>: For ease of exposition, we'll focus on the case that X, Y are compact and c is a continuous cost function. Let $E := C_b(X\times Y)$ be the space of all continuous and bounded functions on $X\times Y$, endowed with the sup-norm $\|\cdot\|_\infty$. By the Riesz Representation Theorem (see Section 19.3 in Royden),

The dual of E is the space of regular Radon measures with total variation norm, i.e. $E^* = \mathcal{M}(X \times Y)$, and a nonnegative linear form is given by a regular nonnegative measure.

Define two functionals on $E = C_b(X \times Y)$ $\quad \Theta, \Xi : E \to \mathbb{R} \cup \{+\infty\}$

$$\Theta(u) := \begin{cases} 0 & \text{if } u(x,y) \geq -c(x,y), \\ +\infty & \text{else,} \end{cases}$$

N.B. This is well defined (exercise)

$$\Xi(u) := \begin{cases} \int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y) & \text{if } u(x,y) = \varphi(x) + \psi(y), \\ +\infty & \text{else.} \end{cases}$$

Note that for these choices of E, Θ, Ξ, the assumptions of Fenchel-Rockafellar duality hold with $z_0 \equiv 1$. Hence, we may invoke the associated minimax principal:

$$\inf_{z \in E} \left\{ \Theta(z) + \Xi(z) \right\} = \sup_{z^* \in E^*} \left\{ -\Theta^*(-z^*) - \Xi^*(z^*) \right\}. \quad (\star)$$

Let us break down $(\star)$. The $\boxed{\text{LHS}}$ is

$$\inf_{(\varphi,\psi) \in C_b(X \times Y)} \left\{ \int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y) \;\middle|\; \varphi(x) + \psi(y) \geq -c(x,y) \right\},$$

Since if that constraint doesn't hold, we get $+\infty$. But this is exactly

$$\boxed{-\sup\left\{ J(\varphi,\psi) \;\middle|\; (\varphi,\psi) \in \Phi_c \right\}}.$$

Let us compute Θ^*, Ξ^*:

$\langle u, -\pi \rangle$ $\qquad$ $-\Theta(u)$

$$\underset{\substack{\text{signed measure}\\ \text{on } X\times Y}}{\Theta^*(-\pi)} = \sup_{u \in C_b(X\times Y)} \left\{ -\int u(x,y)\, d\pi(x,y) \;\middle|\; u(x,y) \geq -c(x,y) \right\}$$

$$= \sup_{u \in C_b(X\times Y)} \left\{ \int u(x,y)\, d\pi(x,y) \;\middle|\; u(x,y) \leq c(x,y) \right\}$$

$\rightarrow$ non-negative measures $\subsetneq M(X\times Y)$

Now, suppose $\pi \notin M_+(X\times Y)$. Then $\exists\, v \in C_b(X\times Y)$ s.t. (i) v is non-positive
(ii) $\int v\, d\pi > 0$.

Then $\forall\, \lambda > 0$, $\lambda \cdot v$ satisfies $\lambda \cdot v(x,y) \leq c(x,y)$ and $\int \lambda v\, d\pi \to +\infty$ as $\lambda \to \infty$.

Conversely, if $\pi \in M_+(X\times Y)$, then

$$\sup_{u \in C_b(X\times Y)} \left\{ \int u(x,y)\, d\pi(x,y) \;\middle|\; u \leq c \right\}$$

$$= \int c(x,y)\, d\pi(x,y).$$

We conclude

$$\Theta^*(-\pi) = \begin{cases} \int c(x,y)\, d\pi(x,y), & \text{if } \pi \in M_+(X\times Y) \\ +\infty, & \text{else.} \end{cases}$$

An analogous argument (exercise) yields a matching result for Ξ^*:

$$\Xi^*(\pi) = \begin{cases} 0 & \text{if } \forall (\varphi, \psi) \in C_b(X)\times C_b(Y),\ \int_{X\times Y} [\varphi(x) + \psi(y)]\, d\pi(x,y) = \int_X \varphi(x)\, d\mu(x) + \int_Y \psi(y)\, d\nu(y) \\ +\infty & \text{else} \end{cases}$$

We conclude $\boxed{\text{RHS}}$ is

$$\boxed{\sup_{\pi \in \Pi(\mu,\nu)} \left\{ -\int c(x,y)\, d\pi(x,y) \right\}}$$

Thus,

$$-\sup_{\substack{(\varphi,\psi) \in \\ \Phi_c \cap C_b}} \left\{ J(\varphi,\psi) \right\} = \sup_{\pi \in \Pi(\mu,\nu)} \left\{ -\int c(x,y)\, d\pi(x,y) \right\}$$

$$\Rightarrow \sup_{(\varphi,\psi) \in \Phi_c \cap C_b} \left\{ J(\varphi,\psi) \right\} = \inf_{\pi \in \Pi(\mu,\nu)} \left\{ \int c(x,y)\, d\pi(x,y) \right\}. \blacksquare$$

- Next time we will think algorithmically about discrete Kantorovich problems.