Optimal Transport: Lecture 2

Recall the two formulations of the OT problem: for μ, ν measures on $\mathbb{R}^d$,

$$\text{(MP)} \qquad T^* = \underset{T \text{ s.t. } T_\# \mu = \nu}{\arg\min} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, T(x))\, d\mu(x)$$

$$\text{(KP)} \qquad \pi^* = \underset{\pi \in \Pi(\mu,\nu)}{\arg\min} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\, d\pi(x,y)$$

Remark: Suppose T^* solves (MP). Then let $\pi_T := (id, T)_\# \mu$. Then $\pi_T \in \Pi(\mu, \nu)$. (Exercise). So, in this sense the (MP) formulation is more rigid than (KP).

Remark: In fact, (MP) is strictly more rigid, in the sense that (MP) may not even have a solution. This is most easily seen with some pathological discrete measures.

Exercise: Let $\delta_x \in \mathcal{P}(\mathbb{R}^d)$ be the Dirac measure supported on the singleton set $x \in \mathbb{R}^d$ (i.e., $\int_{\mathbb{R}^d} f(y)\, d\delta_x(y) = f(x)$ for any test function f). Let $\mu = \delta_0$, $\nu = \frac{1}{2}[\delta_{-1} + \delta_1]$. Show there is no map T pushing μ onto ν, and therefore (MP) has no solution.

- We will see that (i) Solution to (KP) exist in general
  (ii) In "nice" cases (depending on properties of μ, ν and the cost function $c: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$), (KP) $\Leftrightarrow$ (MP).

- Our first task is to understand (KP). Once we have understood its mathematical properties through duality, we will discuss practical implementations to handle (KP) and recent computational advances.

=

- The (KP) problem is to minimize $\displaystyle\int_{\mathbb{R}^{d\times d}} c(x,y)\, d\pi(x,y)$ under some constraints.

Note that this problem is linear in π (imagine no constraints and π as an arbitrary measure rather than a probability measure). So, that suggests the problem should be "easy" in the universe of all optimization problems.

- Moreover, the constraints are linear: $\pi \in \Pi(\mu,\nu)$ means π has marginals μ and ν, respectively. We can formulate these as linear constraints by letting $\mathrm{Proj}_x(x,y) = x$, $\mathrm{Proj}_y(x,y) = y$ be projections onto the first and second factors of $\mathbb{R}^d \times \mathbb{R}^d$, respectively, and noting

$$\pi \in \Pi(\mu,\nu) \Leftrightarrow (\mathrm{Proj}_x)_\# \pi = \mu \quad \text{and} \quad (\mathrm{Proj}_y)_\# \pi = \nu$$

(Exercise).

- So, we can formulate (KP) as
$$\min \int_{\mathbb{R}^{d\times d}} c(x,y)\, d\pi(x,y) \text{ s.t. } (\mathrm{Proj}_x)_\# \pi = \mu$$
$$(\mathrm{Proj}_y)_\# \pi = \nu$$

This makes (KP) into an (infinite-dimensional!) linear optimization subject to linear constraints.

. So, <u>duality</u> is a feasible approach to analyzing (KP). There is a big role played by the choice of cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, but we will aim to work at a high degree of generality. To do so, let's recall the notion of (lower) semi-continuity.

<u>Definition</u>: A function $f: \mathbb{R}^d \to \mathbb{R}$ is <u>lower semicontinuous</u> (lsc) at $x_0 \in \mathbb{R}^d$ if

$$\liminf_{x \to x_0} f(x) \geq f(x_0).$$

The function is lsc in general if it is lsc at every point.

<u>Remark</u>: There are many equivalent properties to lsc; we may call on them as needed (see wikipedia and/or Theorem in 9.6 of Evans).

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

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

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

Let $\Pi(\mu,\nu) := \{ \pi \in \mathcal{P}(X \times Y) \mid \forall$ measurable $A \subset X, B \subset Y$, $\pi(A \times X) = \mu(A)$, $\pi(X \times B) = \nu(B) \}$.

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

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 realized. Moreover, Φ_c may, wlog, be restricted to continuous and bounded (φ,ψ).

Remark: Back to the case that most interests us, $X = Y = \mathbb{R}^d$, the Kantorovich duality tells us

$$\underbrace{\min_{\hat{\pi}\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\, d\hat{\pi}(x,y)}_{\text{"cost minimization over couplings"}} = \underbrace{\sup_{\substack{(\varphi,\psi)\in L^1(d\mu)\times L^1(d\nu)\\ \text{s.t. } \varphi(x)+\psi(y)\le c(x,y)}} \int_{\mathbb{R}^d} \varphi(x)\,d\mu(x) + \int_{\mathbb{R}^d} \psi(y)\,d\nu(y)}_{\text{very clearly a linear optimization problem – think of "playing" with } \varphi \text{ and } \psi \text{ separately}}$$

Proof Sketch of Kantorovich Duality: The main idea is to use a minimax principle, which we will rigorously establish later. In brief, we aim to write $\inf_{\pi\in\Pi(\mu,\nu)} I(\pi)$ as an "inf sup" problem, swap to get "sup inf," and show this is equal to $\sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi)$. We can do the first part by making anything that doesn't satisfy the constraint intolerably costly:

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \inf_{\pi\in\mathcal{M}_+(X\times Y)} \left(I[\pi] + \begin{cases} 0, & \pi\in\Pi(\mu,\nu) \\ \infty, & \pi\notin\Pi(\mu,\nu)\end{cases} \right)$$

Here, $\mathcal{M}_+(X\times Y)$ is the space of non-negative Borel measures, which is of course strictly larger than $\mathcal{P}(X\times Y)$ and is in a sense the right/easy space to optimize over.

We can connected this new penalty term to the RHS of Kantorovich duality by noting

(Exercise)

$$\sup_{(\varphi,\psi)\in C_b(X)\times C_b(Y)} \int \varphi(x)\,d\mu(x) + \int \psi(y)\,d\nu(y) - \int [\varphi(x)+\psi(y)]\,d\pi(x,y) = \begin{cases} 0, & \pi\in\Pi(\mu,\nu) \\ +\infty, & \text{else} \end{cases}$$

Thus,

$$\inf_{\pi\in\Pi(\mu,\nu)} J[\pi] = \inf_{\pi\in\mathcal{M}_+(X\times Y)} \Bigg(\sup_{(\varphi,\psi)\in C_b(X)\times C_b(Y)} \overbrace{\int_{X\times Y} c(x,y)\,d\pi(x,y)}^{J[\pi]} + \int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y) - \int_{X\times Y} [\varphi(x)+\psi(y)]\,d\pi(x,y)$$

Now, if we (non-rigorously!) invoke a minimax principle and swap sup and inf, we got:

$$\inf_{\pi\in\Pi(\mu,\nu)} J[\pi] = \sup_{(\varphi,\psi)\in C_b(X)\times C_b(Y)} \Bigg(\inf_{\pi\in\mathcal{M}_+(X\times Y)} \int_{X\times Y} c(x,y)\,d\pi(x,y) + \int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y) - \int_{X\times Y} [\varphi(x)+\psi(y)]\,d\pi(x,y) \Bigg)$$

$$= \sup_{(\varphi,\psi) \in C_b(X)\times C_b(Y)} \left(\inf_{\pi \in \mathcal{M}_+(X\times Y)} \int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y) - \int_{X\times Y} \left[\varphi(x)+\psi(y) - c(x,y)\right] d\pi(x,y) \right)$$

$$= \sup_{(\varphi,\psi) \in C_b(X)\times C_b(Y)} \left(\int_X \varphi(x)\,d\mu(x) + \int_Y \psi(y)\,d\nu(y) - \sup_{\pi \in \mathcal{M}_+(X\times Y)} \int \left[\varphi(x)+\psi(y) - c(x,y)\right] d\pi(x,y) \right)$$

So, to conclude it suffices to show

$$\sup_{\pi \in \mathcal{M}_+(X\times Y)} \int_{X\times Y} \left[\varphi(x)+\psi(y) - c(x,y)\right] d\pi(x,y) = \begin{cases} 0, & (\varphi,\psi) \in \Phi_c \\ +\infty, & \text{else.} \end{cases}$$

We leave this last calculation as an exercise (or see Villani). ∎