Optimal Transport: Lecture #6

- Let's return to the continuum formulation of the (KP). Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$,

$$\Pi(\mu,\nu) = \left\{ \gamma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0} \;\middle|\; \begin{array}{l} \gamma(A \times \mathbb{R}^d) = \mu(A) \\ \gamma(\mathbb{R}^d \times B) = \nu(B) \end{array} \quad \forall A, B \text{ Borel} \right\}.$$

For a general cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, the (KP) solves

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\, d\pi(x,y). \qquad (KP)$$

- So far, we have established (KP) admits a dual formulation, and that it can be solved exactly or approximately when μ, ν are supported on finite, discrete point clouds.

- Our next major task is to establish the existence of a unique solution to (KP).

- In fact, we will establish a much stronger qualitative property of the solution to (KP). We will show the famous Brenier's Theorem: "the solution to (KP) — under certain mild conditions on μ, ν, c — is unique and in fact is supported on the gradient of a convex function."

- We will become more explicit about these ideas, but an important remark is that when the above holds, the optimal π is not diffuse, but concentrates on a set of the form $\{(x, [\nabla u](x)) \mid x \in \mathbb{R}^d\}$ for some convex function u. In this case, ∇u is the Monge map, so the solution to the (relaxed) Kantorovich problem coincides with the solution to the (hard) Monge problem. Amazing!

- To ease the exposition, we'll assume $c(x,y) = \|x-y\|_2^2$.

//

- We will analyze $$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|_2^2 \, d\pi(x,y) \quad (\star)$$ through duality.

we'll set up what this looks like, the develop some background on convex analysis that may or may not be familiar from PDE. Our aim is:

Theorem (OT for Quadratic Cost): Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ have finite second moments, that is, $\int_{\mathbb{R}^d} \|x\|_2^2 \, d\mu(x)$, $\int_{\mathbb{R}^d} \|y\|_2^2 \, d\nu(y) < \infty$.

(i) $\pi \in \Pi(\mu,\nu)$ is optimal for $(\star)$

$$\Updownarrow$$

$\exists\, \varphi$, convex and lower semi-continuous, such that $\operatorname{Supp}(\pi) \subset \operatorname{Graph}(\partial \varphi)$.

For such a φ, the pair (φ, φ^*) is a minimizer to

$$\inf_{(\varphi,\psi)} \left\{ \int_{\mathbb{R}^d} \varphi(x) \, d\mu(x) + \int_{\mathbb{R}^d} \psi(y) \, d\nu(y) \;\middle|\; x \cdot y \leq \varphi(x) + \psi(y) \right\}$$

(ii) If μ does not give mass to small sets (e.g., is a.c. in $\mathbb{R}^d$), then there exists a unique $\hat{\pi}$, optimal for ⊛, which has the form

$$d\hat{\pi}(x,y) = d\mu(x) \cdot \delta_{\{y = \nabla\varphi(x)\}}(y)$$

where $\nabla\varphi$ is ~~[illegible]~~ (unique up to μ-a.e. equivalence) the gradient of a convex function that pushes μ onto ν: $[\nabla\varphi]_\# \mu = \nu$. Moreover, $\mathrm{supp}(\nu) = \overline{\nabla\varphi(\mathrm{supp}(\mu))}$.

(iii) Under the assumptions of (ii), $\nabla\varphi$ is the unique solution to the Monge problem

$$\inf_{\substack{T \text{ s.t.} \\ T_\# \mu = \nu}} \int_{\mathbb{R}^d} \|T(x) - x\|_2^2 \, d\mu(x) \qquad (MP).$$

(iv) If ν also does not give mass to small sets, then for μ-a.e. x and ν-a.e. y,

$$[\nabla\varphi^* \circ \nabla\varphi](x) = x$$
$$[\nabla\varphi \circ \nabla\varphi^*](y) = y$$

and $\nabla\varphi^*$ solves (MP) for pushing ν onto μ, i.e. $\inf_{\substack{T \text{ s.t.} \\ T_\# \nu = \mu}} \int_{\mathbb{R}^d} \|T(x) - x\|_2^2 \, d\nu(x)$

- The first part (i) is known as "Knott-Smith optimality," while the second part is usually called "Brenier's Theorem."

- We need not require μ, ν be a.c., only that they do not give mass to small sets, that is, $\mu(A) = 0$ and $\nu(A)=0$ for any set $A \subset \mathbb{R}^d$ of Hausdorff dimension $\leq d-1$ (i.e. hypersurfaces). Of course, if μ, ν have densities, this holds immediately.
- Let's look at the primal-dual problem for our specific case (★).
- Following Villani, we'll scale $c(x,y)$ by $\frac{1}{2}$ for elegance of certain calculations: $c(x,y) = \frac{1}{2}\|x-y\|_2^2$. So, our functional in the primal problem is

$$J[\pi] := \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}\|x-y\|_2^2 \, d\pi(x,y).$$

Assumption: $M_2 = \int_{\mathbb{R}^d} \frac{1}{2}\|x\|_2^2 \, d\mu(x) + \int_{\mathbb{R}^d} \frac{1}{2}\|y\|_2^2 \, d\nu(y) < +\infty$. This is the condition that μ, ν have finite second moments. This ensures that for any $\pi \in \Pi(\mu,\nu)$,

$$J[\pi] = \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{1}{2}\|x-y\|_2^2 \, d\pi(x,y) \leq \int \left[\|x\|_2^2 + \|y\|_2^2\right] d\pi(x,y)$$

$$= 2M_2$$

$$< \infty.$$

- We have established, through Kantorovich Duality, that

$$\inf_{\pi \in \Pi(\mu,\nu)} J[\pi] = \sup_{(\varphi,\psi) \in \Phi_c} J(\varphi,\psi),$$

and moreover that the infimum on LHS is realized and is actually a minimum.

- Let's recall the dual problem:

$$\sup_{(\varphi,\psi)\in\Phi_c}\left\{\int_{\mathbb{R}^d}\varphi(x)\,d\mu(x)+\int_{\mathbb{R}^d}\psi(y)\,d\nu(y)\right\}$$

The constraint $(\varphi,\psi)\in\Phi_c$ is $\varphi(x)+\psi(y)\leq\frac{\|x-y\|_2^2}{2}$, μ,ν-a.e.

Breaking this down using $\|x-y\|_2^2$

$$\begin{aligned}&=\langle x-y,\,x-y\rangle_{\mathbb{R}^d}\\&=\|x\|_2^2+\|y\|_2^2-2\langle x,y\rangle_{\mathbb{R}^d}\\&=\|x\|_2^2+\|y\|_2^2-2[x\cdot y],\end{aligned}$$

we get $\varphi(x)+\psi(y)\leq\frac{\|x-y\|_2^2}{2}$

$$\Updownarrow$$

$$\varphi(x)+\psi(y)\leq\frac{1}{2}\|x\|_2^2+\frac{1}{2}\|y\|_2^2-x\cdot y$$

$$\Updownarrow$$

$$x\cdot y\leq\underbrace{\left[\frac{1}{2}\|x\|_2^2-\varphi(x)\right]}_{:=\tilde{\varphi}(x)}+\underbrace{\left[\frac{1}{2}\|y\|_2^2-\psi(y)\right]}_{:=\tilde{\psi}(y)}$$

- Let's just use the new dual variables $(\tilde{\varphi},\tilde{\psi})$ and for convenience drop $(\tilde{\varphi},\tilde{\psi})$.

Then the condition $(\varphi,\psi)\in\Phi_c$ is that $x\cdot y\leq\varphi(x)+\psi(y)$.

So, $$\inf_{\pi \in \Pi(\mu,\nu)} I[\pi] = M_2 - \sup_{\pi \in \Pi(\mu,\nu)} \left\{ \iint_{\mathbb{R}^d \times \mathbb{R}^d} x \cdot y \, d\pi(x,y) \right\}$$

and $$\sup_{(\varphi,\psi) \in \Phi_c} J[\varphi,\psi] = M_2 - \inf_{(\varphi,\psi) \in \tilde{\Phi}} \left\{ J[\varphi,\psi] \right\},$$

where as usual $J[\varphi,\psi] = \int_{\mathbb{R}^d} \varphi(x)\,d\mu(x) + \int_{\mathbb{R}^d} \psi(y)\,d\nu(y)$ but now

$$\tilde{\Phi} = \left\{ (\varphi,\psi) \mid \varphi \in L^1(d\mu),\ \psi \in L^1(d\nu),\ x \cdot y \le \varphi(x) + \psi(y) \text{ a.e.} \right\}.$$