Lecture 13 <u>Announcements</u>:

- Select Wiki Article for revision by Mon 2/21.
- Make up class tomorrow: Friday 2/18, 1:30-2:45pm, SH6635

<u>From Kantorovich back to Monge...</u>

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

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

Now, we have what we need to not only solve the Monge problem, but also to characterize its unique solution.

Remark: Recall that $t\#\mu=\nu$

$$\iff \int \varphi(t(x))\,d\mu(x) = \int \varphi(y)\,\underbrace{d\nu(y)}_{d(t\#\mu)(y)} \quad \forall\, \varphi\in L^1(\nu)$$

Note that if $t=s$ μ-a.e., then $t\#\mu=s\#\mu$.

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. $\gamma_*=(id\times t_*)\#\mu$

In particular, $\inf_{t\#\mu=\nu} M_2(t) = \inf_{\gamma\in\Gamma(\mu,\nu)} K_2(\gamma)$.

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

- φ is differentiable μ-a.e.
- where it is differentiable, $\partial\varphi = \{\nabla\varphi\}$
- $\nabla\varphi$ coincides with the distributional gradient.

Now, we can prove Brenier's theorem!

Pf:

CLAIM: If γ_* is an optimal plan and $\gamma_* = (id \times t)\#\mu$ for some t s.t. $t\#\mu = \nu$, then t is an optimal transport map.

Suppose $\tilde{t}$ is another transport map from μ to ν, i.e., $\tilde{t}\#\mu = \nu$. Then $\tilde{\gamma} = (id \times \tilde{t})\#\mu \in \Gamma(\mu, \nu)$.

$$\mathbb{M}_2(t) = \int |t(x) - x|^2 d\mu(x) = \underbrace{\int |x^2 - x^1|^2 d\gamma_*}_{\mathbb{K}_2(\gamma)}$$

$$\leq \int |x^2 - x^1|^2 d\tilde{\gamma} = \int |\tilde{t}(x) - x|^2 d\mu(x)$$

$$= \mathbb{M}_2(\tilde{t})$$

Now prove ①.
Let γ_* be an OT plan. By K-S Thm,
$\exists\ f_* \in L^1(\mu)$ proper, lsc, convex s.t.

$$x^2 \in \partial f_*(x^1) \qquad \gamma_*\text{-a.e.}$$
$$\Updownarrow \text{ by prev Thm}$$
$$x^2 = \nabla f_*(x^1) \qquad \gamma_*\text{-a.e.}$$

Consequently, $\forall\ \varphi \in L^1(\gamma_*)$

$$\int \varphi(x^1, x^2)\, d\gamma_*(x^1, x^2) = \int \varphi(x^1, \nabla f_*(x^1))\, d\gamma_*(x^1, x^2)$$
$$= \int \varphi(x^1, \nabla f_*(x^1))\, d\mu(x^1)$$

Thus $\gamma_* = (id \times \nabla f) \# \mu$.

$r\#(s\#\sigma) = (r \circ s)\#\sigma$

In particular, $\nu = \pi_2 \# \gamma_* = \nabla f \# \mu$,
so $t = \nabla f$ is a transport map from μ to ν.
By our claim, it is an <u>optimal</u> transport map.

Now, prove ②:

"$\Leftarrow$"

Given $t = \nabla\varphi$ for such a φ, define $\gamma = (id \times t)\#\mu \in \Gamma(\mu,\nu)$.

Then

$$\int |x^2 - \nabla\varphi(x^1)|^2 d\gamma(x^1,x^2) = \int |t(x^1) - \nabla\varphi(x^1)|^2 d\mu(x^1)$$
$$= 0$$

so $x^2 = \nabla\varphi(x^1) \in \partial\varphi(x^1)$ γ-a.e..

By K-S thm, γ is optimal.
So CLAIM ensures t optimal.

"$\Rightarrow$"

Given t optimal, $\gamma = (id \times t)\#\mu$ is an optimal plan.

By K-S thm, $\exists\ \varphi_*$ satisfying hypotheses s.t. $x_2 = \nabla\varphi_*(x^1)$ γ-a.e.

Thus,

$$0 = \int |x^2 - \nabla\varphi_*(x^1)|^2 d\gamma(x^1,x^2) = \int |t(x^1) - \nabla\varphi_*(x^1)|^2 d\mu(x^1)$$

Therefore, $t = \nabla \varphi_*$ μ-a.e.

Now show ③
Suppose $t, \tilde{t}$ are OT maps.

By ②, $\exists$ $\varphi, \tilde{\varphi} \in L^1(\mu)$ convex, lsc, proper
s.t. $t = \nabla \varphi$, $\tilde{t} = \nabla \tilde{\varphi}$.

Arguing as before,

$$\gamma := (id \times t) \# \mu, \quad \tilde{\gamma} : (id \times \tilde{t}) \# \mu$$

then $x_2 \in \partial \varphi(x^1)$ γ-a.e., $x_2 \in \partial \tilde{\varphi}(x^1)$ $\tilde{\gamma}$-a.e.
and

$$\int |x^1|^2 - 2\varphi(x^1) + |x^2|^2 - 2\varphi^*(x^2) \, d\tilde{\gamma}(x^1, x^2)$$
$$\|$$
$$\int |x|^2 - 2\varphi(x) d\mu(x) + \int |x|^2 - 2\varphi^*(x) d\nu(x)$$
$$\|$$
$$-P_0$$
$$\|$$
$$\int |x|^2 - 2\tilde{\varphi}(x) d\mu(x) + \int |x|^2 - 2\tilde{\varphi}^*(x) d\nu(x)$$
$$\|$$
$$\int |x^1|^2 - 2\tilde{\varphi}(x^1) + |x^2|^2 - 2\tilde{\varphi}^*(x^2) \, d\tilde{\gamma}(x^1, x^2)$$

Rearranging and recalling that $x^2 \in \partial\tilde{\varphi}(x^1)$ iff equality holds in Young's inequality,

$$\int \varphi(x^1) + \varphi^*(x^2)\, d\tilde{\gamma}(x^1,x^2) = \int \tilde{\varphi}(x^1) + \tilde{\varphi}^*(x^2)\, d\tilde{\gamma}(x^1,x^2)$$

$$= \int x^1 \cdot x^2 \, d\tilde{\gamma}(x^1,x^2)$$

Rearranging,

$$\int \varphi(x^1) + \varphi^*(x^2) - x^1 \cdot x^2 \quad d\tilde{\gamma}(x^1,x^2) = 0$$

$$\shortparallel$$

$$\int \varphi(x^1) + \varphi^*(\nabla\tilde{\varphi}(x^1)) - x^1 \cdot \nabla\tilde{\varphi}(x^1)\, d\tilde{\gamma}(x^1,x^2)$$

$$\shortparallel$$

$$\int \underbrace{\varphi(x^1) + \varphi^*(\nabla\tilde{\varphi}(x^1)) - x^1 \cdot \nabla\tilde{\varphi}(x^1)}_{\geq 0 \text{ by Young}}\, d\mu(x^1)$$

Thus $\varphi(x^1) + \varphi^*(\nabla\tilde{\varphi}(x^1)) - x^1 \cdot \nabla\tilde{\varphi}(x^1) = 0$ μ-a.e.

$$\Updownarrow$$

$$\nabla\tilde{\varphi}(x^1) \in \partial\varphi(x^1) \quad \mu\text{-a.e.}$$

$$\Updownarrow$$

$$\underbrace{\nabla\tilde{\varphi}(x^1)}_{\tilde{t}} = \underbrace{\nabla\varphi(x^1)}_{t} \quad \mu\text{-a.e.}$$

This shows OT maps are unique μ-a.e. $\square$

In this way, the dual of the Kantorovich problem, with $c(x^1,x^2)=|x^1-x^2|^2$ on $\mathbb{R}^d$, helped us solve the Monge problem in this case.

The solution of the Monge problem will then be a key component in proving

$$W_2(\mu,\nu) := \inf_{\gamma\in\Gamma(\mu,\nu)} \mathbb{K}_2(\gamma)$$

is a <u>metric</u> on the space $\mathcal{P}_2(\mathbb{R}^d)$.

But, before we leave duality behind, one last important application to the case $c(x^1,x^2)=|x^1-x^2|$.

Recall...

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

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

$$\underbrace{\qquad\qquad}_{-D_0} \qquad \underbrace{\varphi \oplus \psi \le c \qquad\qquad}_{-P_0}$$

Furthermore, the maximum is attained.

In the particular case $\mathcal{X} \subset\subset \mathbb{R}^d$, $c(x^1, x^2) = |x^1 - x^2|$, we can rewrite this in a nice way.

<u>Thm</u>: Given $\mathcal{X} \subset\subset \mathbb{R}^d$, $\mu, \nu \in \mathcal{P}(\mathcal{X})$,

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \mathbb{K}_1(\gamma) = \sup_{\varphi \in C(\mathcal{X}),\ \|\varphi\|_{Lip} \le 1} \int \varphi \, d(\mu - \nu)$$

and $\exists\ \varphi^*$ that achieves maximum.

This gives another important way to compare probability measures, known as the 1-Wasserstein or Earth Movers Distance,

$$W_1(\mu,\nu) := \inf_{\gamma \in \Gamma(\mu,\nu)} \mathbb{K}_1(\gamma).$$

Remark: The first part of the theorem continues to hold on any Polish space, under the additional constraint that $\varphi \in L^1(|\mu-\nu|)$

Pf: By our duality theorem, it suffices to show

$$\sup_{\substack{\varphi,\psi \in C(X) \\ \varphi(x^1)+\psi(x^2) \le |x^1-x^2|}} \int \varphi \, d\mu + \int \psi \, d\nu = \sup_{\varphi \in C(X),\ \|\varphi\|_{Lip} \le 1} \int \varphi \, d(\mu-\nu)$$

First "$\ge$"

Note that $\varphi \in C(X)$, $\|\varphi\|_{Lip} \le 1$, then $\varphi(x^1)-\varphi(x^2) \le |x^1-x^2|$, so taking $\psi = -\varphi$ gives (φ,ψ) satisfying constraints on LHS w/ values of objective functions equal.

Next "$\le$"

Take (φ^*,ψ^*) that attain maximum on LHS.