Lecture 12

Office Hours
- Today 11-12
- Tomorrow 10-11

Reminders
- Solutions for 2-3 exercises
- First wiki article due Fri Feb 14
- Select article to revise by Fri, Feb 21

Recall:

Thm: Suppose (X, d_X), (Y, d_Y) are cpt Polish spaces and $c \in C(X \times Y)$, $c \geq 0$. Then $\forall \mu \in \mathbb{P}(X)$, $\nu \in \mathbb{P}(Y)$,

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

Furthermore, the maximum is attained.

Exercise 24: Consider the problem

$$\inf_{\substack{x \in \mathbb{R}^m \\ Ax \geq b}} \sum_{i=1}^{m} x_i c_i$$

Find the dual problem

Exercise 25: Consider metric spaces X, Y and $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ given by

$$\mu = \sum_{i=1}^{m} \delta_{x_i} u_i, \quad \nu = \sum_{j=1}^{n} \delta_{y_j} v_j.$$

If $c : X \times Y \to \mathbb{R}$ is the cost fn in KP, show that

$C_{ij} := c(x_i, y_j)$

$$KP = \inf_{P \in \mathcal{U}(u,v)} \sum_{i,j} C_{ij} P_{ij}$$

$$\mathcal{U}(u,v) = \left\{ P \in \mathbb{R}^{m \times n} : P_{ij} \geq 0, \ \sum_j P_{ij} = u_i, \ \sum_i P_{ij} = v_j \right\}$$

Next, prove that the optimum value coincides with the dual problem

$$\sup_{\substack{f \in \mathbb{R}^n, g \in \mathbb{R}^m \\ f_i + g_j \leq C_{ij} \ \forall ij}} \sum_i f_i u_i + \sum_j g_j v_j$$

From Kantorovich back to Monge

Thm (Knott-Smith Optimality)
Fix $\mathcal{X} \subseteq \mathbb{R}^d$ cpt, $c(x,y) = |x-y|^2$
$\mu, \nu \in \mathcal{P}(\mathcal{X})$.

(i) $\exists\, f_\star \in L^1(\mu)$ proper, lsc, convex s.t.

☆

$\boxed{\text{i.a}}$
$$\sup_{\substack{\varphi,\psi \in C_b(\mathcal{X}) \\ \varphi(x)+\psi(y) \le c(x,y)}} \int \varphi \, d\mu + \int \psi \, d\nu = \int |x|^2 - 2 f_\star(x) \, d\mu + \int |y|^2 - 2 f_\star^*(y) \, d\nu$$

$\boxed{\text{i.b}}$ For <u>any</u> OT plan $\gamma_\star$, $y \in \partial f_\star(x)$ $\gamma_\star$-a.e.

same $f_\star$ forall $\gamma_\star$

(ii) Conversely, if $\gamma \in \Gamma(\mu, \nu)$ and $f \in L^1(\mu)$ proper, lsc, convex for which $\underbrace{y \in \partial f(x)}_{\text{equality holds in Young's inequality } \gamma\text{-a.e.}}$ γ-a.e., then...

$\boxed{\text{ii.a}}$ γ is optimal

$\boxed{\text{ii.b}}$ f satisfies ☆

Remark: Though we prove the result in the case X cpt (so that solns of dual problem exist), the result continues to hold for $X = \mathbb{R}^d$ provided that $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, where

$$\mathcal{P}_2(\mathbb{R}^d) = \left\{ \sigma \in \mathcal{P}(\mathbb{R}^d) : \int |x|^2 d\sigma(x) < +\infty \right\}$$

Pf:

Last time, we showed art (i).

Now, show (ii). Suppose γ and f are as in (ii).

For any $\tilde{\gamma} \in \Gamma(\mu, \nu)$, we have

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

$$= \int -|x|^2 + 2f(x) - |y|^2 + 2f^*(y)\,d\tilde{\gamma}(x,y)$$

Young's Inequality

$$\geq \int -|x|^2 + 2\langle x,y\rangle - |y|^2\,d\tilde{\gamma}(x,y)$$

$$= -\int |x-y|^2\,d\tilde{\gamma}(x,y)$$

However, for γ, equality holds throughout. Thus

$$-\int |x-y|^2 d\gamma(x,y)$$
$$= -\int |x|^2 - 2f(x)\,d\mu + \int |y|^2 - 2f^*(y)\,d\nu$$
$$\geq -\int |x-y|^2 d\tilde{\gamma}(x,y)$$

Since $\tilde{\gamma}$ was arbitrary, this shows γ is optimal.

This also shows $\boxed{\text{ii.b}}$. $\square$

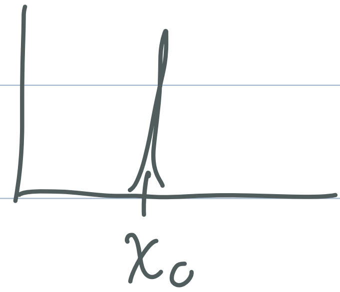
We now have what we need to characterize solns to Monge's Problem.

Thm (Brenier): Given $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, suppose $\mu \ll \mathcal{L}^d$.

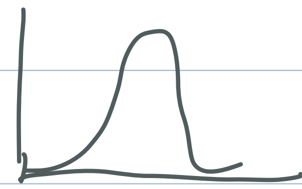
1. If γ_* is an OT plan, $\exists\, t$ meas s.t. $\gamma_* = (id \times t)\#\mu$.
   "Any OT plan is induced by a transport map"
2. t is an OT map $\iff (id \times t)\#\mu$ is an OT plan
3. Given t s.t. $t\#\mu = \nu$, t is optimal $\iff \underbrace{t = \nabla\varphi}_{\mu\text{-a.e.}}$ for $\underbrace{\varphi}_{\text{convex, lsc}} \in L^1(\mu)$
4. the OT map from μ to ν is unique (μ-a.e.)

Remark: While we assume $\mu \ll \mathcal{L}^d$ for our proof, the result continues to hold for any μ s.t. $\mu(S) = 0 \;\forall\, S \in \mathcal{B}(\mathbb{R}^d)$ of finite $d-1$ dim'l Hausdorff measure

Recall:



$\mu = \delta_{x_0}$



ν

We know an OT plan exists.
It's clear that no OT map can exist.
This motivates why μ can't give mass to "small sets"

The proof relies strongly on the following theorems:

<u>Thm</u> (Rademacher): Given $\mathcal{U} \subseteq \mathbb{R}^d$ open and $f: \mathcal{U} \to \mathbb{R}$ Lipschitz, f is differentiable Lebesgue a.e. on $\mathcal{U}$.

Pf: See, e.g., Folland p108 Ex 37

<u>Thm</u>: Given $\mu \in \mathcal{P}(\mathbb{R}^d)$, $\mu \ll \mathcal{L}^d$, $\varphi: \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ convex, $\varphi \in L^1(\mu)$,
(i) φ is differentiable μ-a.e.
(ii) where it is differentiable, $\partial\varphi = \{\nabla\varphi\}$

Exercise 27: Prove part (i).

Now, we can prove Brenier's theorem!

Pf:

First show ①. Suppose $\gamma_\star$ is an OT plan. By K-S thm, $\exists\, f_\star \in L^1(\mu)$ proper, lsc, convex s.t.

$$y \in \partial f_\star(x) \quad \gamma\text{-a.e.}$$

Let $B := \{x \in \mathbb{R}^d : f_\star \text{ is not differentiable}\}$

By previous thm

$0 = \mu(B) = \gamma_\star(B \times \mathbb{R}^d)$,

so $\partial f_\star(x) = \{\nabla f_\star(x)\}$ $\gamma_\star$-a.e.

Thus $\boxed{y = \nabla f_\star(x) \quad \gamma_\star\text{-a.e.}}$ ☆

Consequently, $\forall\, g \in L^1(\gamma_\star)$,

$$\int g(x,y)\,d\gamma^*(x,y) = \int g(x,\nabla f_*(x))\,d\gamma^*(x,y)$$
$$= \int g(x,\nabla f_*(x))\,d\mu(x)$$
$$= \int g(x,y)\,(id\times \nabla f_*)\#\mu$$

Thus, $\gamma^* = (id \times \nabla f_*)\#\mu$.

Note: same f_* for any OT plan γ^*

Now show ②.

"$\Leftarrow$" is exercise 28

"$\Rightarrow$"

Let t be an OT map. Then
$$\mathbb{K}((id\times t)\#\mu) \le \mathbb{K}((id\times s)\#\mu)$$
for any s s.t. $s\#\mu = \nu$.

Let γ_* be an OT plan. Then $\gamma_* = (id\times \nabla f_*)\#\mu$. Thus,

$$\mathbb{K}((id\times t)\#\mu)\le \mathbb{K}(\gamma_*)$$

and $(id\times t)\#\mu$ is an OT plan.

Now show ③. Consider t s.t. $t\#\mu=\nu$.

First, suppose t is optimal. By ②, $\underbrace{(id\times t)\#\mu}_{\gamma}$ is an OT plan. By ①☆,

$$y=\nabla f_*(x) \quad \gamma\text{-a.e.}$$

Thus,

$$0=\int |y-\nabla f_*(x)|\,d\gamma(x,y)$$

$$=\int |t(x)-\nabla f_*(x)|\,d\mu(x),$$

so $t=\nabla f_*$ μ-a.e.

(This shows ④.)

Finally, suppose $t = \nabla\varphi$ for $\varphi \in L^1(\mu)$ convex, lsc.

Define $\gamma := (id \times t)\#\mu \in \Gamma(\mu, \nu)$.
Then,
$\int |y - \nabla\varphi(x)| d\gamma(x,y)$

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

so $y = \nabla\varphi(x) \in \{\partial\varphi(x)\}$ γ-a.e.
By K-S, γ is optimal. By ②, t is an OT map. $\square$

Immediate consequence: Given $\mu \ll \mathcal{L}^d$.
For any $a > 0$, $b \in \mathbb{R}^d$, $t(x) = ax + b$, t is the unique OT map from μ to $t\#\mu$.

We just saw how the dual of KP helped us solve Monge's problem for $c(x,y)=|x-y|^2$.

More generally, similar arguments can be used to show...

<u>Thm</u>: Given $\mathcal{X} \subseteq \mathbb{R}^d$ cpt, $\mu, \nu \in \mathcal{P}(\mathcal{X})$, $c(x,y)=h(x-y)$ for h strictly convex...

- $\exists$ OT plan γ_*
- if $\mu \ll \mathcal{L}^d$, then γ_* is unique and $\gamma_* = (id \times t)\#\mu$ for

OT map

$$t(x) = x - \nabla h^{-1}(\nabla \varphi(x))$$

Pf: Santambrogio, Thm 1.17

One last important application of Kantorovich duality...

$\mathbb{K}_p(\gamma) := \int d(x,y)^p \, d\gamma(x,y)$

<u>Thm:</u> Given $\mathcal{X}$ cpt Polish, $\mu,\nu \in \mathcal{P}(\mathcal{X})$, $c(x,y) = d(x,y)$

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

and $\exists\, \varphi^*$ that achieves max on RHS.

$|\varphi(x) - \varphi(y)| \leq d(x,y) \;\; \forall x,y \in \mathcal{X}$

<u>Remark:</u> The first part of the theorem continues to hold on any Polish space $\mathcal{X}$, under the additional constraint $\varphi \in L^1(|\mu-\nu|)$

Pf: By Kantorovich duality, it suffices to show

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

First "$\geq$":

Suppose $\varphi \in C(X)$, $\|\varphi\|_{Lip} \leq 1$. Then, taking $\psi(y) := -\varphi(y)$, (φ, ψ) satisfies constraint of LHS, and the value of the objective fn is the same.

Other direction next time :)