Lecture 5

Characterising the Optimal Map

RECALL

- Kantorovich prob
  - dual formulation
- using tools from convex analysis to solve dual problem
- symbolic problem:

$$\min L[\phi, \psi] \quad s.t \quad (\phi, \psi) \in \Phi^* \qquad (DP^*)$$

Last time: we can restrict our attn to (ϕ, ψ) which are legendre dual pairs

- we assume that

$$\phi(x) = \sup_{y \in Y} \{ x \cdot y - \psi(y) \}$$

$$\psi(x) = \sup_{x \in X} \{ x \cdot y - \phi(x) \}$$

* Both ϕ, ψ are convex

~~the known~~

QUESTIONS

- we know the feasible set is not empty
- are there minimizers?
  - how can we construct minimizers

To check for a minimizer we first need to check for a finite infimum.

- Is L bounded below?

change of variables

$$L[\phi, \psi] = \frac{1}{2} \int_X |x|^2 d\mu(x) + \frac{1}{2} \int_Y |y|^2 d\nu(y) - J[u, v]$$

~~dual~~ obj. func. for the dual problem

$$\geq \frac{1}{2} \int_X |x|^2 d\mu(x) + \frac{1}{2} \int_Y |y|^2 d\nu(y) - I[\pi]$$

RECALL We said ∃ a relationship b/w values of $J[U, V]$ & values ~~of~~ of our primal obj. func.
→ weak duality

for any feasible $\pi \in \Pi(\mu, \nu)$.

This is my weak duality

NOTE RHS is finite
∴ L is bounded below & infimum will be well-defined

Next want to know if our infimum is a min?

plan
- we'll pick a sequence of φ's & ψ's that cause $L[\varphi, \psi]$ to converge to its infimal value ~~and φ~~
- try to show that these φ & ψ actually converge to something that gives us a concrete minimum

NOTE if φ & ψ are feasible then you can shift things around by a constant

$\rightarrow$ If $(\varphi, \psi) \in \Phi^*$ then $\forall\, a \in \mathbb{R}$

$$(\underbrace{\varphi - a}_{P}, \underbrace{\psi + a}_{R}) \in \Phi^*$$

$\leftarrow$ feasibility cond'n was essentially $P+R$ is a bound on $x \cdot y$. All this NOTE is saying is that if we subtract a & add it back it doesn't change that were satisfying the feasibility ~~condition~~ constraint

$\rightarrow$ This is easy to check
- when considering feasible pairs we can let's only look @ φs that have a min value @ zero
- you can give me any pair I'm just going to shift them around until my φ takes its minimum @ zero

We can restrict to feasible pairs $(\varphi, \psi) \in \Phi$ s.th $\inf_{x \in X} \varphi(x) = 0$

Let's try to actually minimize!

Let $(\varphi_n, \psi_n) \in \Phi^*$ ~~s.t. $L[\varphi_n, \psi_n] \rightarrow \inf$~~

& s.t $L[\varphi_n, \psi_n] \rightarrow \inf_{(\varphi, \psi) \in \Phi} L[\varphi, \psi]$

- (φ_n, ψ_n) are all legendre duals living on bounded sets X, Y

$\Rightarrow$ can say alot about regularity

$\Rightarrow$ uniform Lipschitz bounds which will gives us bounds on the func's themselves

$\rightarrow$ these are equicontinuous (continuous in a strong way)

can use the fact that it's Lipschitz cont. to control amnt. it can increase

$$- 0 \le \varphi \le \underbrace{\mathrm{Lip}(\varphi_n)\, \mathrm{diam}(X)}$$

The Lipschitz const $\times$ diameter of its domain

something to do w/ legendre duals

$$\le \sup_{y \in Y} |y| \cdot \mathrm{diam}(X)$$

- can do the same thing for ψ_n

$\therefore$ we have uniformly bounded sequences

* Since I know my sequence is equicontinuous & uniformly bounded, by Ascoli - Arecla $\exists$ a uniformly convergent subsequence

$$\varphi_{n_k} \to \varphi, \quad \psi_{n_k} \to \psi$$

want to claim that limiting funcs^ of φ, ψ are a minimizer

- need to show that if
  → you plug them into $L[\varphi_n, \psi_n]$ you get the $\inf_{(\varphi_n, \psi_n)} L[\varphi, \psi]$ & need to know that they're feasible

$(\varphi, \psi) \in \Phi^*$ ← from uniform convergence

$$L[\varphi, \psi] = \lim_{k \to \infty} L[\varphi_{n_k}, \psi_{n_k}]$$, we defined these to be minimizing sequences

$$= \inf_{(\varphi, \psi) \in \Phi^*} L(\varphi, \psi]$$

This is the inf over all feasible pairs of my obj. func.

$\Rightarrow \psi$ & φ are minimizers of our transformed dual problem $(DP)^*$

Now, want to pull out this structure & the fact that we can represent these as Legendre duals and try to say something about our original problem.

Let's focus on the (nice enough) case where μ & ν are "nice enough"
i.e. if $|E| = 0$ then $\mu(E) = \nu(E) = 0$ · assuming our measures don't give mass to small sets

We have an optimal pair (φ, ψ) for $(DP)^*$.
- They are convex duals
- They're differentiable a.e

We know that ~~for all x~~ $\forall x \in X$ we can choose some $y \in \partial\varphi(x)$ (subgradient) and have

$$x \cdot y = \varphi(x) + \psi(y)$$

this comes from the fact that φ & ψ are Legendre duals

For a.e $x \in X$, this y is unique: $y = \nabla\varphi(x)$

Let's say $\forall x \in \tilde{X}$, ϕ is differentiable and $|\tilde{X}| = |X|$.

· where $\tilde{X}$ is X ~~minus~~ excluding the points where the func. is not ~~at~~ differentiable.

→ i.e $\tilde{X}$ has full measure. So whatever points I've removed are measure zero

<u>RECALL</u> In the Monge problem we wanted a map that would take a point $x \in X$ and give us a point somewhere y.

* At this point we have ~~somethin~~ something, $y = \nabla\varphi(x)$, that does that almost everywhere

↳ have found a natural relationship b/w points in X & points in Y

Let's define a map $T(x)$ for $x \in X$ so that $T(x) \in \partial\varphi(x)$

- for $x \in \tilde{X}$, $T(x) = \nabla\varphi(x)$

  we have no choice here, only have options @ pts. of measure zero

Does this map solve the Monge formulation (AKA OT problem)?

• Need to check

① mass conservation?

② is it a minimizer?

① Is T measure-preserving?

· Trying to show

$$\int_X h(T(x))\, d\mu(x) = \int_Y h(y)\, d\nu(y) \quad \forall\, h \in C^0(\bar{Y})$$

for all functions, h, the are C-zero (???) on the closure of this set

⇒ somethings that's cont. on a bounded set will be bounded itself.

Introduce:

$$\psi_\varepsilon(y) = \psi(y) + \varepsilon h(y)$$

$$\varphi_\varepsilon(x) = \sup_{y\in Y} \{x\cdot y - \psi(y) - \varepsilon h(y)\}$$

Legendre duals ⇒ $(\varphi_\varepsilon, \psi_\varepsilon) \in \Phi$

So we've started w/ continuous functions and perturbed them in a way that will ~~give us~~ preserve the continuity and that's guaranteed to give us the constraints that ~~—~~ we need

Since (φ, ψ) are an optimal pair:

$$0 \le \frac{L[\varphi_\varepsilon, \psi_\varepsilon] - L[\varphi, \psi]}{\varepsilon}$$

minimizer

dividing by ε

from eqn's we introduced on prev. page

$$= \underbrace{\int_X \frac{\varphi_\varepsilon(x) - \varphi(x)}{\varepsilon} d\mu(x)} + \int_Y h(y)\, d\nu(y)$$

Need to make some argument about limiting value of this integral

NOTE

- we could integrate over $\tilde{X}$ instead of X

- For $x \in \tilde{X}$, $\varphi(x) = \sup_{y \in Y} \{x \cdot y - \psi(y)\}$

$\rightarrow$ there is a unique maximizer @ $y = T(x) = \nabla \varphi(x)$

$$\varphi(x) = x \cdot T(x) - \psi(T(x))$$

*There's only one possible for that we can put for $T(x)$

- For $x \in \tilde{X}$, $\varphi_\varepsilon(x) = \sup_{y \in Y} \{x \cdot y - \psi(y) - \varepsilon h(y)\}$

have perturbed the above by something continuous

$\Rightarrow$ This has a maximizer $y_\varepsilon = T(x) + o(1)$

something small

$\Rightarrow$ new maximizer will be close to that from above

$$\varphi_\varepsilon(x) = x \cdot y_\varepsilon - \psi(y_\varepsilon) - \varepsilon h(y_\varepsilon)$$

- For $x \in \tilde{X}$

$$\varphi_\varepsilon(x) - \varphi(x) = \underbrace{x \cdot y_\varepsilon - \psi(y_\varepsilon)}_{\le \varphi(x)} - \varepsilon h(y_\varepsilon) - \varphi(x)$$

$$\rightarrow \varphi_\varepsilon(x) - \varphi(x) \le -\varepsilon h(y_\varepsilon)$$

$$= -\varepsilon h[T(x) + o(1)]$$

h is cont. $\rightarrow$ pull out $o(1)$

$$\stackrel{?}{=} -\varepsilon h(T(x)) + o(\varepsilon)$$

something smaller than ε

* Need to make this work in the other direction to ensure equality

$$\varphi_\varepsilon(x) - \varphi(x) \geq x \cdot T(x) - \psi(T(x)) - \varepsilon h(T(x))) - \varphi(x)$$

$$= \varphi(x) - \varepsilon h(T(x))) - \varphi(x)$$

$$= -\varepsilon h(T(x))$$

← Now essentially have same thing on RHS & LHS of inequality facing both directions (have small error)

$$\frac{\varphi_\varepsilon(x) - \varphi(x)}{\varepsilon} = -h(T(x)) + o(1)$$

← Now want to integrate & pass through the limit

If we want to integrate $\frac{\varphi_\varepsilon - \varphi}{\varepsilon}$ can use <u>dominated convergence theorem.</u>

bounded func. by def'n

$$\frac{\varphi_\varepsilon(x) - \varphi(x)}{\varepsilon} \leq -h(y_\varepsilon) \leq \|h\|_\infty$$

← infinity norm(???)

other direction: $\frac{\varphi_\varepsilon(x) - \varphi(x)}{\varepsilon} \geq -h(T(x)) \geq \|h\|_\infty$

$$\therefore \int_X \frac{\varphi_\varepsilon(x) - \varphi(x)}{\varepsilon} \, d\mu(x) = -\int_X h(T(x)) \, d\mu(x)$$

Going back to our ~~optimal~~ argument about the limiting value of $\int_X \frac{\varphi_\varepsilon(x) + \varphi(x)}{\varepsilon} \, d\mu(x)$

$$0 \leq \int_X \frac{\varphi_\varepsilon(x) + \varphi(x)}{\varepsilon} \, d\mu(x) + \int_Y h(y) \, d\nu(y)$$

taking $\varepsilon \to 0$ we get

$$0 \leq -\int_X h(T(x)) \, d\mu(x) + \int_Y h(y) \, d\nu(y)$$

↖ want equality here!

- How do I make this inequality work in the other direction
→ repeat w $(-h)$

$$\therefore \int_X h(T(x)) \, d\mu(x) = \int_Y h(y) \, d\nu(y) \quad \text{OR} \quad T_\# \mu = \nu$$

② Is T optimal?

The cost we want to minimize is

$$C(S) = \frac{1}{2}\int_X |x - S(x)|^2 \, d\mu(x)$$

Choose any S s.t $S_\# \mu = \nu$

$$C(S) = \int_X \left(\frac{1}{2}|x|^2 + \frac{1}{2}|S(x)|^2 - \cancel{2}x \cdot S(x)\right) d\mu(x)$$

has to be b/c it's feasible

$$S(x) \in Y \Rightarrow x \cdot S(x) \le \varphi(x) + \psi(S(x))$$

(constraint from $(DP)^*$)

$$C(S) \ge \int_X \left(\frac{1}{2}|x|^2 + \frac{1}{2}|S(x)|^2 - \varphi(x) - \psi(S(x))\right) d\mu(x)$$

measure preservation

$$= \int_X \left(\frac{1}{2}|x|^2 - \varphi(x)\right) d\mu(x) + \int_Y \left(\frac{1}{2}|y|^2 - \psi(y)\right) d\nu(y)$$

$$= \int_X \left[\frac{1}{2}|x^2| + \frac{1}{2}|T(x)|^2 - \varphi(x) - \psi(T(x))\right] d\mu(x)$$

$$= \int_X \left(\frac{1}{2}|x|^2 + \frac{1}{2}|T(x)|^2 - x \cdot T(x)\right) d\mu(x)$$

$$= C(T)$$

$\therefore T$ is optimal

By solving the $(DP)^*$ we've seen there's a relationship b/w the sol'n the $(DP)^*$ & the sol'n to our original Monge problem. Have been able to find some structure.
- T a.e. looks like gradient of convex func

we know: $\exists$ a convex func. φ s.t the optimal map $T(x) = \nabla\varphi(x)$ a.e $x \in X$

*This was needed to give us the Monge-Ampere eqn

If $\underbrace{d\mu(x) = f(x), \; d\nu(y) = g(y)\,dy}_{\text{densities}}$, then mass conservation gives us

$$g(T(x)) \det(\nabla T(x)) = f(x)$$

$$\Rightarrow \begin{cases} \det(\underset{\substack{\uparrow \\ \text{hessian}}}{\nabla^2} \varphi(x)) = \dfrac{f(x)}{g(\nabla \varphi(x))} \\ \varphi \text{ is convex} \end{cases}$$