

Lecture 11

- Midterm week: 9 Nov
- Office hours: TuTh 2-3 pm

Today:

- * The Legendre transform in $\mathbb{R}^n$
- * Brenier's theorem ($c(x,y) = |x-y|^2$)

Assume $\phi = \psi^*$, $\psi = \phi^*$. From

$$\phi(x) = \max \{ x \cdot y - \psi(y) \}$$

we have that

$$\phi(x) + \psi(y) \geq x \cdot y \quad \forall x, y$$

where, for each fixed x_0 (or y_0') there is at least one y_0' (or x_0') for which " $=$ " holds.

Moreover,

$$x \in \partial \psi(y) \iff y \in \partial \phi(x)$$

In what follows, ϕ is assumed to be locally bounded in its domain of definition.

Proposition: The set $\partial \phi(x)$ is always closed

Proof: If $\{y_k\} \in \partial \phi(x_0) \quad \forall k$, and $y_k \rightarrow y_\infty$, then for each k and each x we have

$$\phi(x) \geq \phi(x_0) + y_k \cdot x - y_k \cdot x_0$$

$$\Rightarrow \phi(x) \geq \phi(x_0) + y_\infty \cdot x - y_\infty \cdot x_0$$

$$\Rightarrow y_\infty \in \partial \phi(x_0)$$



Proposition: If $x_k \rightarrow x_\infty$, and $y_k \in \partial \phi(x_k) \quad \forall k$, then any limiting point of $\{y_k\}$ belongs to $\partial \phi(x_\infty)$.

Proof: We have $\forall k, \forall x$

$$\phi(x) \geq \phi(x_k) + y_k \cdot x - y_k \cdot x_k$$

$$\stackrel{(k \rightarrow \infty)}{\uparrow} \Rightarrow \phi(x) \geq \phi(x_\infty) + y_\infty \cdot x - y_\infty \cdot x_\infty$$

along a subsequence

$$\Rightarrow y_\infty \in \partial \phi(x_\infty)$$



Proposition: Suppose $\phi: D \rightarrow \mathbb{R}$, ($D \subset \mathbb{R}^n$)
 is convex. Then ϕ is pointwise C^1 at $x_0 \in D$
 $\Leftrightarrow \partial\phi(x_0)$ has a single element

Proof: ($\Leftarrow$) $\phi(x_k) - p \cdot (x_k - x_0) \geq \lambda |x_k - x_0|$, $x_k \neq x_0$

whos, the $\hat{x}_k (= \frac{x_k - x_0}{|x_k - x_0|})$ converge to some $e \in S^{n-1}$,

Then,

$$\phi(x) \geq \phi(x_0) + y_0 \cdot (x - x_0)$$

Take $x_k \rightarrow x_0$. Then choose $y_k \in \partial\phi(x_k)$, so

$$\phi(x_0) \geq \phi(x_k) + y_k \cdot (x_0 - x_k)$$

$$\text{or,} \quad \phi(x_k) \leq \phi(x_0) + y_k \cdot (x_k - x_0)$$

$$\phi(x_k) \leq \phi(x_0) + y_0 \cdot (x_k - x_0) + (y_k - y_0) \cdot (x_k - x_0)$$

Since $\partial\phi(x_0)$ has just one element, the sequence $\{y_k\}$ converges, and it converges to y_0 .

This shows that

$$\phi(x_k) - (\phi(x_0) + y_0 \cdot (x_k - x_0))$$

vanishes faster than $|x_k - x_0|$

($\Rightarrow$) **Exercise**.

From this proposition we also see that if $x_0 \in \partial\phi(y_0)$, then ϕ is strictly convex at x_0 if and only if ϕ is pointwise C^1 at y_0 .

Suppose : π_0 is optimal for the Kantorovich problem

$$\min \int \int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|^2 d\pi_0(x,y)$$

$$\text{s.t.} \quad \pi_0 \geq 0, \quad \text{Pr}_1(\pi_0) = \mu, \\ \text{Pr}_2(\pi_0) = \nu.$$

Then, we know there are functions ϕ_0, ψ_0 such that $\phi_0 = (\psi_0)^c$ and $\psi_0 = (\phi_0)^c$ with $C(x,y) = |x-y|^2$, and

$$\phi_0(x) + \psi_0(y) = |x-y|^2 \quad \pi_0\text{-a.e.}$$

Observe that if we define

$$u_0(x) := \frac{1}{2}(|x|^2 - \phi_0(x))$$

$$v_0(y) := \frac{1}{2}(|y|^2 - \psi_0(y))$$

then $u_0 = v_0^*$ and $v_0 = u_0^*$.

Then, since ϕ_0, ψ_0 maximize

$$\int_{\mathbb{R}^n} \phi d\mu + \int_{\mathbb{R}^n} \psi d\nu, \quad \phi(x) + \psi(y) \leq |x-y|^2$$

the pair (u_0, v_0) minimizes

$$\int_{\mathbb{R}^n} u d\mu + \int_{\mathbb{R}^n} v d\nu, \quad u(x) + v(y) \geq x \cdot y$$

Here, we reproduce the argument from the last lecture, but this time proving the "working assumption". That we can do this is thanks to Rademacher's theorem.

We recall what we mean by pointwise differentiability: $\phi : D \rightarrow \mathbb{R}$ ($D \subset \mathbb{R}^n$, open) is pointwise differentiable at $x_0 \in D$ if $\exists p \in \mathbb{R}^n$ s.t.

$$\lim_{x \rightarrow x_0} \frac{\phi(x) - \phi(x_0) - p \cdot (x - x_0)}{|x - x_0|} = 0$$

Theorem (Rademacher)

Suppose $\phi: \mathcal{D} \rightarrow \mathbb{R}^2$ is Lipschitz continuous, then the set of points in $\mathcal{D}$ where ϕ is not pointwise differentiable has measure zero.

We can now apply this to u_0 and v_0 . In particular, u_0 is pointwise differentiable a.e., which means $\partial u_0(x)$ has a single element for μ -a.e. x . Here, we are using the assumption $\mu \ll dx$.

In particular, for such x_0 's, there is a unique y_0 s.t.

$$\max_y \{ x_0 \cdot y - V_0(y) \} = x_0 \cdot y_0 - V_0(y_0)$$