OT: Lecture #20

- Last time, we established the continuity (or at least lsc.) properties of certain functionals on Wasserstein space. This is important for showing existence of solutions to problems of the form $\underset{\mu \text{ s.t. constraints hold}}{\operatorname{argmin}} F(\mu)$.

- To develop uniqueness results, we want to understand the convexity of these functionals and how to differentiate them in a sense.

- Let's examine the convexity of a few functionals:

Potential Energy: For $V : \Omega \to \mathbb{R}^d$, $\mathcal{V}(\mu) := \int_\Omega V(x)\,d\mu(x)$. This is linear in μ,

so $\mathcal{V} : \mathcal{P}(\Omega) \to \mathbb{R}$ is automatically convex, but not strictly convex.

Interaction Energy: For $W : \Omega \times \Omega \to \mathbb{R}$, let $\mathcal{W}(\mu) := \int_\Omega \int_\Omega W(x,y)\,d\mu(x)\,d\mu(y)$.

This is not necessarily convex, even for "nice" ~~[illegible]~~ interaction functions W. Indeed, let $W(x,y) = \|x-y\|^2$. Then:

$$\mathcal{W}(\mu) = \int_\Omega \int_\Omega \|x-y\|^2\,d\mu(x)\,d\mu(y)$$

$$= \int_\Omega \int_\Omega \|x\|^2\,d\mu(x)\,d\mu(y) + \int_\Omega \int_\Omega \|y\|^2\,d\mu(x)\,d\mu(y) - 2\int_\Omega \int_\Omega x \cdot y\,d\mu(x)\,d\mu(y)$$

$$= 2\int_\Omega \|x\|^2\,d\mu(x) - 2\left|\int_\Omega x\,d\mu(x)\right|^2 \Big\}, \text{ interpreted as } \cdot \text{ with itself}$$

which we note is a linear functional $\left(\mu \mapsto \int_\Omega \|x\|^2 \, d\mu(x)\right)$ subtracting off the square of a linear functional $\left(\mu \mapsto \iint_{\Omega\,\Omega} x\cdot y \, d\mu(x)\,d\mu(y) = \left[\int_\Omega x \, d\mu(x)\right]^2\right)$, and is in fact concave, not convex.

Transport Functionals: Because they can be written in the dual form as a supremum, these are in fact convex:

$$\mathcal{T}_c(\mu,\nu) = \sup_{\varphi(x)+\psi(y)\le c(x,y)} \int \varphi(x)\,d\mu(x) + \int \psi(y)\,d\nu(y)$$

If the cost function is "nice," we get strict convexity (e.g. if $c(x,y)=\|x-y\|^p$, $p>1$).

Dual norms: Norms are automatically convex but never strictly convex because they are 1-homogeneous: $\|\alpha\mu\|_{\mathcal{X}'} = \alpha\|\mu\|_{\mathcal{X}'}$ $\forall \alpha$ scalar.

// How to Think about "differentiating" such functionals"?

Defn: Let $F:\mathcal{P}(\Omega)\to\mathbb{R}\cup\{+\infty\}$. We say $\rho\in\mathcal{P}(\Omega)$ is regular for F if $\forall\varepsilon\in[0,1]$, $F([1-\varepsilon]\rho + \varepsilon\tilde\rho) < +\infty$ for every $\tilde\rho \in \mathcal{P}(\Omega)\cap L^\infty_c(\Omega)$, (i.e. $\tilde\rho$ is a probability measure that has L^∞ and compactly supported density)

. Note that if F is regular at ρ, then we let $\frac{\delta F}{\delta\rho}(\rho)$ be the (any) measurable function satisfying

$$\frac{d}{d\varepsilon}\left[F(\rho+\varepsilon\chi)\right]\Big|_{\varepsilon=0} = \int \frac{\delta F}{\delta\rho}(\rho)\,d\chi$$

for any $\chi = \tilde{\rho} - \rho$ with $\tilde{\rho} \in L^\infty_c(\Omega) \cap \mathcal{P}(\Omega)$. We call such χ a perturbation and we call $\frac{\delta F}{\delta \rho}(\rho)$ the first variation of F.

__Note__: Since $\rho, \tilde{\rho} \in \mathcal{P}(\Omega)$ we have that $\int_\Omega d\chi = \int_\Omega d\tilde{\rho} - \int_\Omega d\rho = 0$, so that $\frac{\delta F}{\delta \rho}(\rho)$ is defined up to additive constants, and moreover it is unique modulo additive constants.

• Let's examine our functionals of interest and compute their first variations.

__ex__: $\mathcal{V}(\mu) = \int_\Omega V(x)\,d\mu(x)$. Well, $\mathcal{V}(\mu+\epsilon\chi) =$

$$\int_\Omega V(x)\,d[\mu+\epsilon\chi](x)$$

$$= \int_\Omega V(x)\,d\mu(x) + \epsilon \int_\Omega V(x)\,d\chi(x).$$

$$= \mathcal{V}(\mu) + \epsilon\, \mathcal{V}(\chi)$$

$$\Rightarrow \frac{d}{d\epsilon} \mathcal{V}(\mu+\epsilon\chi) = \mathcal{V}(\chi)$$

$$\Rightarrow \left[\frac{d}{d\epsilon} \mathcal{V}(\mu+\epsilon\chi)\right]\bigg|_{\epsilon=0} = \mathcal{V}(\chi)$$

$$= \int_\Omega V(x)\,d\chi(x)$$

$$\Rightarrow \frac{d\cancel{\#}V}{d\rho}(\rho) = V \text{ ; easy!}$$

ex: $\cancel{\#}\underset{W}{}(\mu) = \int_\Omega\int_\Omega W(x,y)\,d\mu(x)\,d\mu(y)$

$$\Rightarrow \overset{W}{\cancel{\#}}(\mu+\varepsilon\chi) = \int_\Omega\int_\Omega W(x,y)\,d[\mu+\varepsilon\chi](x)\,d[\mu+\varepsilon\chi](y)$$

$$= \int_\Omega\int_\Omega W(x,y)\,d\mu(x)\,d\mu(y) + \varepsilon\int_\Omega\int_\Omega W(x,y)\,d\mu(x)\,d\chi(y)$$

$$+ \varepsilon\int_\Omega\int_\Omega W(x,y)\,d\chi(x)\,d\mu(y) + \varepsilon^2\int_\Omega\int_\Omega W(x,y)\,d\chi(x)\,d\chi(y).$$

$$\Rightarrow \frac{d}{d\varepsilon}\left[W(\mu+\varepsilon\chi)\right]\Big|_{\varepsilon=0} = \int_\Omega\int_\Omega W(x,y)\,d\chi(x)\,d\mu(y) + \int_\Omega\int_\Omega W(x,y)\,d\mu(x)\,d\chi(y)$$

$$= \int_\Omega \underbrace{\left[\int_\Omega W(y,y')\,d\mu(y') + \int_\Omega W(x,y)\,d\mu(x)\right]}_{\parallel} d\chi(y)$$

This is therefore the pointwise evaluation of the first variation at y: $\frac{\delta W}{\delta\rho}(\rho)(y)$

If $W(x,y) = W(y,x)$, then $\exists h$ s.t. $W(x,y) = h(x-y)$ for some even function h, and we have

$$\frac{\delta W}{\delta \rho}(\rho) = 2h * \mu$$

Ex: Suppose $\mathcal{T}_c(\mu) = \inf_{\pi \in \Pi(\mu,\nu)} \iint_{\Omega \times \Omega} c(x,y)\, d\pi(x,y)$ is the transportation cost

implicitly depends on ν!!

from μ to a fixed measure ν. We can show (but will not; see OTAM Prop 7.17) that the first variation of $\mathcal{T}_c$ corresponds to the optimal dual variable, in the following sense.

By duality, if Ω is compact, $\mathcal{T}_c(\mu) = \sup_{\varphi \in C^0(\Omega)} \left\{ \int_\Omega \varphi \, d\mu + \int_\Omega \varphi^c \, d\nu \right\}$. Suppose $\exists!\ \varphi_\mu \in C(\Omega)$ s.t. $\sup_{\varphi \in C^0(\Omega)} \left\{ \int_\Omega \varphi\, d\mu + \int_\Omega \varphi^c\, d\nu \right\} = \int_\Omega \varphi_\mu \, d\mu + \int_\Omega \varphi_\mu^c \, d\nu$. Then

$$\frac{\delta \mathcal{T}_c}{\delta \rho}(\mu) = \varphi_\mu .$$

Why care about $\frac{\delta F}{\delta \rho}$? They characterize optimality for F in the following sense.

<u>Proposition (First Variation and Optimality)</u>: Suppose that ρ_0 minimizes $F : \mathcal{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$ and that ρ_0 is regular for F. Suppose $g := \frac{\delta F}{\delta \rho_0}(\rho_0)$ exists, and let $\ell = \operatorname{ess\,inf} g$, the essential infimum of g w.r.t. Lebesgue measure.

(i) If $g \in C(\Omega)$, then $\mathrm{supp}(\rho_0) \subset \{g = \ell\}$ and $g \geq \ell$ with equality exactly on $\mathrm{supp}(\rho_0)$

(ii) Suppose g is measurable and ρ_0 is a.c. Then $g \geq \ell$ with $g = \ell$ a.e. on $\{\rho_0 > 0\}$.

<u>Remark</u>: Since g is only defined up to additive constants, this is basically analagous to saying "the derivative is 0 at optimizers," albeit the situation we are in is a bit more subtle.

<u>Proof</u>: Let $\tilde{\rho} \in L_c^\infty(\Omega)$, and set $\rho_\varepsilon := [1-\varepsilon]\rho_0 + \varepsilon\tilde{\rho}$

$$= \rho_0 + \varepsilon \chi, \quad \chi = \tilde{\rho} - \rho_0$$

Since ρ_0 is minimal, we have

$$\frac{d}{d\varepsilon}\left[F(\rho_\varepsilon)\right]\Big|_{\varepsilon=0} \geq 0$$

$$\Rightarrow \qquad \int g(x)\, d\chi(x) \geq 0$$

$$\Updownarrow$$

$$\int g(x)\, d\tilde{\rho}(x) \geq \int g(x)\, d\rho_0(x)$$

Since this holds for $\tilde{\rho}$, we will pick one carefully as follows. Let $\tilde{\ell} > \ell$. We know $\{x \mid g(x) < \ell'\}$ has positive Lebesgue measure, and we choose an L^∞ density $\tilde{\rho}$ concentrated on it.

It follows that

$\int g \, d\hat{p}$ $\le l' \cdot \int d\hat{p} = l'$, so that

$\int g \, dp_0 < l'$ in fact.

We can then take $l' \to l^-$ and get $\int g \, dp_0 \le l$. We now look at our cases:

(i): the essential infimum coincides with the infimum, so that $g \ge l$ everywhere;

then $\int g \, dp_0 \le l \Rightarrow g = l$ p_0-a.e., and extends to the support of p_0

by continuity of g.

(ii): $g \ge l$ a.e. $\Rightarrow$ $g \ge l$ p_0-a.e. $\Rightarrow g = l$ p_0-a.e. $\Rightarrow$ $g = l$ a.e. on

$\{x \mid p_0(x) > 0\}$. ∎