# (Rough) Introduction to functionals on the space of probabilities

—

May 25, 2020

**Abstract**

These notes are an abridged version of Chapter 7 of "Optimal Transport for Applied Mathematicians", by Filippo Santambrogio, for the Summer 2020 reading group on Topics in Optimal Transport. This is meant to quickly outline some of the fundamental definitions and notions that are necessary for reading papers in the general area.

## Classes of functionals and semi-continuity

Recall that a function $f : X \to \mathbb{R} \cup \{+\infty\}$ is called *lower semicontinuous*, henceforth abbreviated to lsc, if for all $(x_n)_n \subseteq X \to x$, $\liminf_n f(x_n) \geq f(x)$. There are several classes of functionals over the space of probabilities $\mathcal{P}(\Omega)$. We will state the general functional classes and briefly highlight when they are continuous/lsc.

**Potential energy** For a function $V$, the associated potential energy would be $\mathcal{V}(\mu) := \int_\Omega V \, d\mu$. We have that $\mathcal{V}(\mu)$ is continuous if and only if $V \in C_b(\Omega)$, or lsc if and only if $V$ is lsc and bounded from below.

**Interaction energy**: For a function defined on $\Omega \times \Omega$, we define the Interaction energy as $\mathcal{W}(\mu) := \int_{\Omega\times\Omega} W(x,y) \, d\mu(x)d\mu(y)$. The same conditions for continuity and lsc-ness apply as above.

**Wasserstein distances**: For a fixed measure $\nu$, we are often concerned with the mapping $\mu \mapsto W_p^p(\mu,\nu)$. When $\Omega$ is compact, then for any $\nu \in \mathcal{P}(\Omega)$, this functional is continuous. If it is not compact but $\nu \in \mathcal{P}_p(\Omega)$, then we have lsc and finite-valued. More generally, we can consider any cost function of the form $c(x,y)$ and study

$$\mathcal{T}_c(\mu,\nu) = \min_{\gamma\in\Pi(\mu,\nu)} \left\{ \int_{\Omega\times\Omega} c(x,y) \, d\gamma(x,y) \right\}.$$

In this case, compactness of $\Omega$ and continuity of $c$ gives continuity of $\mathcal{T}_c$. When $c$ is lsc and bounded from below, with no assumption on $\Omega$, we get lsc.

**Norm in a dual function space**: For a Banach space of functions on $\Omega$, say $\mathfrak{F}$, we define

$$\|\mu\|_{\mathfrak{F}^*} = \sup_{\varphi\in\mathfrak{F}, \|\varphi\|\leq 1} \int \varphi \, d\mu = \sup_{\varphi\in\mathfrak{F}\setminus\{0\}} \frac{\int \varphi \, d\mu}{\|\varphi\|}.$$

When $\mathfrak{F}$ is such that $\mathfrak{F} \cap C_b(\Omega)$ is dense in $\mathfrak{F}$, then we have lsc of the dual norm.

**Integral of a function of a density**: If $\mu$ has a density function with respect to the Lebesgue measure, $dx$, we would write

$$\mathcal{F}(\mu) := \begin{cases} \int_\Omega f(\rho(x))\ dx & \text{if } \mu = \rho \cdot dx, \\ +\infty & \text{otherwise.} \end{cases}$$

A very common choice of $f(y) = y \log(y)$, which gives the entropy of $\mu$ **(??)**.

**Sum of functions of atomic measures**: Finally, when $\mu$ is an atomic probability measure,

$$\mathcal{G}(\mu) := \begin{cases} \sum_i g(a_i) & \text{if } \mu = \sum_i a_i \delta_{x_i}, \\ +\infty & \text{otherwise.} \end{cases}$$

# Convexity, first variations, and subdifferentials

Denote $L_c^\infty(\Omega)$ as the set of absolutely continuous probabilities with $L^\infty$ densities and compact support. For a functional $F : \mathcal{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$, we say that $\rho \in \mathcal{P}(\Omega)$ is *regular for* $F$ if

$$F((1-\varepsilon)\rho + \varepsilon\tilde{\rho}) < +\infty$$

for every $\varepsilon \in [0,1]$ and every $\tilde{\rho} \in \mathcal{P}(\Omega) \cap L_c^\infty(\Omega)$. If $\rho$ is regular for $F$, we call $\frac{\delta F}{\delta \rho}(\rho)$, if it exists, any measurable function such that

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

where $\chi$ is a perturbation $\tilde{\rho} - \rho$ and $\tilde{\rho} \in L_c^\infty(\Omega) \cap \mathcal{P}(\Omega)$; this is called the *first variation of* $F$. For examples, we refer the interested reader to Section 7.2 of the textbook. We outline a couple examples:

- For a potential energy $\mathcal{V}(\mu)$, the first variation is simply $V$ (follows easily from linearity).
- For $\mu \mapsto \mathcal{T}_c(\mu, \nu)$, the first variation is given by $\varphi$, called a *Kantorovich potential*, which arises from the dual formulation.

## Optimality conditions

**Proposition 1.** *Suppose $\rho^*$ minimizes a functional $F : \mathcal{P}(\Omega) \to \mathbb{R} \cup \{+\infty\}$, is regular for $F$, the first variation at $\rho^*$ exists, called $g$, and $\ell$ to be the essential infimum of $g$,*

$$\ell = \textit{ess inf}\ g := \sup\{b \in \mathbb{R} \mid \textit{Leb}(\{x : g(x) < b\}) = 0\},$$

*when the inner set is non-empty, and $-\infty$ otherwise. In addition, suppose that $g \in C(\Omega)$. Then we have that $\textit{spt}(\rho^*) \subset \textit{argmin}\ g = \{g = \ell\}$, and we have $g \geq \ell$ everywhere.*

*Example* 2. Suppose $\Omega \subseteq \mathbb{R}^d$ is compact and $V$ is continuous. Consider the problem

$$\min_{\rho \in \mathcal{P}(\Omega)} F(\rho) := W_2^2(\rho, \nu) + \int V\ d\rho.$$

From the assumptions on $\Omega$ and $V$, we guarantee the existence of a minimizer and that it is regular for $F$. The first variation of this objective function is just $\varphi + V$ (up to an additive constant), where $\varphi$ is a Kantorovich potential.