Serie 11

Optimal transport, Fall semester

EPFL, Mathematics section, Dr. Xavier Fernández-Real

Exercise 11.1. Given a probability measure $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, let us denote by $x_\mu \in \mathbb{R}^d$ its barycenter (that is, $x_\mu := \int x d\mu(x)$). Show that, for any pair of probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and any $p \geq 1$, it holds

$$W_p(\mu, \nu) \geq |x_\mu - x_\nu|.$$

In particular, among all measures with fixed barycenter $\bar{x}$, which is one that minimizes the W_p-distance from a fixed delta $\delta_{\bar{y}}$?

Solution: Observe that, by Hölder's inequality, for any $p \geq 1$ we have, on the one hand

$$\int_{\mathbb{R}^{2d}} |x-y| d\gamma(x,y) \leq \left(\int_{\mathbb{R}^{2d}} |x-y|^p d\gamma(x,y) \right)^{\frac{1}{p}}.$$

On the other hand, we have

$$\int_{\mathbb{R}^{2d}} |x-y| d\gamma(x,y) \geq \left| \int_{\mathbb{R}^{2d}} (x-y) d\gamma(x,y) \right| = \left| \int_{\mathbb{R}^{2d}} x d\gamma(x,y) - \int_{\mathbb{R}^{2d}} y d\gamma(x,y) \right| = |x_\mu - x_\nu|,$$

where in the last step we used that $\int_{\mathbb{R}^{2d}} x d\gamma(x,y) = \int_{\mathbb{R}^d} x d\mu(x) = x_\mu$. Taking the infimum among transport plans γ we reach the desired result.

Finally, observe that when μ and ν are both deltas ($\mu = \delta_{x_\mu}$ and $\nu = \delta_{x_\nu}$) the equality holds for any $p \geq 1$, so among measures with fixed barycenter, one that minimizes the distance to a fixed delta is a delta.

Exercise 11.2. Given a probability measure $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ and a point $x_0 \in \mathbb{R}^d$, compute $W_p(\mu, \delta_{x_0})$.

Solution:

Claim: $\gamma = \mu \otimes \delta_{x_0}$ (indeed $\Gamma(\mu, \delta_{x_0}) = \{\mu \otimes \delta_{x_0}\}$). To show the claim we compute $\gamma(B)$, where $B = A_1 \times A_2$, $A_1 \subseteq \Omega$ and $A_2 \subseteq \Omega$. It suffices to consider sets of this form since the σ-algebra $\mathcal{A}$ on $\Omega \times \Omega$ is generated by these sets (indeed $\mathcal{A}$ is the smallest σ-algebra that contains these sets). We distinguish two cases:

- $x_0 \notin A_2$:

$$\mu \otimes \delta_{x_0}(B) = \int_B d(\mu \otimes \delta_{x_0}) = \int_{A_1} \int_{A_2} d\delta_{x_0} d\mu = \int_{A_1} 0 \, d\mu = 0$$

  We used Fubini in the second passage to separate the integral. For γ:

$$\gamma(B) = \gamma(A_1 \times A_2) \leq \gamma(\Omega \times A_2) = \delta_{x_0}(A_2) = 0$$

Since γ is a welldefined positive measure we can conclude $\gamma(B) = 0$, hence $\gamma = \mu \otimes \delta_{x_0}$.

- $x_0 \in A_2$:

$$\begin{aligned}\gamma(B) &= \gamma(A_1 \times A_2) \leq \gamma(A_1 \times \Omega) = \mu(A_1) = \int_{A_1} d\mu = \int_{A_1} 1\, d\mu \\ &= \int_{A_1}\int_{A_2} d\delta_{x_0} d\mu \overset{\text{Fubini}}{=} \int_B d(\mu \otimes \delta_{x_0}) = \mu \otimes \delta_{x_0}(B)\end{aligned}$$

Hence $\gamma(B) \leq \mu \otimes \delta_{x_0}(B)$. Consider now $\mu(\Omega \setminus A_1) = 1 - \mu(A_1)$. Following the same reasoning as above: $\gamma((\Omega \setminus A_1) \times A_2) \leq \mu(\Omega \setminus A_1) = \mu \otimes \delta_{x_0}((\Omega \setminus A_1) \times A_2)$. Therefore again by Fubini:

$$1 - \gamma(A_1 \times A_2) = \gamma((\Omega \setminus A_1) \times A_2) \leq \mu \otimes \delta_{x_0}((\Omega \setminus A_1) \times A_2) = 1 - \mu \otimes \delta_{x_0}(A_1 \times A_2)$$

Now we can conclude the inverse inequality $\gamma(B) \geq \mu \otimes \delta_{x_0}(B)$, which finishes the proof.

Now we can use the claim to compute $W_p(\mu, \delta_{x_0})$:

$$\begin{aligned}W_p(\mu, \delta_{x_0}) &= \left(\int_{\Omega^2} |x-y|^p d\gamma\right)^{\frac{1}{p}} = \left(\int_{\Omega^2} |x-y|^p d(\mu \otimes \delta_{x_0})\right)^{\frac{1}{p}} \\ &\overset{\text{Fubini}}{=} \left(\int_\Omega \int_\Omega |x-y|^p d\delta_{x_0}(y) d\mu(x)\right)^{\frac{1}{p}} \overset{x_0 \in \Omega}{=} \left(\int_\Omega |x-x_0|^p d\mu(x)\right)^{\frac{1}{p}} = \|x - x_0\|_p^\mu\end{aligned}$$

Where the norm of $x - x_0$ is defined since $\mu \in \mathcal{P}_p(\Omega)$.

Exercise 11.3. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a λ-Lipschitz function and let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$. Prove that

$$\int_{\mathbb{R}^d} f d\mu - \int_{\mathbb{R}^d} f d\nu \leq \lambda W_1(\mu, \nu).$$

Solution:

$$\begin{aligned}\lambda \cdot W_1(\mu, \nu) &\overset{\gamma \text{ optimal}}{=} \int_{\Omega \times \Omega} \lambda |x-y| d\gamma(x,y) \overset{\text{Lipschitz}}{\geq} \int_{\Omega \times \Omega} |f(x) - f(y)| d\gamma(x,y) \\ &\geq \int_{\Omega \times \Omega} (f(x) - f(y)) d\gamma(x,y) \overset{\gamma \in \Gamma(\mu,\nu)}{=} \int_\Omega f(x) d\mu(x) - \int_\Omega f(y) d\nu(y).\end{aligned}$$

Exercise 11.4. Let $1 \leq p < \infty$, and let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$. Consider a family of non-negative mollifiers $(\rho_\varepsilon)_{\varepsilon > 0} \subset C^\infty(\mathbb{R}^d)$ such that

$$\rho_\varepsilon(x) := \varepsilon^{-d} \rho\left(\frac{x}{\varepsilon}\right), \quad \int_{\mathbb{R}^d} \rho(x) dx = 1, \quad m_p^p(\rho) := \int_{\mathbb{R}^d} |x|^p \rho(x) dx < +\infty.$$

Then, if $\mu_\varepsilon := \mu * \rho_\varepsilon$ and $\nu_\varepsilon := \nu * \rho_\varepsilon$[1], show that

(i) $W_p(\mu, \mu_\varepsilon) \le \varepsilon m_p(\rho)$, and therefore, μ_ε converges to μ in $\mathcal{P}_p(\mathbb{R}^d)$ as $\varepsilon \downarrow 0$.

(ii) $W_p(\mu_\varepsilon, \nu_\varepsilon) \le W_p(\mu, \nu)$.

(iii) $W_p(\mu_\varepsilon, \nu_\varepsilon) \to W_p(\mu, \nu)$ as $\varepsilon \downarrow 0$.

Solution: Let us show first (i). Let us consider the coupling $\gamma_\varepsilon \in \Gamma(\mu, \mu_\varepsilon)$ defined by

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y) d\gamma_\varepsilon(x,y) := \int_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y)\rho_\varepsilon(y-x)\,dy\,d\mu(x).$$

Notice that indeed $\gamma_\varepsilon \in \Gamma(\mu, \mu_\varepsilon)$ since

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \phi(x) d\gamma_\varepsilon(x,y) = \int_{\mathbb{R}^d} \phi(x) d\mu(x),$$

and

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \phi(y) d\gamma_\varepsilon(x,y) = \int_{\mathbb{R}^d} (\phi * \rho_\varepsilon)(x) d\mu(x) = \int_{\mathbb{R}^d} \phi(x) d(\rho_\varepsilon * \mu)(x).$$

In particular, we have that

$$\begin{aligned} W_p^p(\mu,\mu_\varepsilon) &\le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\gamma_\varepsilon(x,y) = \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} |x-y|^p \rho_\varepsilon(y-x)\,dy\right) d\mu(x) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |z|^p\rho_\varepsilon(z)\,dz\,d\mu(x) = \varepsilon^p \int_{\mathbb{R}^d} |z|^p \rho(z)\,dz \end{aligned}$$

as we wanted to see.

For the second point, (ii), given any coupling $\gamma \in \Gamma(\mu, \nu)$, let us define γ_ε as

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \varphi(x,y) d\gamma_\varepsilon(x,y) := \int_{\mathbb{R}^d\times\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(x-z, y-z)\rho_\varepsilon(z)\,dz\,d\gamma(x,y).$$

In particular, as before we can check that $\gamma_\varepsilon \in \Gamma(\mu_\varepsilon, \nu_\varepsilon)$:

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \phi(x) d\gamma_\varepsilon(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d} (\phi*\rho_\varepsilon)(x) d\gamma(x,y) = \int_{\mathbb{R}^d} (\phi*\rho_\varepsilon)(x) d\mu(x) = \int_{\mathbb{R}^d} \phi(x) d(\rho_\varepsilon * \mu)(x),$$

and the same holds for ν. Thus,

$$\begin{aligned} W_p^p(\mu_\varepsilon,\nu_\varepsilon) &\le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\gamma_\varepsilon(x,y) = \int_{\mathbb{R}^d\times\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|^p \rho_\varepsilon(z)\,dz\,d\gamma(x,y) \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\gamma(x,y). \end{aligned}$$

Taking now the infimum in $\gamma \in \Gamma(\mu, \nu)$ we reach the desired result.

[1] The convolution of a measure μ with a smooth function ρ is the measure defined as follows:

$$\mu * \rho(A) := \int_A \int_{\mathbb{R}^d} \rho(x-y) d\mu(y) dx.$$

Finally, for point (iii), we combine the previous two points with the triangular inequality. Indeed,

$$W_p(\mu, \nu) \leq W_p(\mu, \mu_\varepsilon) + W_p(\mu_\varepsilon, \nu_\varepsilon) + W_p(\nu_\varepsilon, \nu) \leq 2\varepsilon m_p(\rho) + W_p(\mu_\varepsilon, \nu_\varepsilon).$$

That is,

$$W_p(\mu, \nu) - 2\varepsilon m_p(\rho) \leq W_p(\mu_\varepsilon, \nu_\varepsilon) \leq W_p(\mu, \nu)$$

and the desired result follows.