Serie 7

Optimal transport, Fall semester

EPFL, Mathematics section, Dr. Xavier Fernández-Real

Exercise 7.1 (Counterexamples)**.** For any of the following statements, find two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with compact support such that the statement holds (you can choose also the dimension $d \in \mathbb{N}$). Each of the statements should be treated independently.

(i) There is more than one[1] optimal transport map from μ to ν with respect to the linear cost $|x-y|$.

(ii) There is more than one optimal transport map from μ to ν with respect to the quadratic cost $\frac{1}{2}|x-y|^2$.

(iii) There is not an optimal transport plan between μ and ν with respect to the cost $c(x,y) = \lfloor |x-y| \rfloor$ (the floor function[2] of the distance).

(iv) There is an optimal transport map from μ to ν with respect to the linear cost, but there is none with respect to the quadratic cost.

Hint: To solve (iii), show that the infimum of the Kantorovich problem for $\mu = \chi_{[0,1]}\mathcal{L}^1$, $\nu = \chi_{[1,2]}\mathcal{L}^1$ is 0 but any transport plan has *strictly* positive cost.

Solution:

(i) Let $d = 1$ and $\mu := \frac{1}{2}(\delta_0 + \delta_1)$, $\nu := \frac{1}{2}(\delta_1 + \delta_2)$. Every transport map from μ to ν has the same cost (cf. Remark 2.7.5), thus the two maps $T_1, T_2 : \{0,1\} \to \{1,2\}$ given by

$$T_1(0) = 1, \quad T_1(1) = 2, \qquad T_2(0) = 2, \quad T_2(1) = 1,$$

both send μ to ν and are both optimal.

(ii) Let $d = 2$ and let $p_1, p_2, p_3 \subset \mathbb{R}^2$ be the three vertices of an equilateral triangle. Let $\mu := \frac{1}{2}(\delta_{p_1} + \delta_{p_2})$ and $\nu := \frac{1}{2}(\delta_{p_3} + \delta_{\frac{p_1+p_2+p_3}{3}})$. One can explicitly check that every transport plan from μ to ν has the same quadratic cost. Thus the two maps from μ to ν (namely, $T_1(p_1) = p_3$, $T_1(p_2) = \frac{p_1+p_2+p_3}{3}$, and $T_2(p_1) = \frac{p_1+p_2+p_3}{3}$, $T_2(p_2) = p_3$) are both optimal.

(iii) Let $d = 1$ and $\mu := dx|_{[0,1]}$, $\nu := dy|_{[1,2]}$. We will show that the infimum of the Kantorovich problem is 0, but any transport plan has *strictly* positive cost.

Given $\varepsilon > 0$, consider the map $T_\varepsilon : [0,1] \to [1,2]$ defined as

$$T_\varepsilon(x) := \begin{cases} x + 2 - \varepsilon & \text{if } 0 \le x \le \varepsilon, \\ x + 1 - \varepsilon & \text{if } \varepsilon < x \le 1. \end{cases}$$

[1] Uniqueness should be understood in the μ-almost everywhere sense.
[2] Given $t \ge 0$, $\lfloor t \rfloor$ is the largest integer n such that $n \le t$.

One can check that $(T_\varepsilon)_\#\mu = \nu$. Also, it holds

$$\int_{[0,1]} c(x, T_\varepsilon(x))\, d\mu(x) = \int_0^\varepsilon \lfloor 2-\varepsilon \rfloor \, dx + \int_\varepsilon^1 \lfloor 1-\varepsilon \rfloor \, dx = \varepsilon.$$

Since ε can be chosen arbitrarily small, we have proven that the infimum of the Kantorovich problem is 0.

Let us assume by contradiction that $\gamma \in \Gamma(\mu,\nu)$ achieves cost 0. Hence γ must have all his mass on the set

$$\{(x,y) \in \mathbb{R}^2 : \ c(x,y) = 0\} = \{(x,y) \in \mathbb{R}^2 : \ |x-y| < 1\}. \tag{1}$$

However, since $\gamma \in \Gamma(\mu,\nu)$, by the marginal condition we get

$$1 = \int_{\mathbb{R}^2} (y-x) d\gamma(x,y) = \int_{\mathbb{R}^2} |y-x| d\gamma(x,y) < 1,$$

a contradiction.

(iv) Let $d = 1$ and $\mu := dx|_{[0,\frac{1}{2}]} + \frac{1}{2}\delta_1$, $\nu := dy|_{[\frac{5}{2},3]} + \frac{1}{2}\delta_2$. Since μ is supported on $\{x \leq 2\}$, whereas ν is supported on $\{x \geq 2\}$, Remark 2.7.5 implies that any admissible plan has the same cost with respect to the linear cost, and this cost is given by

$$\int_{\mathbb{R}} x\, d\nu - \int_{\mathbb{R}} x\, d\mu = \frac{7}{4}.$$

In particular, any admissible map is optimal with respect to the linear cost (and it is clear that there is at least one, take for instance $T(x) = x + 5/2$ for $x \in [0, 1/2)$ and $T(1/2) = 2$).

On the other hand, the optimal map with respect to the quadratic cost (if it exists) must be nondecreasing. Let us assume by contradiction that there is an admissible transport map $T : \mathbb{R} \to \mathbb{R}$ that is nondecreasing. Since it is an admissible map, it must hold $T(1) = 2$. The monotonicity then implies that $T(x) \leq 2$ for any $x \leq 1$ and thus $\nu = T_\#\mu$ is supported on $x \leq 2$, that is a contradiction.

Exercise 7.2. Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be two compactly supported probability measures invariant under rotations (that is, $\mu(L(E)) = \mu(E)$ and $\nu(L(E)) = \nu(E)$ for any Borel set E and any orthogonal transformation $L \in O(d)$). Assume that $\mu \ll \mathcal{L}^d$, and let T be the unique optimal transport map from μ to ν with respect to the quadratic cost (see Theorem 2.5.9). Show that T can be written as $x \to \tau(|x|)\frac{x}{|x|}$, where $\tau : [0, +\infty) \to [0, +\infty)$ is a nondecreasing function.

Hint: The function τ is the monotone transport map between two suitable 1-dimensional measures. Also, one may want to use (and prove) the following lemma:

Lemma 1. Let $\mu^0, \mu^1 \in \mathcal{P}(\mathbb{R}^d)$ be two rotationally invariant probability measures, and let $\Phi(x) := |x|$. If $\Phi_\#\mu^0 = \Phi_\#\mu^1$ then $\mu^0 = \mu^1$.

Solution:

Sketch of the proof of Lemma 1. We need to show that, given any smooth function $\varphi \in C_c^\infty(\mathbb{R}^d)$,

$$\int \varphi \mu^0 = \int \varphi \mu^1.$$

Since μ_0 is rotationally symmetric, we know that $\int_{\mathbb{R}^d} \varphi(Lx) d\mu^0(x) = \int_{\mathbb{R}^d} \varphi(x) d\mu^0(x)$ for any othonormal transformation $L \in O(d)$. In particular, we can take all symmetries $L \in O(d)$ and we have

$$\int_{\mathbb{R}^d} \varphi(x) d\mu^0(x) = \frac{1}{|O(d)|} \int_{O(d)} \int_{\mathbb{R}^d} \varphi(Lx) d\mu^0(x) dL = \int_{\mathbb{R}^d} \bar{\varphi}(x) d\mu^0(x)$$

where

$$\bar{\varphi}(x) = \frac{1}{|O(d)|} \int_{O(d)} \varphi(Lx) dL$$

is rotationally invariant. That is, $\bar{\varphi}(x) = \phi(|x|)$ for some $\phi : [0, \infty) \to \mathbb{R}$, and since $\Phi_\# \mu^0 = \Phi_\# \mu^1$ we have

$$\int \varphi \mu^0 = \int_{\mathbb{R}^d} \phi(|x|) d\mu^0(x) = \int_{\mathbb{R}} \phi(x) d\Phi_\# \mu^0(x) = \int_{\mathbb{R}} \phi(x) d\Phi_\# \mu^1(x) = \int_{\mathbb{R}^d} \phi(|x|) d\mu^0(x) = \int \varphi \mu^1$$

as we wanted to see. $\square$

Let us now solve the exercise. Let $\Phi(x) := |x|$, and define the one-dimensional compactly supported measures

$$\tilde{\mu} := \Phi_\# \mu \in \mathcal{P}([0, +\infty)), \qquad \tilde{\nu} := \Phi_\# \nu \in \mathcal{P}([0, +\infty)).$$

From the identity $\Phi_\# dx = \omega_d r^{d-1} dr|_{[0,+\infty)}$ (use polar coordinates), where ω_d is the measure of the unit sphere in $\mathbb{R}^d$, and the fact that $\mu \ll \mathcal{L}^d$, it follows that $\tilde{\mu} \ll \Phi_\# \mathcal{L}^d = \omega_d r^{d-1} dr|_{[0,+\infty)} \ll dr$. Hence, applying Theorem 2.5.9 from $\tilde{\mu}$ to $\tilde{\nu}$, there exists a convex function $\psi : [0, +\infty) \to \mathbb{R}$ such that $\tau := \psi'$ is the optimal transport map from $\tilde{\mu}$ to $\tilde{\nu}$. Notice that ψ is nondecreasing ($\psi' \geq 0$).

Let us show that $\mathcal{T}(x) := \tau(|x|) \frac{x}{|x|}$ is the optimal transport map from μ to ν, so $\mathcal{T} = T$. We begin by noticing that $\mathcal{T}$ is the gradient of the function $\Psi(x) := \psi(|x|)$, $\mathcal{T} = \nabla \Psi$, and that Ψ is convex (since ψ is convex and non-decreasing). By Corollary 2.5.10, $\mathcal{T}$ is an optimal transport map. It remains only to prove $\mathcal{T}_\# \mu = \nu$.

First, we prove that $\mathcal{T}_\# \mu$ is rotationally invariant and $\Phi_\# \mathcal{T}_\# \mu = \Phi_\# \nu$. Given $L \in O(d)$,

$$L \circ \mathcal{T}(x) = L\Big(\tau(|x|) \frac{x}{|x|}\Big) = \tau(|x|) \frac{Lx}{|x|} = \tau(|Lx|) \frac{Lx}{|Lx|} = \mathcal{T} \circ Lx.$$

Since μ is rotationally invariant, we get

$$L_\# T_\# \mu = (L \circ \mathcal{T})_\# \mu = (\mathcal{T} \circ L)_\# \mu = \mathcal{T}_\# L_\# \mu = \mathcal{T}_\# \mu,$$

and thus $\mathcal{T}_\# \mu$ is rotationally invariant. Also, from the identity $\Phi \circ \mathcal{T} = \tau \circ \Phi$, we deduce that

$$\Phi_\# T_\# \mu = (\Phi \circ T)_\# \mu = (\tau \circ \Phi)_\# \mu = \tau_\# \tilde{\mu} = \tilde{\nu} = \Phi_\# \nu.$$

Hence, applying the Lemma stated in the hint we conclude that $\mathcal{T}_{\#}\mu = \nu$, as desired.

Exercise 7.3. Find the optimal transport map for the quadratic cost $c(x,y) = \frac{1}{2}|x-y|^2$ between $\mu = f \cdot \mathcal{L}^2$ and $\nu = g \cdot \mathcal{L}^2$ in $\mathbb{R}^2$, where $f(x) = \frac{1}{\pi}\mathbb{1}_{B_1}(x)$ and $g(x) = \frac{1}{8\pi}(4-|x|^2)\mathbb{1}_{B_2}(x)$.

Solution: Let us use Exercise 7.2. We know that the optimal transport map will be of the form $T(x) = \tau(|x|)\frac{x}{|x|}$ for some τ nondecreasing. We need to choose τ such that $T_{\#}\mu = \nu$. It is enough to check this condition on balls. Indeed, we need to find τ such that for any $r \in (0,1)$,

$$\mu(B_r) = \nu(T(B_r)) = \nu(B_{\tau(r)}).$$

We have

$$\mu(B_r) = \frac{1}{\pi}\int_{B_r} dx = r^2$$

and, knowing that $\tau(r) < 2$,

$$\nu(B_{\tau(r)}) = \frac{1}{8\pi}\int_{B_{\tau(r)}}(4-|x|^2)dx = \frac{1}{4}\int_0^{\tau(r)}(4-t^2)tdt = \frac{1}{2}(\tau(r))^2 - \frac{1}{16}(\tau(r))^4.$$

That is, for all $r \in (0,1)$ solving $\frac{1}{2}(\tau(r))^2 - \frac{1}{16}(\tau(r))^4 = r^2$ for $\tau(r)$ positive we obtain $\tau(r) = 2\sqrt{1+\sqrt{1-r^2}}$. That is,

$$T(x) = 2\sqrt{1+\sqrt{1-|x|^2}}\,\frac{x}{|x|}\mathbb{1}_{B(0,1)}$$

is the optimal transport map we are looking for.

Remark 7.1. We gave in class a counterexample for the following statement: let us fix a cost c lower semicontinuous and $\gamma \in \Gamma(\mu,\nu)$ such that suppγ is c-cyclically monotone, then γ is optimal.

The example is the following: let $\mu = \mathcal{L}$ be the lebesgue measure on the one dimensional torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, $\alpha \in \mathbb{R}\setminus\mathbb{Q}$, $c : \mathbb{T}\times\mathbb{T} \to \mathbb{R}$ be defined as follows

$$c(x,y) = \begin{cases} 1 & \text{if and only if } y = x-\alpha, \\ 0 & \text{if and only if } y = x, \\ \infty & \text{otherwise ,} \end{cases}$$

$\gamma = (Id, T)_{\#}\mu$ and $T(x) = x - \alpha$. We proved in class that suppγ is c-cyclically monotone but γ is not optimal.

Exercise 7.4. (💣) In this exercise we want to go deeper in the understanding of the counterexample in Remark 7.1.

(i) Find as many points as you can in the proof of the statement "suppγ is c-cyclically monotone, then γ is optimal for a continuous cost" where it could fail.

(ii) If we define for any $n \in \mathbb{N}$ the cost

$$c_n(x,y) = \begin{cases} 1 & \text{if and only if } y = x - \alpha, \\ 0 & \text{if and only if } y = x, \\ n & \text{otherwise.} \end{cases}$$

Is it true that γ is optimal for the cost c_n (where γ was defined above $\gamma = (Id, T)_{\#}\mu$)? Is it true that suppγ is c_n-cyclically monotone for the cost c_n?

Solution:

(i) One point is when we defined

$$\varphi(x) = \sup_{N \geq 1, (x_i, y_i) \in S} \{-c(x, y_N) + c(x_N, y_N) - c(x_N, y_{N-1}) + c(x_{N-1}, y_{N-1}) + \cdots - c(x_1, y_0)\}$$

we should verify that $\varphi(x) \not\equiv -\infty$ (which is direct if we use the quadratic cost, because φ is a convex function which is finite in a point) otherwise the definition of

$$\varphi^c(y) = \sup_z \{-c(y,z) - \varphi(z)\} \equiv +\infty$$

(or is not well defined, what is $-\infty + \infty = ?$) which compromises the proof of

$$c(x,y) + \varphi(x) + \varphi^c(y) = 0 \quad \Leftarrow \quad y \in \partial_c \varphi(x)$$

used in the proof of "suppγ c-cyclically monotone implies γ optimal". Indeed if x is such that $\varphi(x) = -\infty$ then $\partial\varphi(x) = \mathbb{R}^d$, but $c(x,y) + \varphi(x) + \varphi^c(y)$ is not well defined and in particular cannot be equal to 0.

Remark 7.2. As an extra exercise you can verify that φ defined above for the cost function c and the c-cyclically monotone set $S =$ suppγ defined in Remark 7.1 we do not have the following property: $\varphi(x) \neq -\infty$ for any point $x \in \mathbb{T}$. This is one point where the proof for a continuous cost function fails in the particular case of the example in Remark 7.1.

(ii) For any $n \in \mathbb{N}$, γ is not optimal because (KP)=0, but $\int_{\mathbb{T}^2} c(x,y) d\gamma(x,y) = 1$.

For any $n \in \mathbb{N}$ the support of γ is not c_n-cyclically monotone.

Construct the same sequence as in class, but now the closing condition cost n instead of $+\infty$, more precisely, let us fix $n \in \mathbb{N}$ and we prove that suppγ is not c_n-cyclically monotone: let $x_1 \in \mathbb{T}$, for any $i \in \mathbb{N}$ we iteratively define

$$x_{i+1} = y_i = x_i - \alpha.$$

It is clear that the sequence $(x_i, y_i) \in$ supp(γ) for any i and if we take the sequence $\{x_i\}_{i=1,..,n+1}$ we have that

$$n+1 = \sum_{i=1}^{n+1} c(x_i, y_i) > \sum_{i=1}^{n+1} c(x_{i+1}, y_i) = c(x_{n+2}, y_{n+1}) = c(x_1, y_{n+1}) = c(x_1, x_1 - (n+1)\alpha) = n$$

where we imposed the closing condition $x_{n+2} = x_1$.