

Serie 6

Optimal transport, Fall semester

EPFL, Mathematics section, Dr. Xavier Fernández-Real

Exercise 6.1 (Translations are optimal). Let $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the translation map $T(x) := x + x_0$, where $x_0 \in \mathbb{R}^d$. For any probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, show that T is an optimal transport map from μ to $T_{\#}\mu$ with respect to the quadratic cost.

Solution: We provide two solutions.

- (i) We just need to show that T is the gradient of a convex function (see Corollary 2.5.10). Let $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ be the convex function $\varphi(x) := \frac{1}{2}|x + x_0|^2$. Since $T = \nabla\varphi$, the optimality of T follows.
- (ii) By the general case of Kantorovich duality (Theorem 2.6.6) we know that for any ϕ and ψ continuous and bounded such that $\phi(x) + \psi(y) \leq c(x, y)$, we have

$$\int_X \phi d\mu + \int_Y \psi d\nu \leq \min_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y)$$

where X and Y need not be compact and $\nu = T_{\#}\mu$.

Consider $\phi(x) = -\langle x, x_0 \rangle$ and $\psi(y) = \langle y - x_0, x_0 \rangle + \frac{1}{2}\|x_0\|^2$. Then

$$\begin{aligned} \phi(x) + \psi(y) &= -\langle x, x_0 \rangle + \langle y - x_0, x_0 \rangle + \frac{1}{2}\|x_0\|^2 = \langle y - x, x_0 \rangle - \frac{1}{2}\|x_0\|^2 \\ &\leq \|y - x\|\|x_0\| - \frac{1}{2}\|x_0\|^2 \leq \frac{1}{2}\|y - x\|^2 = c(x, y). \end{aligned}$$

Hence, recalling $\nu = T_{\#}\mu$ with $T(x) = x + x_0$

$$\begin{aligned} \min_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) &\geq \int_{\mathbb{R}^n} -\langle x, x_0 \rangle d\mu(x) + \int_{\mathbb{R}^n} \left(\langle y - x_0, x_0 \rangle + \frac{1}{2}\|x_0\|^2 \right) d\nu(y) \\ &= - \int_{\mathbb{R}^n} \langle x, x_0 \rangle d\mu(x) + \int_{\mathbb{R}^n} \left(\langle x, x_0 \rangle + \frac{1}{2}\|x_0\|^2 \right) d\mu(x) \\ &= \int_{\mathbb{R}^n} \frac{1}{2}\|x_0\|^2 d\mu(x) = \int_{\mathbb{R}^n} c(x, T(x)) d\mu(x) \end{aligned}$$

which means that T is optimal.

Exercise 6.2 (Homoteties are optimal). Let $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the homotety map $T(x) := \lambda x$, where $\lambda > 0$. For any compactly supported probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$, show that T is an optimal transport map from μ to $T_{\#}\mu$ with respect to the quadratic cost.

Solution: We provide two solutions.

- (i) It is sufficient to show that the homothety T is the gradient of a convex function (see Corollary 2.5.10). Let $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ be the convex function $\varphi(x) := \frac{\lambda}{2}|x|^2$. Since $T = \nabla\varphi$, the optimality of T follows.
- (ii) We consider the potential $\phi(x) = \frac{1}{2}(1-\lambda)\|x\|^2$ together with its c-transform ϕ^c . Since μ is compactly supported, we can work on a compact domain $\Omega = \max(1, \lambda)\text{spt}(\mu) \subset \mathbb{R}^n$, so that ϕ is bounded in Ω . We compute

$$\phi^c(y) = \inf_{x \in \mathbb{R}^n} c(x, y) - \phi(x) = \inf_{x \in \mathbb{R}^n} \frac{1}{2}\|x - y\|^2 - \frac{1}{2}(1-\lambda)\|x\|^2 \geq \inf_{x \in \mathbb{R}^n} \frac{1}{2}(\|x\| - \|y\|)^2 - \frac{1}{2}(1-\lambda)\|x\|^2$$

The previous is a quadratic expression and the infimum is achieved when $\|x\| = \frac{1}{\lambda}\|y\|$. In particular, we find

$$\phi^c(y) = -\frac{(1-\lambda)}{2\lambda}\|y\|^2.$$

The definition of c-transform implies $\phi(x) + \phi^c(y) \leq c(x, y)$. Thus, if we denote $\nu = T_{\#}\mu$ with $T(x) = \lambda x$,

$$\begin{aligned} \min_{\gamma \in \Gamma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) &\geq \int_{\Omega} \phi(x) d\mu(x) + \int_{\Omega} \phi^c(x) d\nu(x) \\ &= \int_{\Omega} \frac{1}{2}(1-\lambda)\|x\|^2 d\mu(x) - \int_{\Omega} \frac{\lambda(1-\lambda)}{2}\|x\|^2 d\mu(x) \\ &= \int_{\Omega} \frac{(1-\lambda)^2}{2}\|x\|^2 d\mu(x) = \int_{\mathbb{R}^n} c(x, T(x)) d\mu(x), \end{aligned}$$

which means that T is optimal.

Exercise 6.3. Let $S : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the function $S(x) := -x$. Characterize the probability measures $\mu \in \mathcal{P}(\mathbb{R}^d)$ with compact support such that S is an optimal transport map between μ and $S_{\#}\mu$ with respect to the quadratic cost.

Solution: Assume that S is optimal from μ to $S_{\#}\mu$, and let $\gamma := (\text{id} \times S)_{\#}\mu$ be the associated coupling. Since γ is optimal, Corollary 2.5.7 implies the existence of a cyclically monotone set $A \subset \mathbb{R}^d \times \mathbb{R}^d$ such that γ is supported on A . Since γ is also supported on $\text{graph}(S)$, we can assume without loss of generality that $A \subset \text{graph}(S)$.

Take two points $x, y \in \mathbb{R}^d$ such that $(x, S(x)), (y, S(y)) \in A$. By cyclical monotonicity, it holds

$$\frac{1}{2}|x - S(x)|^2 + \frac{1}{2}|y - S(y)|^2 \leq \frac{1}{2}|x - S(y)|^2 + \frac{1}{2}|y - S(x)|^2.$$

Developing the squares and rearranging terms, this is equivalent $|x - y|^2 \leq 0$, thus $x = y$. Hence, this implies that A contains only one point $(x_0, S(x_0))$, and therefore $\mu = \delta_{x_0}$.

On the other hand, if μ is δ_{x_0} for some $x_0 \in \mathbb{R}^d$, then S is optimal from μ to $S_{\#}\mu = \delta_{S(x_0)}$ (since there is only one transport map).

Exercise 6.4. Let B_1 be the unit ball in \mathbb{R}^2 . Write it as $B_1^+ \cup B_1^-$, where $B_1^+ = B_1 \cap \{(x, y) : x \geq 0\}$ and $B_1^- = B_1 \cap \{(x, y) : x < 0\}$. Find a convex function $\varphi : B_1 \rightarrow \mathbb{R}$ such that $\nabla\varphi$ is the optimal transport between $\mathbb{1}_{B_1}\mathcal{L}^2$ and $(\mathbb{1}_{(-1,0)+B_1^-} + \mathbb{1}_{(1,0)+B_1^+})\mathcal{L}^2$ with the quadratic cost.

- *Observation:* Recall that $\mathbb{1}_A$ denotes the indicator function of a set $A \subset \mathbb{R}^2$.
- (♣): For every small $\epsilon > 0$, define the strip

$$S_\epsilon := [-1, 1] \times [-\epsilon, \epsilon] \subset \mathbb{R}^2.$$

Let $\mu = \mathbb{1}_{B_1}\mathcal{L}^2$, and consider the measures

$$\nu_\epsilon := \frac{1}{1+4\epsilon}(\mathbb{1}_{(-1,0)+B_1^-} + \mathbb{1}_{(1,0)+B_1^+})\mathcal{L}^2 + \frac{1}{1+4\epsilon}\mathbb{1}_{S_\epsilon}\mathcal{L}^2.$$

Let $T_\epsilon : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the optimal transport map from μ to ν_ϵ with respect to the quadratic cost. Prove that for ϵ sufficiently small, T_ϵ is discontinuous in the interior of B_1 .

Solution:

We may conjecture that the optimal transport moves each point horizontally by +1 or -1 depending on which side of B_1 they are. Namely, we claim that the optimal transport map $T = \nabla\varphi$ is given by $\nabla\varphi(x, y) = (\text{sign } x + x, y)$. Observe that this map is transporting the right measures, since

$$\begin{aligned} (\nabla\varphi(x, y))_{\#}\mathbb{1}_{B_1}\mathcal{L}^2 &= ((x, y) \rightarrow (1+x, y))_{\#}\mathbb{1}_{B_1^+}\mathcal{L}^2 + ((x, y) \rightarrow (-1+x, y))_{\#}\mathbb{1}_{B_1^-}\mathcal{L}^2 \\ &= (\mathbb{1}_{(-1,0)+B_1^-} + \mathbb{1}_{(1,0)+B_1^+})\mathcal{L}^2. \end{aligned}$$

From here, we observe that the corresponding φ then could be given by (integrating in x and y the expression of $\nabla\varphi$)

$$\varphi(x, y) = |x| + \frac{1}{2}(x^2 + y^2)$$

which is a convex function.

By Theorem 2.5.9 and Corollary 2.5.10, every gradient of a convex function is the unique optimal map between its own marginals in the setting of this problem, so φ is the unique optimal transport map that we are looking for.

For the proof of the second point see the discussion at pages 58-59 in Ambrosio's book.

Exercise 6.5. Let $S : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be the reflection with respect to the last variable, i.e.

$$S(x_1, \dots, x_{d-1}, x_d) = (x_1, \dots, x_{d-1}, -x_d).$$

Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ be a probability measure with finite quadratic moment. Prove that if S is an optimal transport map between μ and $S_{\#}\mu$, then there exists a function $f : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$ with Lipschitz constant $\text{Lip}(f) \leq 1$ such that

$$\text{supp}\mu \subseteq \text{Graph}(f) := \{x \in \mathbb{R}^d : x_d = f(x_1, \dots, x_{d-1})\}.$$

Solution: If S is optimal, then the support of the coupling $(id, S)_{\#}\mu$ has to be cyclically monotone. In particular, for every $x, y \in \text{supp}\mu$, calling $x = (x', x_d), y = (y', y_d)$ we must have

$$0 \leq \langle S(y) - S(x), y - x \rangle = |y' - x'|^2 - (y_d - x_d)^2.$$

Thus, $|y_d - x_d| \leq |y' - x'|$ for all $x, y \in \text{supp}\mu$. Let $K \subset \mathbb{R}^{d-1}$ be the set given by the projection on the first $d - 1$ variables of the support of μ . Notice that, by the inequality we proved above, for each $x' \in K$ there is exactly one value $x_d =: \bar{f}(x')$ for which $(x', x_d) \in \text{supp}\mu$. Moreover, the function $\bar{f} : K \rightarrow \mathbb{R}$ so defined is 1-Lipschitz and $\text{supp}\mu \subset \text{Graph}(\bar{f})$. By Kirszbraun's extension theorem (see https://en.wikipedia.org/wiki/Kirszbraun_theorem) \bar{f} can be extended to a 1-Lipschitz map $f : \mathbb{R}^{d-1} \rightarrow \mathbb{R}$, which concludes the proof.

Exercise 6.6. Let (X, d) be a compact metric space. Consider the cost function $c : X \times X \rightarrow [0, \infty)$ defined as $c(x, y) = d(x, y)^\alpha$, for some $\alpha \in (0, 1)$. Prove that, for every $\mu, \nu \in \mathcal{P}(X)$,

$$\min_{\gamma \in \Gamma(\mu, \nu)} \left\{ \int_{X \times X} c(x, y) d\gamma(x, y) \right\} = \max \left\{ \int_X \varphi d(\mu - \nu) : \varphi \in C^\alpha(X), [\varphi]_\alpha \leq 1 \right\},$$

where $C^\alpha(X)$ is the space of α -Hölder continuous functions, i.e., functions $\varphi : X \rightarrow \mathbb{R}$ such that

$$[\varphi]_\alpha := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d(x, y)^\alpha} < \infty.$$

Solution: By Kantorovich duality, it is sufficient to prove that $\varphi : X \rightarrow \mathbb{R}$ is c -convex if and only if $[\varphi]_\alpha \leq 1$, and that if this is the case, then the c -conjugate of φ is $-\varphi$.

Let φ be c -convex. Then, by definition,

$$\varphi(x) = \sup_{z \in X} (-d(x, z)^\alpha + \psi(z))$$

for some function $\psi : X \rightarrow \mathbb{R}$. Fix $x, y \in X$, $x \neq y$ and $\varepsilon > 0$. We can find $z \in X$ such that $\varphi(x) \leq -d(x, z)^\alpha + \psi(z) + \varepsilon$. Having also $\varphi(y) \geq -d(y, z)^\alpha + \psi(z)$, we deduce

$$\varphi(x) - \varphi(y) \leq d(y, z)^\alpha - d(x, z)^\alpha + \varepsilon \leq d(x, y)^\alpha + \varepsilon,$$

where in the second step we used triangle inequality and the concavity of $r \mapsto r^\alpha$. By the arbitrariness of $\varepsilon > 0$, x , and y we obtain $[\varphi]_\alpha \leq 1$.

Conversely, if $[\varphi]_\alpha \leq 1$, then for all $x, y \in X$ we have $\varphi(x) \geq -d(x, y)^\alpha + \varphi(y)$, with equality if $x = y$. This implies that

$$\varphi(x) = \sup_{y \in X} (-d(x, y)^\alpha + \varphi(y)).$$

Therefore, φ is c -convex, and $\varphi^c = -\varphi$, as desired.