Serie 8

Optimal transport, Fall semester

EPFL, Mathematics section, Dr. Xavier Fernández-Real

Exercise 8.1. Assume $\mu, \nu \in \mathcal{P}(\mathbb{R})$ and μ be non atomic. We define

$$T_{mon}(x) := F_\nu^{[-1]} \circ F_\mu(x),$$

where F denotes the cumulative distribution of a probability measure and $F^{[-1]}$ its pseudo-inverse. Show that the nondecreasing map $T_{mon} : \mathbb{R} \to \mathbb{R}$ constructed in class is optimal with respect to the quadratic cost.

Solution: We first show that any monotone map $T : \mathbb{R} \to \mathbb{R}$ pushing μ to ν which is

$$T_\# \mu(A) = \nu(A) \ \forall A \ \nu \text{ measurable}$$

must be equal to T_{mon}. Note that we have the following inclusion for any $x \in \mathbb{R}$

$$(-\infty, x] \subset T^{-1}((-\infty, T(x))])$$

since x is clearly in the preimage and hence any smaller number by monotonicity of T. This gives:

$$\begin{aligned} F_\mu(x) = \mu((-\infty, x]) &\leq \mu(T^{-1}((-\infty, T(x))])) = T_\# \mu((-\infty, T(x))]) \\ &= \nu((-\infty, T(x))]) = F_\nu(T(x)). \end{aligned}$$

Hence $F_\nu^{[-1]} \circ F_\mu(x) \leq T(x)$. Assume now by contradiction that the inequality is strict at some fixed x. This gives, by definition of the pseudo inverse, the existence of some $\epsilon_0 > 0$ such that for all $\epsilon \in (0, \epsilon_0)$,

$$F_\mu(x) \leq F_\nu(T(x) - \epsilon).$$

However $T^{-1}((-\infty, T(x))]) \subset (-\infty, x]$ by monotonicity again, and reasoning as above one obtains the opposite inequality for all $\epsilon \in (0, \epsilon_0)$, namely:

$$F_\mu(x) \geq F_\nu(T(x) - \epsilon).$$

This shows that F_ν is constant in $(T(x) - \epsilon_0, T(x))$. By monotonicity and upper semi-continuity of the distribution function this can happen at countably many times, call those constant values y_i. One then has:

$$\{x \mid F_\nu^{[-1]} \circ F_\mu(x) < T(x)\} \subset \bigcup_i \{x \mid F_\mu(x) = y_i\}$$

we claim that since μ is non-atomic the set $\{x \mid F_\mu(x) = l\}$ is a μ null-set for any $l \in [0, 1]$, which would allow us to conclude $T = T_{mon}$ μ-a.e. We will prove the slightly stronger implication

$$\mu \in \mathcal{P}(\mathbb{R}) \text{ atomless} \Rightarrow F_{\mu \#}\mu = \mathscr{L}|_{[0,1]}.$$

The conclusion would then follow quickly from

$$\mu(\{x \mid F_\mu(x) = l\}) = \mu(F_\mu^{-1}(l)) = \mathscr{L}|_{[0,1]}(l) = 0.$$

To see that the above must be true, we observe that F_μ is continuous, being μ atomless. Indeed, if $x_n \uparrow x$,

$$F(x) - F_\mu(x_n) = \mu((-\infty, x]) - \mu((-\infty, x_n]) = \mu((x_n, x]) \to \mu(x) = 0$$

where the last part holds since we have a nested sequence of sets and μ is finite being a probability measure. Continuity togheter with monotonicity give that $\{x \mid F_\mu(x) \leq t\} = (-\infty, x_t]$, with $F_\mu(x_t) = t$. Now let $a \in [0, 1]$:

$$F_{\mu\#}\mu([0, a]) = \mu(\{x \mid F_\mu(x) \leq a\}) = \mu((-\infty, x_a]) = F_\mu(x_a) = a = \mathscr{L}|_{[0,1]}([0, a]),$$

thus the pushforward of μ via its distribution function is indeed the restriceted lebesgue measure as claimed. Now that we have uniqueness of T_{mon} it suffices to show that any optimal transport map $T_{opt} : \mathbb{R} \to \mathbb{R}$ for the quadratic cost must be monotone. This will be a consequence of c-CM. Indeed, consider $\gamma := (id, T)_{\#}\mu(\mathbb{R}^2)$. This must have a c-CM support and so if $(x, y), (x', y') \in \text{Supp}(\gamma)$, one has $y = T(x)$ and $y' = T(x')$ and

$$c(x, y) + c(x', y') \leq c(x, y') + c(x', y)$$

from which one deduces

$$(T(x') - T(x))(x' - x) \geq 0.$$

The rest of the exercise sheet is devoted to the disintegration theorem. The next exercise shows the intuition behind the disintegration, whereas the following ones are increasingly general statements of the theorem. The main difficulty of the proof is contained entirely in Exercise 8.3.

Exercise 8.2. Let $\mu \in \mathscr{M}(\mathbb{R}^2)$ be a finite measure on $\mathbb{R}^2$ that is absolutely continuous with respect to the Lebesgue measure with density $\rho : \mathbb{R}^2 \to \mathbb{R}$. Let $\nu \in \mathscr{M}(\mathbb{R})$ be the measure with density $\eta(x) = \int_{\mathbb{R}} \rho(x, y)dy$. For any $x \in \mathbb{R}$ such that $\eta(x) \neq 0$, let μ_x be the measure with density $\rho_x(y) = \frac{\rho(x,y)}{\eta(x)}$. If $\eta(x) = 0$, then simply set $\mu_x = 0$.

Show that for any $g \in L^1(\mu)$ it holds

$$\int_{\mathbb{R}^2} g(x, y)d\mu(x, y) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x, y)d\mu_x(y)d\nu(x).$$

Solution: Let $g \in L^1(\mu)$. Notice that for all $x \in \mathbb{R}$ such that $\eta(x) = 0$, we have

$$\int_{\mathbb{R}} g(x,y) d\mu_x(y) = 0.$$

Moreover, ρ is a density, hence $\rho(x,y) \geq 0$ almost everywhere. Therefore, given $x_0 \in \mathbb{R}$, we see that $\eta(x_0) = 0 \iff \rho(x_0, y) = 0$ for almost all $y \in \mathbb{R}$. Hence

$$\int_{\{\eta(x)=0\}} g(x,y)\rho(x,y) dy = 0.$$

We then get

$$\begin{aligned}\int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y) d\mu_x(y) d\nu(x) &= \int_{\mathbb{R}\setminus\{\eta(x)=0\}} \int_{\mathbb{R}} g(x,y) \frac{\rho(x,y)}{\eta(x)} dy \; \eta(x) dx \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y)\rho(x,y) dy dx = \int_{\mathbb{R}^2} g(x,y) d\mu(x,y).\end{aligned}$$

Exercise 8.3 (Disintegration for product of compact spaces)**.** Let X, Y be two compact spaces and let $\mu \in \mathscr{M}(X \times Y)$ be a finite measure on the product $X \times Y$. Let us denote $\nu = (\pi_1)_{\#}\mu$ where $\pi_1 : X \times Y \to X$ is the projection on the first coordinate. Prove that there exists a family of probabilities $(\mu_x)_{x \in X} \subseteq \mathcal{P}(Y)$ such that for any $g \in L^1(\mu)$ it holds

$$\int_{X \times Y} g(x,y) d\mu(x,y) = \int_X \int_Y g(x,y) d\mu_x(y) d\nu(x). \tag{1}$$

Do it by following the next steps:

(i) Given $\psi \in C^0(Y)$, consider the map $A_\psi : L^1(X, \nu) \to \mathbb{R}$ given by the formula $A_\psi(\phi) := \int_{X \times Y} \phi(x)\psi(y) d\mu(x,y)$. Prove that the said map is linear continuous and therefore A_ψ can be represented by a function in $L^\infty(X, \nu)$. As an abuse of notation, we denote by $A_\psi(x) \in L^\infty(X, \nu)$ such function, so that the previous map is $\phi \mapsto \int_X \phi(x) A_\psi(x) d\nu(x)$.

(ii) Fix a countable dense subset $S \subseteq C^0(Y)$. Prove that for ν-almost every $x \in X$ the map $\mu_x : S \to \mathbb{R}$ given by $\mu_x(\psi) = A_\psi(x)$ is linear continuous and therefore $\mu_x \in \mathcal{P}(Y)$. Assume that the said family $(\mu_x)_{x \in X}$ satisfies that for any Borel set $E \subset Y$, the map $x \mapsto \mu_x(E)$ is ν-measurable (this is necessary to give a meaning to the integral in the statement).

(iii) Show that the desired statement holds when $g \in L^1(X, \nu) \times S$. Show that this implies that it holds also when $g \in L^1(X, \nu) \times C^0(Y)$. Finally show that this implies it holds also for any $g \in L^1(\mu)$.

Solution:

(i) For $\psi \in C^0(Y)$, define $A_\psi : L^1(X, \nu) \to \mathbb{R}$ by $A_\psi(\phi) := \int_{X \times Y} \phi(x)\psi(y) d\mu(x,y)$. Notice that, since Y is compact, $\|\psi\|_\infty < +\infty$.

We first show that this is well-defined. Notice that for $\phi \in L^1(X,\nu)$ we have

$$\begin{aligned}|A_\psi(\phi)| &= \left|\int_{X\times Y} \phi(x)\psi(y)d\mu(x,y)\right| \leq \int_{X\times Y} |\phi(x)||\psi(y)|d\mu(x,y) \\ &\leq \int_{X\times Y} |\phi(x)|||\psi||_\infty d\mu(x,y) \leq ||\psi||_\infty \int_{X\times Y} |\phi(x)|d\mu(x,y) \\ &= ||\psi||_\infty \int_X |\phi(x)|d\nu(x) = ||\psi||_\infty \|\phi\|_{L^1(X,\nu)} < +\infty.\end{aligned}$$

Notice that linearity of A_ψ follows directly from linearity of the integral, and boundedness from the calculations above. So A_ψ is a linear bounded operator and, since the dual of $L^1(X,\nu)$ is $L^\infty(X,\nu)$, we can now see A_ψ as an $L^\infty(X,\nu)$ function. Observe, also, that $A_\psi \geq 0$ for all $\psi \geq 0$.

(ii) Notice that $C^0(Y)$ is separable (since Y is compact). Fix then a countable dense subset $S = (\psi_n)_n \subset C^0(Y)$, and let $x \in X$. Define $\mu_x : S \to \mathbb{R}$ by $\mu_x(\psi) = A_\psi(x)$ for all $\psi \in S$. Notice that for each $n \in \mathbb{N}$, $\mu_x(\psi_n)$ is defined for $x \in X \setminus E_n$ for some $E_n \subset X$ with $\nu(E_n) = 0$. By taking $E = \bigcup_{n\in\mathbb{N}} E_n$, we have that μ_x is well defined for any $\psi \in S$, and for $x \in X \setminus E$, where $\nu(E) = 0$.

By linearity of the integral, A_ψ satisfies $A_{a\psi_1+b\psi_2} = aA_{\psi_1} + bA_{\psi_2}$ for all $a, b \in \mathbb{R}, \psi_1, \psi_2 \in S$, ν-a.e.. Furthermore, $|\mu_x(\psi)| = |A_\psi(x)| \leq ||\psi||_\infty < \infty$ for all $\psi \in S \subset C^0(Y)$. Therefore μ_x is a linear and bounded operator on S ν-a.e., and extends to a linear and bounded operator on $C^0(Y)$. This means that we can now consider (as an abuse of notation) μ_x as a finite Radon measure on Y by the Riesz representation theorem. That is, the operator becomes $\psi \mapsto \int \psi d\mu_x$. We can extend the map to all $x \in X$ by choosing an arbitrary $\tilde{y} \in Y$ and defining $\mu_x(\psi) = \psi(\tilde{y})$ for all $\psi \in C^0(Y)$ and $x \in E$. By assumption, then, the integrals in the statement are well defined.

Notice, also, that $\mu_x \geq 0$, since $A_\psi \geq 0$ for $\psi \geq 0$, so μ_x is a finite measure. Finally, observe that if $\psi \equiv 1$ in Y, $A_\psi \equiv 1$ in X and $\mu_x(Y) = 1$, that is, $\mu_x \in \mathcal{P}(Y)$.

(iii) If $g \in L^1(X,\nu) \times C(y)$, then $g(x,y)$ can be written as $g(x,y) = g_1(x)g_2(y)$, $g_1 \in L^1(X,\nu)$, $g_2 \in C(y)$.

By unraveling the definitions, and seeing μ_x as a functional on S,

$$\begin{aligned}A_\psi(\phi) &= \int_{X\times Y} \phi(x)\psi(y)d\mu(x,y) = \int_X A_\psi(x)\phi(x)d\nu(x) = \\ &= \int_X \left(\int_Y \psi(y)d\mu_x(y)\right)\phi(x)d\nu(x) = \int_X\int_Y \psi(y)\phi(x)d\mu_x(y)d\nu(x).\end{aligned}$$

This applies to all $g \in L^1(X,\nu) \times S$. By continuous extension (S is dense in $C^0(Y)$), we obtain that this also applies to functions $g \in L^1(X,\nu) \times C^0(Y)$. Again, by density, we have that the previous result holds for all simple functions in the product space $X \times Y$. Since simple functions can monotonically approximate any non-negative measurable function, we are done by splitting any $g \in L^1(\mu)$ into positive and negative parts, and by the monotone convergence theorem.

Exercise 8.4 (💣). [Disintegration for product of Polish spaces] Show the statement of the previous exercise when X and Y are Polish spaces, i.e. they are complete and separable.

Hint: Use Prokhorov's theorem (and Lemma 2.1.9) to find a suitable exhaustion in compact sets that allows to apply the previous exercise.

Solution: Let us consider an exhaustion $X_m \times Y_m \subset X \times Y$ for $m \in \mathbb{N}$, and let us define $\mu_m = \mu \mathbb{1}_{X_m \times Y_m}$ (the restriction to $X_m \times Y_m$). Let us define $\nu_m := (\pi_1)_{\#}\mu_m$, so that it is clear that, as measures, $\nu_m \le \nu_{m'} \le \nu$, with $m \le m'$. Then we have that $\mu_m \rightharpoonup \mu$, and also notice that $\nu_m \rightharpoonup \nu$.

From Exercise 8.3 we also have probability measures μ_x^m such that (1) is satisfied, with μ_m, ν_m, and μ_x^m, and for all $g \in L^1(\mu_m)$.

Let now $m \le m'$. Observe that, in this case, for any $g \in C^0(X_m \times Y_m)$,

$$\int_{X_m} \int_{Y_m} g(x,y) d\mu_x^m(y) d\nu_m(x) = \int_{X_m} \int_{Y_m} g(x,y) d\mu_x^{m'}(y) d\nu_{m'}(x).$$

This implies that, as measures, we have

$$\mu_x^m(y)\nu_m(x) = \mu_x^{m'}(y)\nu_{m'}(x) \quad \text{in} \quad X_m \times Y_m.$$

Now take any fixed $\bar{m}$, and consider $Y_{\bar{m}}$. Observe that, if $m \le \bar{m}$, $\mu_x^m(Y_{\bar{m}}) = 1$, since μ_x^m is a probability measure with support in $Y_m \subset Y_{\bar{m}}$. On the other hand, if $m > \bar{m}$, from the previous equality we have

$$\nu_{\bar{m}}(x) = \mu_x^{\bar{m}}(Y_{\bar{m}})\nu_{\bar{m}}(x) = \mu_x^m(Y_{\bar{m}})\nu_m(x)$$

for $\nu_{\bar{m}}$-a.e. $x \in X_{\bar{m}}$. Notice that, since $\nu_m \uparrow \nu$, this implies that for $\nu_{\bar{m}}$-a.e. $x \in X_{\bar{m}}$, $\mu_x^m(Y_{\bar{m}}) \ge 1 - \omega_x(\bar{m})$, where $\omega_x(\bar{m}) \to 0$ as $\bar{m} \to \infty$. In particular, for ν-a.e. $x \in X$, μ_x^m is tight, and μ_x^m weakly converges to some measure, that we define as $\mu_x^m \rightharpoonup \mu_x$.

Now, from the weak convergence of ν_m and μ_x^m to ν and μ_x for ν-a.e. $x \in X$ we deduce the desired result. Indeed, on the one hand we have that for ν-a.e. $x \in X$,

$$f_m(x) := \int_{Y_m} g(x,y) d\mu_x^m(y) \to f(x) := \int_{Y_m} g(x,y) d\mu_x(y)$$

pointwise. Notice that if g is bounded then they are also uniformly bounded, since μ_x^m are all probability measures. On the other hand, we have to prove

$$\int_{X_m} f_m(x) d\nu_m(x) \to \int_X f(x) d\nu(x).$$

Observe that we can consider the measure $\delta_m := \nu - \nu_m$, such that $\delta_m(X) \to 0$ as $m \to \infty$. In particular, we have by the dominated convergence theorem that

$$\int_{X_m} f_m(x) d\nu(x) \to \int_X f(x) d\nu(x),$$

and on the other hand,

$$\int_{X_m} f_m(x)d\delta_m(x) \leq \|f_m\|_{L^\infty(X_m)}\delta_m(X) \to 0.$$

Exercise 8.5 (Disintegration for fibers of a map)**.** Let X, Y be two Polish spaces, let $f : Y \to X$ be a Borel map and let $\mu \in \mathscr{M}(Y)$ be a finite measure on Y. Let us denote $\nu := f_\#\mu$. Show that there exists a family of probabilities $(\mu_x)_{x\in X} \subseteq \mathcal{P}(Y)$ such that:

(i) For ν-almost every $x \in X$ the measure μ_x is supported on the fiber $f^{-1}(x)$.

(ii) For any $g \in L^1(\mu)$ it holds

$$\int_Y g(y)d\mu(y) = \int_X \int_{f^{-1}(x)} g(y)d\mu_x(y)d\nu(x).$$

Hint: Apply the previous exercise on the measure $(f \times \mathrm{id})_\#\mu$.

Solution: Let $\gamma = (f \times \mathrm{id})_\#\mu \in \mathscr{M}(X \times Y)$. Notice that $\nu = (\pi_1)_\#\gamma$. We can then use Exercise 8.4 to find a family of probabilities $(\mu_x)_{x\in X} \subset \mathcal{P}(Y)$ such that for all $g \in L^1(\gamma)$,

$$\int_X \int_Y g(x,y)d\mu(y)d\nu(x) = \int_{X\times Y} g(x,y)d\gamma(x,y) = \int_X \int_Y g(x,y)d\mu_x(y)d\nu(x).$$

In particular, if we define the set

$$E := \bigcup_{x\in X} \{x\} \times (Y \setminus f^{-1}(x)) \subset X \times Y,$$

then

$$\gamma(E) = \int_E d\gamma(x,y) = \int_X \int_{Y\setminus f^{-1}(x)} d\mu_x(y)d\nu(x).$$

However, notice that $\gamma(E) = \mu((f \times \mathrm{id})^{-1}(E)) = 0$, since

$$(f \times \mathrm{id})^{-1}(E) = \{y \in Y : (f(y), y) \in E\} = \{y \in Y : y \in Y \setminus f^{-1}(f(y))\} = \emptyset.$$

That is,

$$\int_X \int_{Y\setminus f^{-1}(x)} d\mu_x(y)d\nu(x) = 0,$$

which means that $d\mu_x(Y \setminus f^{-1}(x)) = 0$ for ν-a.e. $x \in X$. That is, $\mathrm{supp}(\mu_x) \subset f^{-1}(x)$ for ν-a.e. $x \in X$. Let now $g \in L^1(\mu)$ and see it as $g \in L^1(\gamma)$, in the sense that $g(x,y) = g(y)$ for all $x \in X$. Then,

$$\begin{aligned}\int_Y g(y)d\mu(y) &= \int_{X\times Y} g(x,y)d\gamma(x,y) = \int_X \int_Y g(x,y)d\mu_x(y)d\nu(x)\\ &= \int_X \int_{f^{-1}(x)} g(y)d\mu_x(y)d\nu(x)\end{aligned}$$

by Exercise 8.4 and the fact that μ_x is concentrated on $f^{-1}(x)$.