Density of transport maps
and relaxation from Monge to Kantorovitch

This document aims at clarifying the fact that the formulation by Kantorovitch of the optimal transport problem is nothing but the relaxation of the one by Monge. The key point will be the density of the plans induced by transport maps in the set of all transport plans.

We call *transport plans* all the elements of $\Pi(\mu,\nu) := \{\gamma \in \mathcal{P}(\Omega\times\Omega) : (\pi_x)_\#\gamma = \mu, (\pi_y)_\#\gamma = \nu, \}$. Among these plans there are those which are induced by a transport map T, i.e. those of the form $\gamma_T := (id \times T)_\#\mu$, where $T : \Omega \times \Omega$ is a transport map from μ to ν, i.e. it satisfies $T_\#\mu = \nu$.

Let us set $J(\gamma) := \int_{\Omega\times\Omega} c\, d\gamma$. Since we know that $\int_\Omega c(x, T(x))d\mu = \int_{\Omega\times\Omega} c\, d\gamma_T = J(\gamma_T)$, Monge's problem may be re-written as

$$\min \tilde{J}(\gamma) : \gamma \in \Pi(\mu,\nu),$$

where

$$\tilde{J}(\gamma) = \begin{cases} J(\gamma) & \text{if } \gamma = \gamma_T, \\ +\infty & \text{otherwise.} \end{cases}$$

This is simple to understand : the definition of $\tilde{J}$ forces to restrict the minimization to those plan induced by a transport map. This fact is useful in order to consider Monge's and Kantorovitch's problems as two problems on the *same set of admissible objects*, where the only difference is the functional to be minimized, $\tilde{J}$ or J.

The question is now : why did Kantorovitch decided to replace $\tilde{J}$ with J? Can we easily prove that $\inf J = \inf \tilde{J}$? this is obviously true when, by chance, the minimizer of J is of the form $\gamma = \gamma_T$, since in this case we would have equality of the two minima. But is it possible to justifiy the procedure in general?

The main mathematical justification comes from the following notion of *relaxation*.

Definition 1. Lef $F : X \to \mathbb{R} \cup \{+\infty\}$ be a given functional on a metric space X. We define the relaxation of F as the functional $\overline{F} : X \to \mathbb{R} \cup \{+\infty\}$ which is the maximal functional among those $G : X \to \mathbb{R} \cup \{+\infty\}$ hich are lower semicontinuous and such that $G \le F$. This functional exists since the supremum of an arbitrary family of l.s.c. functions is also l.s.c.. Moreover, we also have a representation formula, which is easy to prove:

$$\overline{F}(x) = \inf\{\liminf_n F(x_n) : x_n \to x\}.$$

A consequence of the definition is also that $\inf F = \inf \overline{F}$ (this latter infimum, that of $\overline{F}$, being often a minimum, when X is compact). This is easy to check if one considers that $F \ge \overline{F}$, which implies $\inf F \ge \inf \overline{F}$, but we also have that F is larger than the constant $l := \inf F$, and a constant function is l.s.c.. Hence $\overline{F} \ge l$ and $\inf \overline{F} \ge \inf F$.

Here we claim that, under some assumptions, J is actually the relaxation of $\tilde{J}$. It will happen, in this case, by chance, that this relaxation is also continuous, instead of only semi-continuous, and that it coincides with $\tilde{J}$ on $\{\tilde{J} < +\infty\}$.

The assumptions are the following: we take $\Omega \subset \mathbb{R}^n$ to be compact, c continuous and μ atomless (i.e. for every $x \in \Omega$ we have $\mu(\{x\}) = 0$).

We need some preliminary results.

Lemma 0.1. *If μ, ν are two probability measures on the real line $\mathbb{R}$ and μ is atomless, then there exists at least a transport map T such that $T_{\#}\mu = \nu$.*

Proof. Just consider the monotone increasing map T, which is well defined $\mu-$a.e. (thanks to the absence of atoms) and which, by the way, optimizes the quadratic cost (but here we don't care about it). □

Lemma 0.2. *There exists a Borel map $\sigma_n : \mathbb{R}^n \to \mathbb{R}$ which is injective, its image is a Borel subset of $\mathbb{R}$, and its inverse map is Borel measurable as well.*

Proof. First notice that it is sufficient to prove this result for $n = 2$, since then one can proceed by induction: if a map σ_{n-1} is given, defined on $\mathbb{R}^{n-1}$, then one can produce a map σ_n by considering $\sigma_n(x_1, x_2, dots, x_n) = \sigma_2(x_1, \sigma_{n-1}(x_2, x_3, \ldots, x_n))$.

Then, notice also that it is enough to define a map from $]0,1[^2$, since one can go from $\mathbb{R}^2$ to $]0,1[^2$ by considering the map $(x, y) \mapsto (\frac{1}{2} + \frac{1}{\pi} \arctan x, \frac{1}{2} + \frac{1}{\pi} \arctan y)$.

Then, consider the map which associates to the pair (x, y), where $x = 0, x_1x_2x_3 \ldots$ and $y = 0, y_1y_2y_3 \ldots$ in decimal (or binary) notation, the point $0, x_1y_1x_2y_2x_3y_3 \ldots$. In order to avoid ambiguities, we can decide that no decimal notation is allowed to end with a periodic 9 (i.e. $0, 347299999 \ldots$ is to be written as $0, 3473$). This is why the image of this map will not be the whole interval, since the points like $0, 39393939 \ldots$ are not obtained through this map. But this set of points is actually Borel measurable.

It is not difficult neither to check that the map is Borel measurable, as well as its inverse, since the pre-image of every interval defined by prescribing the first $2k$ digits of a number in $\mathbb{R}$ is just a rectangle in $\mathbb{R}^2$, the product of two intervals defined by prescribing the first k digits of every component. These particular intervals being a base for the Borel tribe, this proves the measurability we need. □

Corollary 0.3. *If μ, ν are two probability measures on $\mathbb{R}^n$ and μ is atomless, then there exists at least a transport map T such that $T_{\#}\mu = \nu$.*

Proof. This is just obtained by considering a transport map T from $(\sigma_n)_{\#}\mu$ to $(\sigma_n)_{\#}\mu$ and then composing with σ_n and σ_n^{-1}. □

A last lemma

Lemma 0.4. *Consider on a compact metric space X, endowed with a probability $\lambda \in \mathcal{P}(X)$, a sequence of partitions G_n, each G_n being a family of disjoint subsets $C_{i,n}$ such that $\bigcup_{i \in I_n} C_{i,n} = X$ for every n. Suppose that size$(G_n) := \max_i \operatorname{diam}(C_{i,n})$ tends to 0 and consider a sequence of probability* *measures λ_n on X such that, for every n and $i \in I_n$, the equality $\lambda_n(C_{i,n}) = \lambda(C_{i,n}) = m_{i,n}$. Then $\lambda_n \rightharpoonup \lambda$.*

Proof. It is sufficient to take a continuous function $\phi \in C(X)$ and notice that

$$\begin{aligned}\left|\int_X \phi d\lambda_n - \int_X \phi d\lambda\right| &\leq \sum_{i\in I_n}\left|\int_{C_{i,n}} \phi d\lambda_n - \int_{C_{i,n}} \phi d\lambda\right| \\ &\leq \omega(\mathrm{diam}(C_{i,n}))\sum_{i\in I_n} m_{i,n} = \omega(\mathrm{diam}(C_{i,n})) \to 0,\end{aligned}$$

where ω is the modulus of continuity of ϕ. This is justified by the fact that, whenever two measures have the same mass on a set $C \subset X$, since the oscillation of ϕ on the same set does not exceed $\omega(\mathrm{diam}(C))$, the difference of the two integrals is no more than this number times the common mass.

This proves $\int \phi d\lambda_n \to \int \phi d\lambda$ and hence $\lambda_n \to \lambda$. □

We can now prove the following

Theorem 0.5. *On a compact subset of $\mathbb{R}^n$, the set of plans γ_T induced by a transport is dense in the set of plans $\Pi(\mu,\nu)$ whenever μ is atomless.*

Proof. Fix n, and consider any partition of Ω into sets $K_{i,n}$ of diameter smaller than $1/(2n)$ (for instance, small cubes). The sets $C_{i,j,n} := K_{i,n} \times K_{j,n}$ make a partition of $\Omega \times \Omega$ with size smaller than $1/n$.

Let us now take any measure $\gamma \in \Pi(\mu,\nu) \subset \mathcal{P}(\Omega \times \Omega)$. Thanks to Lemma 0.4, we will get the desired density if we are able to build a transport T sending μ to ν such that γ_T gives the same mass as γ to each one of the sets $C_{i,j,n}$. To do this, define the columns $Col_{i,n} := K_{i,n} \times \Omega$ and denote by $\gamma_{i,n}$ the restriction of γ on $Col_{i,n}$. Its marginal will be denoted by $\mu_{i,n}$ and $\nu_{i,n}$. Consider now, for each i, a transport map $T_{i,n}$ sending $\mu_{i,n}$ to $\nu_{i,n}$. It exists thanks to Corollary 0.3, since each $\mu_{i,n}$ is a submeasure of μ and is atomless as well. Since the $\mu_{i,n}$ are concentrated on disjoint sets, by "gluing" the transports $T_{i,n}$ we get a transport T sending μ to ν (using $\sum_i \mu_{i,n} = \mu$ and $\sum_i \nu_{i,n} = \nu$).

It is enough to check that γ_T gives the same mass as γ to every $C_{i,j,n}$, but it is easy to prove. Actually, this mass equals that of $\gamma_{T_{i,n}}$ and $\gamma_{T_{i,n}}(C_{i,j,n}) = \mu_{i,n}(\{x : x \in K_{i,n}, T_{i,n}(x) \in K_{j,n}\}) = \mu_{i,n}(\{x : T_{i,n}(x) \in K_{j,n}\}) = \nu_{i,n}(K_{j,n}) = \gamma(K_{i,n} \times K_{j,n})$. □

The relaxation result is just a consequence.

Theorem 0.6. *Under the abovementioned assumptions, J is the relaxation of $\tilde{J}$.*

Proof. First notice that, since J is continuous, then it is l.s.c. and since, due to the definition, we have $J \leq \tilde{J}$, then J is necessarily smaller than the relaxation of $\tilde{J}$. We only need to prove that, for each γ, we can find a sequence of transports T_n such that $\gamma_{T_n} \rightharpoonup \gamma$ and $\tilde{J}(\gamma_{T_n}) \to J(\gamma)$, so that the infimum in the sequential characterization of the relaxed functional (see definition) will be smaller than J, thus proving the equality. Actually, since for $\gamma = \gamma_{T_n}$ the two functionals J and $\tilde{J}$ coincide, and since J is continuous, we only need to produce a sequence T_n such that $\gamma_{T_n} \rightharpoonup \gamma$. This is possible thanks to Theorem 0.5 □