

c –Cyclical Monotonicity of the support of optimal transport plans

This document aims at clarifying the proof of the necessary condition for the optimality of a transport plan $\gamma \in \Pi(\mu, \nu)$ stating that its support must be c –Cyclically Monotone, when the cost c is continuous. The proof will try to be easier than what I presented in class (I thank Patrick Gerard and Vianney Perchet for some observations during the pause). Some applications and remarks will also be included.

First, some definitions.

Definition 1. Once a function $c : \Omega \times \Omega \rightarrow \mathbb{R} \cup +\infty$ is given, we say that a set $\Gamma \subset \Omega \times \Omega$ is c –cyclically monotone (briefly c –CM) if, for every $k \in \mathbb{N}$, every permutation σ and every finite family of points $(x_1, y_1), \dots, (x_k, y_k) \in \Gamma$ we have

$$\sum_{i=1}^k c(x_i, y_i) \leq \sum_{i=1}^k c(x_i, y_{\sigma(i)}).$$

The word “cyclical” refers to the fact that, since every σ is the disjoint composition of cycles, it is enough to check this property for cyclical permutations, i.e. replacing $\sum_{i=1}^k c(x_i, y_{\sigma(i)})$ with $\sum_{i=1}^k c(x_i, y_{i+1})$ in the definition (with the obvious convention $y_{k+1} = y_1$). The word “monotone” refers instead to the behavior of those sets when $c(x, y) = -x \cdot y$.

It is useful to recall the following theorem, which is a generalization of a theorem by Rockafellar in convex analysis, and whose proof may be found in [1], even if it is originally taken from [2]

Theorem 0.1. *If Γ is a c –CM set and c is real valued, then there exist a pair of functions ϕ, ψ such that $\phi(x) = \inf_y c(x, y) - \psi(y)$, $\psi(y) = \inf_x c(y, x) - \phi(x)$ and*

$$\Gamma \subset \{(x, y) \in \Omega \times \Omega : \phi(x) + \psi(y) = c(x, y)\}$$

(these functions are c –concave functions and Γ is included in the graph G_ϕ^c).

To introduce the following theorem we recall the definition

Definition 2. On a separable metric space X , the support of a measure γ is defined as the smallest closed set on which γ is concentrated, i.e.

$$\text{spt}(\gamma) := \bigcap \{A : A \text{ is closed and } \gamma(X \setminus A) = 0\}.$$

This is well defined since the intersection may be taken countable, due to the assumption on the space X . Moreover, there exists also this characterization

$$\text{spt}(\gamma) = \{x \in X : \gamma(B(x, r)) > 0 \text{ for all } r > 0\}.$$

We can now prove

Theorem 0.2. *If γ is an optimal transport plan for the cost c (i.e. it minimizes $\int c d\gamma$ over $\Pi(\mu, \nu)$) and c is continuous, then $\text{spt}(\gamma)$ is a c -CM set.*

Proof. Suppose by contradiction that there exist k, σ and $(x_1, y_1), \dots, (x_k, y_k) \in \text{spt}(\gamma)$ such that

$$\sum_{i=1}^k c(x_i, y_i) > \sum_{i=1}^k c(x_i, y_{\sigma(i)}).$$

Take now $\varepsilon < \frac{1}{2k} \left(\sum_{i=1}^k c(x_i, y_i) - \sum_{i=1}^k c(x_i, y_{\sigma(i)}) \right)$. By continuity of c , there exists r such that for all $i = 1, \dots, k$ and for all $(x, y) \in B(x_i, r) \times B(y_i, r)$ we have $c(x, y) > c(x_i, y_i) - \varepsilon$ and for all $(x, y) \in B(x_i, r) \times B(y_{\sigma(i)}, r)$ we have $c(x, y) < c(x_i, y_{\sigma(i)}) + \varepsilon$.

Now consider $V_i := B(x_i, r) \times B(y_i, r)$ and notice that $\gamma(V_i) > 0$ for every i , due to the condition $(x_i, y_i) \in \text{spt}(\gamma)$. Define the measures $\gamma_i := \gamma \llcorner V_i / \gamma(V_i)$ and $\mu_i := (\pi_x)_{\#} \gamma_i$, $\nu_i := (\pi_y)_{\#} \gamma_i$. Take $\alpha < \frac{1}{k} \min_i \gamma(V_i)$.

For every i , build a measure $\tilde{\gamma}_i \in \Pi(\mu_i, \nu_{\sigma(i)})$ at will (for instance take $\tilde{\gamma}_i = \mu_i \otimes \nu_i$).

Now define

$$\tilde{\gamma} := \gamma - \alpha \sum_{i=1}^k \gamma_i + \alpha \sum_{i=1}^k \tilde{\gamma}_i.$$

We want to find a contradiction by proving that $\tilde{\gamma}$ is a better competitor than γ in the transport problem, i.e. $\tilde{\gamma} \in \Pi(\mu, \nu)$ and $\int c d\tilde{\gamma} < \int c d\gamma$.

First we check that $\tilde{\gamma}$ is a positive measure. It is sufficient to check that $\gamma - \alpha \sum_{i=1}^k \gamma_i$ is positive, and, for that, the condition $\alpha \gamma_i < \frac{1}{k} \gamma$ will be enough. This condition is satisfied since $\alpha \gamma_i = (\alpha / \gamma(V_i)) \gamma \llcorner V_i$ and $\alpha / \gamma(V_i) \leq \frac{1}{k}$.

Now, let us check the condition on the marginals of $\tilde{\gamma}$. We have

$$\begin{aligned} (\pi_x)_{\#} \tilde{\gamma} &= \mu - \alpha \sum_{i=1}^k (\pi_x)_{\#} \gamma_i + \alpha \sum_{i=1}^k (\pi_x)_{\#} \tilde{\gamma}_i = \mu - \alpha \sum_{i=1}^k \mu_i + \alpha \sum_{i=1}^k \mu_i = \mu, \\ (\pi_y)_{\#} \tilde{\gamma} &= \nu - \alpha \sum_{i=1}^k (\pi_y)_{\#} \gamma_i + \alpha \sum_{i=1}^k (\pi_y)_{\#} \tilde{\gamma}_i = \nu - \alpha \sum_{i=1}^k \nu_i + \alpha \sum_{i=1}^k \nu_{\sigma(i)} = \nu. \end{aligned}$$

Finally, let us estimate $\int c d\gamma - \int c d\tilde{\gamma}$ and prove that it is positive, thus concluding the proof. We have

$$\begin{aligned} \int c d\gamma - \int c d\tilde{\gamma} &= \alpha \sum_{i=1}^k \int c d\gamma_i - \alpha \sum_{i=1}^k \int c d\tilde{\gamma}_i \\ &\geq \alpha \sum_{i=1}^k (c(x_i, y_i) - \varepsilon) - \alpha \sum_{i=1}^k (c(x_i, y_{\sigma(i)}) + \varepsilon) \\ &= \alpha \left(\sum_{i=1}^k c(x_i, y_i) - \sum_{i=1}^k c(x_i, y_{\sigma(i)}) - 2k\varepsilon \right) > 0, \end{aligned}$$

where we used the fact that γ_i is concentrated on $B(x_i, r) \times B(y_i, r)$, $\tilde{\gamma}_i$ on $B(x_i, r) \times B(y_{\sigma(i)}, r)$, and that they have unit mass (due to the rescaling by $\gamma(V_i)$). \square

Notice that the previous theorem, together with Theorem 0.1, guarantees the existence of a solution to the dual problem

$$\max \left\{ \int \phi d\mu + \int \psi d\nu : \phi, \psi \in C(\Omega), \phi(x) + \psi(y) \leq c(x, y) \right\},$$

and the equality between this maximum and the minimum of $\int cd\gamma$ over $\Pi(\mu, \nu)$. If one follows this way, it is not necessary to evoke abstract convex analysis theorems guaranteeing the equality when one exchanges inf and sup. Notice also that the opposite path could also be possible, i.e. using the existence of an optimal pair (ϕ, ψ) (with ϕ a Kantorovitch potential and $\psi = \phi^c$) to prove that $\text{spt}(\gamma)$ is c -CM. Anyway, the two concepts are quite important so that it is worthwhile to prove them in two different ways. It is also clear that, besides the case where c is continuous and real-valued, the two approaches (looking at the support or looking at the potential) could not be equivalent, even if we do not develop here these more difficult cases.

An application of the previous theorem, for c continuous, is the following result concerning the one-dimensional case. To introduce it, we recall that, for every pair $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with μ non-atomic, there exists a unique non-decreasing map $T : \mathbb{R} \rightarrow \mathbb{R}$ such that $T_{\#}\mu = \nu$ (defined uniquely μ -a.e.). Let us denote this map by T_{mon} . This map is, by the way, the optimal transport map for the quadratic cost (since in this case we know that the optimal transport is the gradient of a convex function, which means, in dimension one, a non-decreasing function). It is important to notice that, any transport plan $\gamma \in \Pi(\mu, \nu)$, satisfying the property

$$(x, y), (x', y') \in \text{spt}(\gamma) \quad x < x' \Rightarrow y \leq y'$$

actually coincides with $\gamma_{T_{mon}}$. This is easy to check: for any point x one can define the interval I_x as the minimal interval I such that $\text{spt}(\gamma) \cap \{x\} \times \mathbb{R} \subset \{x\} \times I$. The above property implies that the interior of all these intervals are disjoint (and ordered). In particular, there can be at most a countable quantity of points such that I_x is not a singleton. Since these points are μ -negligible (as a consequence of μ being atomless), we can define μ -a.e. a map T such that γ is concentrated on the graph of T . This map will be monotone non-decreasing as a consequence of the implication above, and this gives $T = T_{mon}$ since we already know the uniqueness of a non-decreasing map with fixed marginals.

Theorem 0.3. *Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a strictly convex function, and $\mu, \nu \in \mathcal{P}(\mathbb{R})$ be compactly supported measures with μ non-atomic. Then the optimal transport problem with cost $c(x, y) = h(y - x)$ over $\Pi(\mu, \nu)$ has a unique solution, which is given by $\gamma_{T_{mon}}$.*

Moreover, if the strict convexity assumption is withdrawn and h is only convex, then the same T_{mon} is actually an optimal transport map, but no uniqueness is guaranteed anymore.

Proof. We will use the fact that the support of any optimal γ is a c -CM set Γ . This means in particular that $(x, y), (x', y') \in \Gamma$ implies

$$h(y - x) + h(y' - x') \leq h(y' - x) + h(y - x'). \tag{0.1}$$

We only need to show that this implies (in the strictly convex case) a monotone behavior : we will actually deduce from (0.1) that $x < x'$ implies $y \leq y'$, and this will allow to conclude as we noticed above.

To prove $y \leq y'$ suppose by contradiction $y > y'$ and denote $a = y - x$, $b = y' - x'$ and $c = x' - x$. Condition (0.1) reads $h(a) + h(b) \leq h(b + c) + h(a - c)$. Moreover, the assumption $y < y'$ implies $b + c < a$. We also need to recall that $c > 0$ ($x < x'$) : this implies that $b + c$ and $a - c$ are located in the segment between b and a (and $b < a$). More precisely, we have

$$b + c = (1 - t)b + ta, \quad a - c = tb + (1 - t)a, \quad \text{for } t = \frac{c}{a - b} \in]0, 1[.$$

Thus, convexity yields

$$\begin{aligned} h(a) + h(b) &\leq h(b + c) + h(a - c) \\ &< (1 - t)h(b) + th(a) + th(b) + (1 - t)h(a) = h(a) + h(b). \end{aligned}$$

This gives a contradiction and proves the thesis in the strictly convex case.

The statement when h is only convex is obtained by approximation through the transport problem for the cost $c_\varepsilon(x, y) := h(y - x) + \varepsilon|y - x|^2$. In this case we can say that $\gamma_{T_{mon}}$ optimizes the cost $\int c_\varepsilon d\gamma$ and hence

$$\int (h(y - x) + \varepsilon|y - x|^2) d\gamma_{T_{mon}} \leq \int (h(y - x) + \varepsilon|y - x|^2) d\gamma,$$

for all $\gamma \in \Pi(\mu, \nu)$. Passing to the limit as $\varepsilon \rightarrow 0$, since $\int |x - y|^2 d\gamma, \int |x - y|^2 d\gamma_{T_{mon}} < +\infty$, we get that $\gamma_{T_{mon}}$ also optimizes the cost c . \square

Notice that the assumptions on the compactness of the support has exactly been put so as to guarantee the finiteness of the integral for the quadratic cost. Yet, it is possible to choose other strictly convex approximation in much more general situations, but we will not enter into details about these technicalities.

After this application to the one-dimensional case, it is worthwhile to give some details about the case where c is not continuous. This could be, by the way, applied exactly as in the previous theorem to the one dimensional situation for $c(x, y) = h(x - y)$ and h is convex but not real-valued, which was the assumption guaranteeing its continuity. We will give two results about l.s.c. costs.

The first concerns the validity of the duality formula. This means the equality

$$\min \left\{ \int c d\gamma, \gamma \in \Pi(\mu, \nu) \right\} = \sup \left\{ \int \phi d\mu + \int \psi d\nu : \phi, \psi \in C(\Omega), \phi(x) + \psi(y) \leq c(x, y) \right\}. \quad (0.2)$$

By now, we have established this equality when c is continuous, also proving that the dual problem admits a maximizing pair. We also know that an inequality is always true : the minimum on the left is always larger than the maximum on the right (just integrate w.r.t. γ the condition on (ϕ, ψ)). More precisely, we are able to deal with the uniformly continuous case (since we want to guarantee

continuity of c -concave functions of the form $\phi(x) = \inf_y c(x, y) - \psi(y)$. This means that we can handle the case where Ω is compact, or we need to add the uniform continuity assumption on c .

To deal with a l.s.c. cost c bounded from below, we will use the fact there exists a sequence c_k of continuous functions (each one being k -Lipschitz) increasingly converging to c . We need the following lemma.

Lemma 0.4. *Suppose that c_k and c are l.s.c. and bounded from below and that c_k converges increasingly to c . Then*

$$\lim_{k \rightarrow \infty} \min \left\{ \int c_k d\gamma, \gamma \in \Pi(\mu, \nu) \right\} = \min \left\{ \int c d\gamma, \gamma \in \Pi(\mu, \nu) \right\}.$$

Proof. Due to the increasing limit condition, we have $c_k \leq c$ and hence the limit on the left (which exists by monotonicity) is obviously smaller than the quantity on the right. Now consider a sequence $\gamma_k \in \Pi(\mu, \nu)$, built by picking an optimizer for each cost c_k . Up to subsequences, due to the tightness of $\Pi(\mu, \nu)$, we can suppose $\gamma_k \rightharpoonup \bar{\gamma}$. Fix now an index j . Since for $k \geq j$ we have $c_k \geq c_j$, we have

$$\lim_k \min \left\{ \int c_k d\gamma, \gamma \in \Pi(\mu, \nu) \right\} = \lim_k \int c_k d\gamma_k \geq \liminf_k \int c_j d\gamma_k.$$

By semicontinuity of the integral cost c_j we have

$$\liminf_k \int c_j d\gamma_k \geq \int c_j d\bar{\gamma}.$$

Hence we have obtained

$$\lim_k \min \left\{ \int c_k d\gamma, \gamma \in \Pi(\mu, \nu) \right\} \geq \int c_j d\bar{\gamma}.$$

Since j was arbitrary and $\lim_j \int c_j d\bar{\gamma} = \int c d\bar{\gamma}$ by monotone convergence, we also have

$$\lim_k \min \left\{ \int c_k d\gamma, \gamma \in \Pi(\mu, \nu) \right\} \geq \int c d\bar{\gamma} \geq \min \left\{ \int c d\gamma, \gamma \in \Pi(\mu, \nu) \right\}.$$

This concludes the proof. Notice that it also gives, as a byproduct, the optimality of $\bar{\gamma}$ for the limit cost c . \square

We can now establish the validity of the duality formula for semi-continuous costs.

Theorem 0.5. *If c is l.s.c. and bounded from below, then (0.2) holds.*

Proof. Consider a sequence c_k of k -Lipschitz functions approaching c increasingly. Then the same duality formula holds for c_k , and hence we have

$$\begin{aligned} \min \left\{ \int c_k d\gamma, \gamma \in \Pi(\mu, \nu) \right\} &= \max \left\{ \int \phi d\mu + \int \psi d\nu : \phi, \psi \in C(\Omega), \phi(x) + \psi(y) \leq c_k(x, y) \right\} \\ &\leq \sup \left\{ \int \phi d\mu + \int \psi d\nu : \phi, \psi \in C(\Omega), \phi(x) + \psi(y) \leq c(x, y) \right\}, \end{aligned}$$

where the inequality is justified by the fact that $c_k \leq c$ and hence every pair (ϕ, ψ) satisfying $\phi(x) + \psi(y) \leq c_k(x, y)$ also satisfies $\phi(x) + \psi(y) \leq c(x, y)$. The conclusion follows by letting $k \rightarrow +\infty$, using Lemma 0.4. Notice that for the cost c we cannot guarantee the existence of a maximizing pair (ϕ, ψ) . \square

The duality formula also allows to prove the following c -cyclical monotonicity theorem.

Theorem 0.6. *If c is l.s.c. and γ is an optimal transport plan, then γ is concentrated on a c -CM set Γ (which will not be closed in general).*

Proof. Thanks to the previous theorem the duality formula holds, which means that, if we take a maximizing pair (ϕ_h, ψ_h) in the dual problem, we have

$$\int (\phi_h(x) + \psi_h(y)) d\gamma = \int \phi_h d\mu + \int \psi_h d\nu \rightarrow \int c d\gamma,$$

since the value of $\int c d\gamma$ is the minimum of the primal problem, which is also the maximum of the dual. Yet, we also have $c(x, y) - \phi_h(x) + \psi_h(y) \geq 0$, which implies that the functions $f_h := c(x, y) - \phi_h(x) - \psi_h(y)$, defined on $\Omega \times \Omega$, converge to 0 in $L^1(\Omega \times \Omega, \gamma)$ (since they are positive and their integral tends to 0). As a consequence, up to a subsequence (not relabeled) they also converge pointwisely γ -a.e. to 0. Let $\Gamma \subset \Omega \times \Omega$ be a set with $\gamma(\Gamma) = 1$ where the convergence happens. Let us prove that this set is c -CM. This is true since, for any k, σ and $(x_1, y_1), \dots, (x_k, y_k) \in \Gamma$ we have

$$\sum_{i=1}^k c(x_i, y_i) = \lim_h \sum_{i=1}^k \phi_h(x_i) + \psi_h(y_i) = \lim_h \sum_{i=1}^k \phi_h(x_i) + \psi_h(y_{\sigma(i)}) \leq \sum_{i=1}^k c(x_i, y_{\sigma(i)}). \quad \square$$

We conclude these notes by stressing that the problems about the c -cyclical monotonicity, the duality formula and the existence of optimal potentials (ϕ, ϕ^c) in general situations are a matter of current investigation. Typical questions are the equivalence between the optimality of γ and the fact that γ is concentrated on a c -CM set, as well as the existence of a solution to the dual problem in a suitable functional class (probably not C^0 nor Lipschitz functions but rather BV) for costs taking the value $+\infty$. Possible answers to these questions could find applications also in easier problems for “normal” costs (i.e. uniformly continuous and real-valued, say), since some strategies about these costs pass through decompositions that add additional constraints or through secondary variational problems.

References

- [1] L. Ambrosio, *Lecture Notes on Optimal Transport Problems, Mathematical aspects of evolving interfaces*, CIME Summer School in Madeira, vol. 1812, Springer, 2003.
- [2] L. Ruschendorf, On c -optimal random variables, *Statistics and Probability Letters*, Vol. 27, 1996, 267–270