

A LAND OF MONOTONE PLENTY, *BIS REPETITA*: FROM CLASSICAL TO WEAK OPTIMAL TRANSPORT

VIRGINIE EHRLACHER, RODRIGUE LELOTTE, AND LUCA NENNA

ABSTRACT. The celebrated c -cyclical monotonicity property is shown to boil down to the zeroth-order optimality condition for the optimal transport problem. More precisely, we show that optimality is equivalent to the non-negativity of the linear transport cost functional on the radial cone of admissible perturbations. We then utilise this point of view to extend the c -cyclical monotonicity property to the weak optimal transport problem, for which it corresponds to the first-order optimality condition, namely to the non-negativity of the linearisation of the weak transport cost functional near the optimiser. Altogether, this sheds new light on this monotonicity concept. For both classical and weak optimal transport, we show that this property characterises (under suitable assumptions) optimal transport plans. In the classical case, we recover known results of the literature but with revisited proofs.

CONTENTS

1. Introduction	2
1.1. Contributions	2
1.2. <i>The gist of it</i>	2
1.3. Previous works	4
1.4. Outline of the paper	4
2. A perturbative approach to the classical optimal transport	4
2.1. Setting and notations	4
2.2. The zeroth-order optimality condition for (OT)	5
2.3. On c -cyclical monotonicity and its related notions	7
2.4. <i>Optimality implies c-cyclical monotonicity</i> (\Rightarrow)	9
2.5. <i>c-cyclical monotonicity implies optimality</i> (\Leftarrow)	13
3. Extension of cyclical monotonicity to the weak optimal transport	20
3.1. Setting and notations	20
3.2. The first-order optimality conditions for (WOT)	23
3.3. <i>C-cyclical monotonicity characterises optimality</i>	28
3.4. Removing differentiability: the case of the barycentric cost	30
3.5. A comment on the entropy-regularised problem	39
Appendix A. Proof of Lemma 2	40
References	41

1. INTRODUCTION

The notion of *c-cyclical monotonicity* occupies a central role in optimal transport theory. A transport plan is said to be *c-cyclically monotone* if, in essence, local re-routing of the mass on the support of the transport plan leads to an increase of the overall cost of transport. Under suitable assumptions, this property is both necessary and sufficient for optimality. We refer the reader to the recent survey of De Pascale, Kausamo, and Wyczesany [14] for a broad overview of the subject.

1.1. Contributions. The purpose of the present work is twofold. First, we revisit cyclical monotonicity from a perturbative perspective based on elementary optimality considerations. In particular, we show that this property characterises under suitable assumptions optimal transport plans. Although we recover known results of the literature, the proofs are revisited in light of our approach. Second, building on this approach, we extend the notion of *c-cyclical monotonicity* to the case of the *weak* optimal transport problem by introducing a notion of monotonicity which, contrary to previously existing notions in the literature, genuinely recovers the classical *c-cyclical monotonicity* property as a special case.

1.2. The gist of it. Given two Polish spaces X and Y , two probability measures $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, and a Borel cost function $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$, the classical optimal transport problem reads

$$\inf_{\pi \in \Pi(\mu, \nu)} \mathcal{E}(\pi), \quad \mathcal{E}(\pi) := \int_{X \times Y} c(x, y) \pi(dx, dy),$$

where $\Pi(\mu, \nu)$ denotes the set of transport plans with marginals μ and ν . Our starting point is deceptively simple. If π is an optimal transport plan and if η is an admissible perturbation preserving the marginal constraints, that is a signed Radon measure $\eta \in \mathcal{M}(X \times Y)$ so that $\pi + t\eta \in \Pi(\mu, \nu)$ for $t > 0$ small enough, then it obviously holds

$$\mathcal{E}(\pi + t\eta) \geq \mathcal{E}(\pi).$$

This is just the **zeroth-order optimality condition**. In the classical optimal transport problem, the cost functional is linear, and therefore optimality is equivalent to the non-negativity condition

$$\mathcal{E}(\eta) \geq 0$$

for every admissible perturbation η . Now, it will follow from elementary manipulations that the *c-cyclical monotonicity* exactly rephrases as demanding that \mathcal{E} be non-negative for a special kind of “cyclical” perturbations η . *The gist of our work is to show that \mathcal{E} being non-negative on the set made of these special perturbations is sufficient for it to be non-negative over the entire set of admissible perturbations and vice-versa*, thus yielding sufficiency and necessity of *c-cyclical monotonicity* for optimality in the optimal transport problem. This gives a new and conceptually transparent route to the

equivalence between optimality and cyclical monotonicity. Moreover, the same approach can be used to derive the analogue of c -cyclical monotonicity in other types of optimal transport problems.

Indeed, a central contribution of the paper concerns the weak optimal transport problem. We recall that this problem reads

$$\inf_{\pi \in \Pi(\mu, \nu)} \int_X C(x, \pi_x) \mu(dx),$$

where $(\pi_x)_{x \in X}$ denotes the disintegration of π with respect to its first marginal and the function $C : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ is (typically) convex in its second argument.

A notion coined as C -monotonicity was introduced in the literature by Backhoff-Veraguas, Beiglböck, and Pammer [2] and shown to be necessary and sufficient for optimality under suitable assumptions. However, this notion does not fully recover classical cyclical monotonicity when the weak transport problem reduces to the ordinary optimal transport problem (see Example 2 *infra*). One of the contributions of the present work is to introduce a notion of monotonicity for weak optimal transport, which we call C -cyclical monotonicity, that genuinely extends the classical notion of c -cyclical monotonicity to the weak setting. This notion is formulated in terms of the gradients (or subgradients) of the weak cost with respect to its measure argument. Concretely, if $\rho \mapsto C(x, \rho)$ is differentiable in an appropriate sense (see Section 3.1.1), then our monotonicity condition takes the form

$$\boxed{\sum_{i=1}^n \nabla_{\rho} C(x_i, \pi_{x_i})(y_i) \leq \sum_{i=1}^n \nabla_{\rho} C(x_i, \pi_{x_i})(y_{i+1})}$$

for all $n \in \mathbb{N}$ and all families $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$ where $\Gamma \subset X \times Y$ is a measurable set so that $\pi(\Gamma) = 1$. Here, we let $y_{n+1} := y_1$. This property enjoys several important features. First, it exactly recovers the classical c -cyclical monotonicity when the weak problem boils down to the classical one, namely when the cost of transport is given by

$$C(x, \rho) := \int_Y c(x, y) \rho(dy).$$

Indeed, in this case the gradient is given by $\nabla_{\rho} C(x, \rho)(y) = c(x, y)$, and one recognises that the above condition is exactly the c -cyclical monotonicity property. Second, we show that C -cyclical monotonicity is equivalent to optimality under assumptions comparable to those used in the classical setting.

Our C -cyclical monotonicity property is derived by looking at the **first-order optimality condition** for the weak optimal transport. Indeed, the objective of this problem is

$$\mathcal{E}(\pi) := \int_X C(x, \pi_x) \mu(dx),$$

which, contrary to the classical optimal transport problem, is no longer linear, so we need to go to the first-order to say anything meaningful. Whereas in the classical problem, optimality is encoded by the non-negativity of the cost on admissible perturbations, in the weak setting, the same mechanism survives after replacing the transport cost by the linearised quantity

$$\eta \mapsto \int_{X \times Y} \nabla_{\rho} C(x, \pi_x)(y) \eta(dx, dy).$$

The C -cyclical monotonicity introduced above is then derived from the non-negativity of the above functional once evaluated at these special “cyclical” perturbations already mentioned above.

1.3. Previous works. It is possible that part of our arguments *may* not be completely new to some readers already familiar with the subject. In a way, a part of what we present in this work can be seen as a way of explaining what cyclical monotonicity is all about in a different perspective with a different terminology, thus shedding a new kind of light on this concept. Whether or not these ideas are known to specialists, they are certainly not widely familiar to a broader audience. During our investigation, we came across a couple of works of S. Bianchini and L. Caravenna [9, 10] which share some common features with our own work here and that should be readily advertised. We shall also mention the work of Beiglböck [4] which utilises the ergodic theorem to prove sufficiency of the c -cyclical monotone property for optimality somehow in the same way as we do. Also – as the title of our paper is obviously a nod to this work — we shall also mention the seminal paper [5] — and also [21]. However, the unified formalism presented here and its use as a systematic recipe to derive the monotonicity property to other type of optimal transports is decidedly new — at least to the best of our knowledge.

1.4. Outline of the paper. Section 2 revisits the c -cyclical monotonicity property in the classical optimal transport through the perturbative point of view. We introduce the radial cone of admissible perturbations and show that optimality is equivalent to the zeroth-order optimality condition on this cone. We then relate this condition to finite optimality and to c -cyclical monotonicity. Section 3 is devoted to weak optimal transport. We derive the first-order optimality conditions for weak transport problems, and use them to define the notion of C -cyclical monotonicity and establish its equivalence with optimality under suitable assumptions. We also discuss the case of the barycentric cost (which is non-differentiable).

2. A PERTURBATIVE APPROACH TO THE CLASSICAL OPTIMAL TRANSPORT

2.1. Setting and notations. For a Polish space Z , we write $\mathcal{M}(Z)$ the set of finite (and signed) Radon measures on Z . By a *finite* Radon measure η , it is meant that $\|\eta\|_{TV} < \infty$ where the total variation norm $\|\eta\|_{TV}$ is defined as $\|\eta\|_{TV} := |\eta|(Z)$ where $|\eta| := \eta_+ + \eta_-$ with η_+ and η_- being respectively

the positive and negative parts of the signed measure μ . The space $\mathcal{M}(Z)$ is equipped with the *vague topology*, that is in duality with $C_0(Z)$, *i.e.* the set of real-valued continuous functions on Z vanishing at infinity — when Z is compact, this set is simply the set of all continuous functions. We recall that finite radii balls $\{\eta \in \mathcal{M}(Z) : |\eta|(Z) \leq r\}$ for $0 \leq r < \infty$ are metrisable in this topology provided that Z is Polish and furthermore locally compact, see *e.g.* [12, Theorem 3.29]. Finally, we write $\mathcal{P}(Z) \subset \mathcal{M}(Z)$ be the subset of probability measures.

Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be probability measures on Polish (and locally compact) spaces X and Y , and let $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ be a Borel-measurable cost of transport possibly assuming $+\infty$ as value. We consider the classical optimal transport problem

$$(OT) \quad \min_{\pi \in \Pi(\mu, \nu)} \mathcal{C}(\pi), \quad \mathcal{C}(\pi) := \int_{X \times Y} c(x, y) \pi(dx, dy)$$

where $\Pi(\mu, \nu)$ is the standard set of *transport plans*, that is those probability measures $\pi \in \mathcal{P}(X \times Y)$ so that the first marginal of π (*i.e.* with respect to X) is μ and such that the second marginal of π (*i.e.* with respect to Y) is ν . This also rewrites $\text{pr}_1^\# \pi = \mu$ and $\text{pr}_2^\# \pi = \nu$ where $\text{pr}_1 : X \times Y \rightarrow X$ and $\text{pr}_2 : X \times Y \rightarrow Y$ are the canonical surjections of the Cartesian product $X \times Y$ onto its factors X and Y — and where we use the standard notation $f^\# \pi$ for any measurable map $f : X \times Y \rightarrow Z$ to mean the push-forward measure defined as $f^\# \pi(A) := \pi(f^{-1}(A))$ for all Borel sets $B \subset Z$.

2.2. The zeroth-order optimality condition for (OT). We start with the following simple observation. If $\pi \in \Pi(\mu, \nu)$ is optimal for the optimal transport problem (OT) and π has finite transport cost, meaning that $\mathcal{C}(\pi) < \infty$, then it must be that $\mathcal{C}(\pi + t\eta) \geq \mathcal{C}(\pi)$ for all $t \in \mathbb{R}$ and $\eta \in \mathcal{M}(X \times Y)$ such that $\pi + t\eta$ is still an admissible transport plan, that is $\pi + t\eta \in \Pi(\mu, \nu)$. By linearity of the cost functional \mathcal{C} , this is equivalent to saying that $\mathcal{C}(\eta) \geq 0$ for such perturbations η provided that $t > 0$. Let us then introduce for $\pi \in \Pi(\mu, \nu)$ the set $\mathcal{R}_\pi \subset \mathcal{M}(X \times Y)$ defined as

$$(1) \quad \boxed{\mathcal{R}_\pi := \left\{ \eta \in \mathcal{M}(X \times Y) : \pi + t\eta \in \Pi(\mu, \nu) \text{ for some } t > 0 \right\}.}$$

Then, the previous statement rephrases as follows: *If π is optimal for the optimal transport problem (OT) and has finite transport cost, that is $\mathcal{C}(\pi) < \infty$, then $\mathcal{C}(\eta) \geq 0$ for all $\eta \in \mathcal{R}_\pi$.* In the context of convex analysis, the set \mathcal{R}_π is referred to as the *radial cone* to the (convex) set $\Pi(\mu, \nu)$ at the point π . We refer to Remark 1 and (3) *infra* for a general definition of the radial cone to a convex set. Furthermore, the converse statement holds as well: *If $\pi \in \Pi(\mu, \nu)$ is such that \mathcal{C} is non-negative on \mathcal{R}_π , then π is optimal in (OT) — again, provided that $\mathcal{C}(\pi) < \infty$.* Indeed, for all $\gamma \in \Pi(\mu, \nu)$, one has $\mathcal{C}(\gamma) = \mathcal{C}(\gamma - \pi + \pi) = \mathcal{C}(\gamma - \pi) + \mathcal{C}(\pi) \geq \mathcal{C}(\pi)$ since $\eta := \gamma - \pi$ is obviously an element of \mathcal{R}_π . Thus π is optimal. Altogether, we

obtain the following proposition which is just the **zeroth-order optimality condition** for (OT):

Proposition 1. *A transport plan $\pi \in \Pi(\mu, \nu)$ with finite cost $\mathcal{C}(\pi) < \infty$ is optimal in (OT) if and only if $\mathcal{C}(\eta) \geq 0$ for all $\eta \in \mathcal{R}_\pi$, where \mathcal{R}_π is the radial cone to the feasible set $\Pi(\mu, \nu)$ at the point π as defined in (1).*

Remark 1. *Proposition 1 is just one particular instance of a much general situation. Let X be a normed vector space, $L : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a linear function on X and $C \subset X$ be a convex set. Let us consider the minimisation problem*

$$(2) \quad \min_{x \in C} Lx.$$

Then, $x \in C$ is optimal for (2) if and only if (provided that $Lx < \infty$) $Lh \geq 0$ for all $h \in R_C(x)$, where the radial cone $R_C(x) \subset X$ (“to the convex set C at the point x ”) is defined as

$$(3) \quad \mathcal{R}_C(x) := \{h \in X : x + th \in C \text{ for some } t > 0\}.$$

(We remark passing by that by convexity of C , if there exists $t_0 > 0$ such that $x + t_0h \in C$ then $x + th \in C$ for all $0 \leq t \leq t_0$.) Now, let us assume that x is optimal, and let $h \in \mathcal{R}_C(x)$. Then, there exists a small enough $t > 0$ such that $x + th \in C$. By optimality of x (and the fact that $Lx < \infty$), we have $L(x + th) \geq L(x)$ and thus $L(h) \geq 0$ by linearity of L . The other direction is also straightforward. Indeed, assume that $L(h) \geq 0$ for all $h \in \mathcal{R}_C(x)$. Then for all $y \in C$, we have $L(y) = L(x) + L(y - x) \geq L(x)$ since $y - x \in \mathcal{R}_C(x)$, and thus x is optimal for the problem (2).

Let us now characterise the radial cone \mathcal{R}_π defined above in (1) explicitly. We have:

Lemma 1 (Characterisation of the radial cone \mathcal{R}_π). *Let any $\pi \in \Pi(\mu, \nu)$. The radial cone \mathcal{R}_π to the set of transport plan $\Pi(\mu, \nu)$ at the point π — as defined in (1) — is characterised as the set of signed Radon measure $\eta = \eta^+ - \eta^-$ such that η^+ and η^- share common marginals, that is $\text{pr}_i^\# \eta^+ = \text{pr}_i^\# \eta^-$ for $i = 1, 2$, and such that $\eta^- \ll \pi$ and $\frac{d\eta^-}{d\pi} \in L^\infty(\pi)$.*

Here, we write η^- the negative part of η and correspondingly η^+ its non-negative part. We use the standard notation $\eta^- \ll \pi$ to mean that η^- is absolutely continuous with respect to π , and we write $\frac{d\eta^-}{d\pi}$ the Radon-Nykodym density of η^- with respect to π . We remark that, in Lemma 1, the fact that η^+ and η^- share common marginals also rephrases as $\text{pr}_i^\# \eta = 0$ for $i = 1, 2$, that is the (signed) measure η has *zero* marginals.

Proof of Lemma 1. Let $\mathcal{R}'_\pi \subset \mathcal{M}(X \times Y)$ be the set

$$(4) \quad \mathcal{R}'_\pi = \left\{ \eta \in \mathcal{M}(X \times Y) \left| \text{pr}_i^\# \eta = 0 \text{ for } i = 1, 2, \quad \frac{d\eta^-}{d\pi} \in L^\infty(\pi) \right. \right\}.$$

— where it is implicitly understood that $\eta^- \ll \pi$. Let us show that $\mathcal{R}_\pi = \mathcal{R}'_\pi$, which is exactly the thesis of Lemma 1. Let $\eta \in \mathcal{R}'_\pi$ and let us define $\pi_t := \pi + t\eta$. We need to show that for a small enough $t > 0$, the measure π_t is in the set of transport plans $\Pi(\mu, \nu)$. This means showing that π_t is positive and that it verifies the marginal constraints — note that, if we show these two properties hold, it will immediately follow that this measure is in fact a *probability* measure. Let us decompose $\eta := \eta^+ - \eta^-$. By assumption, the negative part η^- is absolutely continuous with respect to π and its Radon-Nykodym density is in $L^\infty(\pi)$. Therefore, π_t is positive for all $0 \leq t \leq t_0$ where we define $t_0^{-1} := \|\frac{d\eta^-}{d\pi}\|_{L^\infty(\pi)} > 0$. Now, the fact that π_t verifies the marginal constraints is immediate from the fact that $\text{pr}_i^\# \eta = 0$ for all $i = 1, 2$. This shows that $\mathcal{R}'_\pi \subset \mathcal{R}_\pi$. We now prove the converse inclusion. Let $\eta \in \mathcal{R}_\pi$. Then, there exists a (small enough) $t > 0$ such that $\pi_t \in \Pi$, where we let π_t as above. This (again) immediately implies that $\text{pr}_i^\# \eta = 0$ for all $i = 1, 2$. Furthermore, because π_t must be a positive measure, it must be that $t\eta^-(A) \leq \pi(A)$ for all Borel sets $A \subset X \times Y$, which means that η^- is absolutely continuous with respect to π and that its Radon-Nykodym density is uniformly bounded with respect to π — and moreover $\|\frac{d\eta^-}{d\pi}\|_{L^\infty(\pi)} \leq t^{-1}$. Hence $\mathcal{R}_\pi = \mathcal{R}'_\pi$ and the thesis of Lemma 1 follows. \square

2.3. On c -cyclical monotonicity and its related notions. In this section, let us recall what c -cyclical monotonicity is all about and how it can be related to the perturbative approach outlined above. First, let us recall the mere definition of c -cyclical monotonicity:

Definition 1. A set $\Gamma \subset X \times Y$ is said to be *c -cyclically monotone* if for all couples of points $(x_1, y_1), \dots, (x_n, y_n)$ in Γ — where $n \in \mathbb{N}$ is arbitrary (but finite) — we have

$$(5) \quad \sum_{i=1}^n c(x_i, y_i) \leq \sum_{i=1}^n c(x_i, y_{i+1}),$$

where we let $y_{n+1} := y_1$. A transport plan $\pi \in \Pi(\mu, \nu)$ is said to be *c -cyclically monotone* if it is *concentrated* on a c -cyclically monotone set, that is if there exists $\Gamma \subset X \times Y$ which is c -cyclically monotone such that $\pi(\Gamma) = 1$.

It is known that a property equivalent to that of c -cyclical monotonicity is the following:

Definition 2. A transport plan $\pi \in \Pi(\mu, \nu)$ is said to be *finitely optimal* if for all couples of positive measures α and $\alpha' \in \mathcal{M}(X \times Y)$ with *finite* supports sharing common marginals — that is $\text{pr}_i^\# \alpha = \text{pr}_i^\# \alpha'$ for $i = 1, 2$ — and such that the support of α is contained in that of π , it holds that $\mathcal{C}(\alpha) \leq \mathcal{C}(\alpha')$.

The fact that this property¹ implies c -cyclical monotonicity is rather straightforward. Indeed, let us assume that $\pi \in \Pi(\mu, \nu)$ is finitely optimal, and let $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$ for $n \in \mathbb{N}$ where $\Gamma := \text{supp}(\pi)$. Then, let us define the positive and finitely supported measures $\alpha := \sum_{i=1}^n \delta_{(x_i, y_i)}$ and $\alpha' := \sum_{i=1}^n \delta_{(x_i, y_{i+1})}$ where we let $y_{n+1} := y_1$. By construction, we have $\text{supp}(\alpha) \subset \Gamma$. It is also evident that α and α' share the same marginals — since $y_{n+1} = y_1$. Therefore, finite optimality says that $\mathcal{C}(\alpha) \leq \mathcal{C}(\alpha')$. But $\mathcal{C}(\alpha) = \sum_{i=1}^n c(x_i, y_i)$ and $\mathcal{C}(\alpha') = \sum_{i=1}^n c(x_i, y_{i+1})$ so that we exactly recover the c -cyclical monotonicity property. The proof of the converse statement, namely that c -cyclical monotonicity is equivalent to the finite optimality property is to be found in [18, Prop. 1.6] or yet [14] — see also Remark 2.

Finite optimality of Definition 2 and the *zeroth-order optimality condition* shown in Proposition 1 are obviously resembling. In the case of finite optimality, if we let $\eta := \alpha' - \alpha$ where α and α' are as in Definition 2, then we see that η is a specific kind of perturbation of the constraint set $\Pi(\mu, \nu)$ at the point π . Indeed, since α and α' share common marginals, the measure $\pi + \eta$ is a “coupling” of μ and ν , in the sense that its marginals are precisely μ and ν . Nevertheless, we stress out that it need *not* be a rightful element of the radial cone \mathcal{R}_π , because α is purely atomic and therefore has no reason *a priori* to be absolutely continuous with respect to π . Another way to express this is by the fact that $\pi + \eta$ may not be a *positive* measure in this situation.

On the other side, in terms of the vocabulary of [5], this means that the elements of the radial cone $\eta \in \mathcal{R}_\pi$ are precisely those (signed) measure such that $\eta = \eta^+ - \eta^-$ and η^+ is a **competitor** for η^- . In essence, η^+ represents a re-routing of some amount of mass taken from the transport plan π , which is understood from the fact that η^- is absolutely continuous with respect to π , that is $\eta^- \ll \pi$. Again, the difference here with the notion of competitors of [5] is that it is demanded here that η^- be absolutely continuous, whereas in the case of finite optimality, it is demanded that η^- be finitely supported inside the support of π .

Definition 3 (Finite perturbations). Let us define $\mathcal{F}_\pi \subset \mathcal{M}(X \times Y)$ the set of perturbations pertaining to finite optimality, namely the set of (signed) purely atomic measures η with *finite* support — meaning that $\text{supp}(|\eta|)$ is a finite set where we recall that $|\eta| = \eta^+ + \eta^-$ — such that $\text{pr}_i^\# \eta = 0$ for $i = 1, 2$ and such that $\text{supp}(\eta^-) \subset \text{supp}(\pi)$. In the language of [5], this means that η^+ is a competitor to η^- . We will refer to elements of \mathcal{F}_π as *finite perturbations*.

¹This notion goes back to at least the work of Beiglböck and Griessler [5] — and [6] in the case of the martingale optimal transport setting. Nevertheless, the journal version (& first preprinted version) of [5] refers the reader to [20, Ex. 2.21] — whereas the second preprinted version of same article also refers to [1] when introducing this notion. Bottom line, finite optimality has probably been around as *folkloric knowledge* for quite a while.

By construction, it is obvious that ***finite optimality as in Definition 2 is equivalent to demanding that $\mathcal{C}(\eta) \geq 0$ for all $\eta \in \mathcal{F}_\pi$*** . We can also define the subset \mathcal{C}_π of \mathcal{F}_π of *cyclical perturbations*, to wit the set of (signed) purely atomic measures η of the form

$$\eta = \frac{1}{n} \sum_{i=1}^n (\delta_{(x_i, y_{i+1})} - \delta_{(x_i, y_i)}).$$

for some $n \in \mathbb{N}$ and family of points $(x_1, y_1), \dots, (x_n, y_n) \in \text{supp}(\pi)$ with $y_{n+1} := y_1$. It is then obvious that the ***c-cyclical monotonicity property is exactly equivalent to the fact that $\mathcal{C}(\eta) \geq 0$ for all $\eta \in \mathcal{C}_\pi$*** .

Remark 2 (Equivalence between finite optimality and *c-cyclical monotonicity*). *We explained above that finite optimality implies c-cyclical monotonicity. Within our language, this trivially follows from the fact that \mathcal{C}_π is a subset of \mathcal{F}_π . The reverse statement also holds, and we refer the reader to [14, Appendix D]. We stress out that this necessitates no assumptions whatsoever on the cost of transport c . In fact, the rationale found in [14] relies on purely algebraic manipulations, and in fact it can be stated abstractly as follows: If $L : \mathcal{M}(X \times Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ is a (not necessarily continuous) linear functional, and if $L(\eta) \geq 0$ for all cyclical perturbations $\eta \in \mathcal{C}_\pi$ — provided $L\eta^- < \infty$ so that the quantity $L(\eta)$ makes sense as an element of $\mathbb{R} \cup \{+\infty\}$ — then $L(\eta) \geq 0$ for all finite perturbations $\eta \in \mathcal{F}_\pi$ — again, so that $L(\eta^-) < \infty$. Here, in our present context, the linear function L is simply given by \mathcal{C} the cost functional of (OT).*

The game will be to show that (under suitable assumptions) ***the cost functional \mathcal{C} being non-negative over the radial cone \mathcal{R}_π is equivalent to it being non-negative over the set of finite perturbations \mathcal{F}_π or equivalently on the set of cyclical perturbations \mathcal{C}_π*** . We will first show (under suitable assumptions) that *c-cyclical monotonicity is necessary for optimality in Section 2.4 (“If \mathcal{C} is non-negative over \mathcal{R}_π , then it is non-negative over \mathcal{C}_π (or equivalently over \mathcal{F}_π)”)* then prove that it is sufficient in Section 2.5 (“If \mathcal{C} is non-negative over \mathcal{C}_π (or equivalently over \mathcal{F}_π), then it is non-negative over \mathcal{R}_π ”). In fact, we will show every element of the radial cone \mathcal{R}_π is an accumulation point of \mathcal{F}_π (see Proposition 3 *infra*). This last property is actually somehow completely agnostic to the cost of transport, but allows to retrieve the equivalence between optimality and *c-cyclical monotonicity* under appropriate continuity assumptions on the cost of transport c (see Proposition 4 *infra*).

2.4. Optimality implies *c-cyclical monotonicity* (\Rightarrow). We start by showing in our framework that optimality implies *c-cyclical monotonicity*. The most general result in this direction is that of Beiglböck et al. [7] under the sole assumption that the cost of transport $c : X \times Y \rightarrow [0, \infty]$ be Borel-measurable. Here, we revisit the proof that optimality implies *c-cyclical*

monotonicity under the slightly stronger assumption that c is integrable with respect to $\mu \otimes \nu$:

Theorem 1. *Let $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ be Borel-measurable so that c is (locally) integrable with respect to $\mu \otimes \nu$. If $\pi \in \Pi(\mu, \nu)$ is an optimal transport plan with $\mathcal{C}(\pi) < \infty$, then π is c -cyclically monotone.*

We stress out that c is allowed to assume $+\infty$ as value (and in fact we do not demand that c be either non-negative nor bounded from below at this stage). In the following proposition, we first prove Theorem 1 in the special case where the marginal spaces are Euclidean, that is $X = \mathbb{R}^d$ and $Y = \mathbb{R}^m$ for $d, m \in \mathbb{N}$. This assumption will be removed afterwards.

Proposition 2. *Let $c : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be Borel-measurable so that c is (locally) integrable with respect to $\mu \otimes \nu$. If $\pi \in \Pi(\mu, \nu)$ is an optimal transport plan such that $\mathcal{C}(\pi) < \infty$ then π is c -cyclical monotone.*

Proof of Proposition 2. Let us consider a (finite) sequence of points $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$ for a set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^m$ such that $\pi(\Gamma) = 1$. The precise definition of the set Γ is to be given at the end of the proof. Here, recall that we let $y_{n+1} := y_1$. Let $\varepsilon > 0$ and let $B_\varepsilon^i \subset \mathbb{R}^d \times \mathbb{R}^m$ be the ball of radius ε centered at (x_i, y_i) . Again, let $B_\varepsilon^{n+1} := B_\varepsilon^1$. Up to refining the set Γ , we may (and will) assume without loss of generality that $\pi(B_\varepsilon^i) > 0$ for all $\varepsilon > 0$ and all $i = 1, \dots, n$. Let us then define σ_i^ε to be the restriction of π to the ball B_ε^i normalised so as to make it a probability measure, that is $\sigma_i^\varepsilon = \pi(B_\varepsilon^i)^{-1} \pi \upharpoonright_{B_\varepsilon^i}$. We then let

$$\sigma_\varepsilon := \sum_{i=1}^n \sigma_i^\varepsilon.$$

We note that $\sigma_\varepsilon \ll \pi$ and that the Radon-Nykodym derivative of σ_ε with respect to π is uniformly bounded since $\pi(B_\varepsilon^i) > 0$ for all $i = 1, \dots, n$. More precisely, we have $\|\frac{d\sigma_\varepsilon}{d\pi}\|_{L^\infty(\pi)}^{-1} = \min_{i=1, \dots, n} \pi(B_\varepsilon^i)$. Now, we introduce $\xi_i^\varepsilon := \text{pr}_1^\# \sigma_i^\varepsilon \otimes \text{pr}_2^\# \sigma_{i+1}^\varepsilon$ for all $i = 1, \dots, n$, where again we let $\sigma_{n+1}^\varepsilon := \sigma_1^\varepsilon$ and let us define

$$\xi_\varepsilon := \sum_{i=1}^n \xi_i^\varepsilon.$$

By construction σ_ε and ξ_ε share common marginals. Therefore $\eta_\varepsilon := \xi_\varepsilon - \sigma_\varepsilon \in \mathcal{R}_\pi$ is an admissible perturbation at π — since $\pi + t(\xi_\varepsilon - \sigma_\varepsilon) \in \Pi(\mu, \nu)$ for all $0 \leq t < \|\frac{d\sigma_\varepsilon}{d\pi}\|_{L^\infty(\pi)}^{-1}$ — and by optimality (and finiteness) of π it must be that $\mathcal{C}(\xi_\varepsilon) \geq \mathcal{C}(\sigma_\varepsilon)$. Let us show that $\mathcal{C}(\sigma_\varepsilon) \rightarrow \sum_{i=1}^n c(x_i, y_i)$ and $\mathcal{C}(\xi_\varepsilon) \rightarrow \sum_{i=1}^n c(x_i, y_{i+1})$ as $\varepsilon \rightarrow 0$, so that the thesis of Proposition 2 will follow.

We now recall that by virtue of *Lebesgue differentiation theorem* for π -a.s. all point $(x_0, y_0) \in \mathbb{R}^d \times \mathbb{R}^m$ one has

$$(6) \quad \frac{1}{\pi(B_\varepsilon)} \int_{B_\varepsilon} c(x, y) \pi(dx, dy) \xrightarrow{\varepsilon \rightarrow 0} c(x_0, y_0)$$

where B_ε is the ball of radius ε centred at (x_0, y_0) . This theorem is valid for any Radon measure on Euclidean spaces, see *e.g.* [8, Theorem 8.4.6]. This is veracious since π has finite transport cost, meaning that $c \in L^1(\pi)$. Let then $\Gamma_0 \subset \mathbb{R}^d \times \mathbb{R}^m$ be the set of points $(x_0, y_0) \in \mathbb{R}^d \times \mathbb{R}^m$ such that (6) holds — which has therefore full π -measure. Then — provided $(x_i, y_i) \in \Gamma_0$ for all $i = 1, \dots, n$ — we have

$$\begin{aligned} \mathcal{C}(\sigma_\varepsilon^i) &= \int_{\mathbb{R}^d \times \mathbb{R}^m} c(x, y) \sigma_\varepsilon^i(dx, dy) \\ &= \frac{1}{\pi(B_\varepsilon^i)} \int_{B_\varepsilon^i} c(x, y) \pi(dx, dy) \xrightarrow{\varepsilon \rightarrow 0} c(x_i, y_i). \end{aligned}$$

We have therefore obtain the sought-for claim that

$$\mathcal{C}(\sigma_\varepsilon) = \sum_{i=1}^n \mathcal{C}(\sigma_\varepsilon^i) \rightarrow \sum_{i=1}^n c(x_i, y_i)$$

as $\varepsilon \rightarrow 0$. It remains to take care of the other limit.

We have

$$\begin{aligned} \mathcal{C}(\xi_\varepsilon^i) &= \int_{\mathbb{R}^d \times \mathbb{R}^m} c(x, y) \text{pr}_1^\# \sigma_\varepsilon^i \otimes \text{pr}_2^\# \sigma_{\varepsilon^{i+1}}(dx, dy) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^m} c(x, y) d \left(\frac{1}{\pi(B_\varepsilon^i)} \int_{\mathbb{R}^m} \pi|_{B_\varepsilon^i}(x, dy') \right) \otimes \left(\frac{1}{\pi(B_\varepsilon^{i+1})} \int_{\mathbb{R}^d} \pi|_{B_\varepsilon^{i+1}}(dx', y) \right) \\ &= \frac{1}{(\pi \otimes \pi)(B_\varepsilon^i \times B_\varepsilon^{i+1})} \int_{B_\varepsilon^i \times B_\varepsilon^{i+1}} c(x, y) \pi \otimes \pi(dx, dy', dx', dy). \end{aligned}$$

We can then appeal again to Lebesgue differentiation theorem, applied this time to the probability measure $\pi \otimes \pi \in \mathcal{P}((\mathbb{R}^d \times \mathbb{R}^m) \times (\mathbb{R}^d \times \mathbb{R}^m))$ and for the integrand c seen as the function $c(x, y', x', y) := c(x, y)$. Note that this is righteous since c is assumed to be (locally) integrable with respect to $\mu \otimes \nu$. For all $i = 1, \dots, n$, we have $(x_i, y_{i+1}) \in B_\varepsilon^i \times B_\varepsilon^{i+1}$ and letting $\varepsilon \rightarrow 0$ yields that $\mathcal{C}(\xi_\varepsilon^i) \rightarrow c(x_i, y_{i+1})$, and therefore the claim that $\mathcal{C}(\xi_\varepsilon) = \sum_{i=1}^n \mathcal{C}(\xi_\varepsilon^i) \rightarrow \sum_{i=1}^n c(x_i, y_{i+1})$. This terminates (in essence) the proof of Proposition 2.

To finish, let us discuss how to construct the set Γ introduced at the beginning of this proof. Let $\Gamma_1 := \bigcap_{i=1,2} \text{pr}_i(\Sigma)$ where $\Sigma \subset (\mathbb{R}^d \times \mathbb{R}^m) \times (\mathbb{R}^d \times \mathbb{R}^m)$ is the set of points for which Lebesgue differentiation theorem holds this time for $\pi \otimes \pi$ and where evidently we define $\text{pr}_1(x, y, z, t) = (x, y)$ and $\text{pr}_2(x, y, z, t) = (z, t)$ for all $(x, y, z, t) \in (\mathbb{R}^d \times \mathbb{R}^m) \times (\mathbb{R}^d \times \mathbb{R}^m)$. It is well-known that the sets $\text{pr}_1(\Sigma)$ and $\text{pr}_2(\Sigma)$ need not be Borel-measurable, but that they are certainly Lebesgue-measurable [13, Theorem 2.12]. It then suffices to consider $\Gamma := \Gamma_0 \cap \Gamma_1$ where Γ_0 was introduced above as the (Borel-measurable) set of points for which Lebesgue differentiation theorem holds (6) in the case of π . Note that Γ is *a priori* only Lebesgue-measurable. Nonetheless, as such there exists a Borel set $\Gamma' \subset \mathbb{R}^d \times \mathbb{R}^m$ and a π -null set $N \subset \mathbb{R}^d \times \mathbb{R}^m$ so that $\Gamma = \Gamma' \cup N$. Then Γ' is the sought-for c -cyclically monotone Borel-measurable set over which π concentrates. \square

In the preceding proposition, we have assumed that the marginal spaces where Euclidean so as to appeal to the standard Lebesgue differentiation theorem, see [8, Theorem 8.4.6]. In the *non*-Euclidean setting, this theorem is still much valid but requires additional assumptions either on the underlying space or on the probability measure under consideration. Typically, one demands that this measure be *doubling*, see *e.g.* [16, Chapter 2.9]. But this is completely inadequate, since there are no reasons for a (optimal) transport plan to be doubling — and in fact it is easy to build counterexamples. The gap between the Euclidean and non-Euclidean setting in the Lebesgue differentiation theorem boils down to the choice of the *differentiation basis* — *i.e.* whether or not we can consider “nice” shrinking sets like balls or cubes. But for what matters to us, this is rather unimportant. More precisely, we can attach to each point a sequence of measurable sets specifically tailored to this point so that the Lebesgue differentiation theorem holds:

Lemma 2. *Let $\gamma \in \mathcal{P}(Z)$ where Z is a Polish space. Then, for γ -a.s all $z \in Z$, there exists a sequence $E_k(z)$ of Borel sets all containing z such that $\gamma(E_k(z)) > 0$ for all $k \in \mathbb{N}$ and such that for any function $\phi : Z \rightarrow \mathbb{R}$ integrable with respect to γ , we have*

$$(7) \quad \frac{1}{\gamma(E_k(z))} \int_{E_k(z)} \phi(w) \gamma(dw) \xrightarrow[k \rightarrow \infty]{} \phi(z)$$

for γ almost-surely all $z \in Z$.

The proof of the technical Lemma 2 is postponed to Appendix A. We can prove Theorem 1 in full generality using this lemma:

Proof of Theorem 1. The proof is identical to that of Proposition 2, but this time we need to care for which kind of neighbourhood we are restricting π onto. Let $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$, where Γ is a set such that $\pi(\Gamma) = 1$ that will be constructed again *a posteriori*. For each $z \in X \times Y$, let $E_k(z)$ be the Borel sets of Lemma 2. Let us then write $E_k^i = E_k(z_i)$ where $z_i := (x_i, y_i)$. Let then π_i^k be the normalised restriction of π to E_k^i and $\sigma_k := \sum_{i=1}^n \sigma_i^k$. Then, we let $\xi_i^k := \text{pr}_1^\# \sigma_i^k \otimes \text{pr}_2^\# \sigma_{i+1}^k$ for all $i = 1, \dots, n$ (where $\sigma_{n+1}^k := \sigma_1^k$) and let $\xi_k := \sum_{i=1}^n \xi_i^k$. In virtue of Lemma 2, just as in the proof Proposition 2, we have $\mathcal{C}(\sigma_k) \rightarrow \sum_{i=1}^n c(x_i, y_i)$ as $k \rightarrow \infty$. Also, again using Fubini’s theorem in the same spirit as in the proof of Proposition 2, write

$$(8) \quad \mathcal{C}(\xi_i^k) = \frac{1}{\pi \otimes \pi(E_k^i \times E_{k+1}^i)} \int_{E_k^i \times E_{k+1}^i} c(x, y) \pi \otimes \pi(dx, dy', dx', dy)$$

Now, the thesis of Lemma 2 still holds here — we refer to Remark 7 in the Appendix. In particular, we have $\mathcal{C}(\xi_i^k) \rightarrow c(x_i, y_{i+1})$ for all $i = 1, \dots, n$ and therefore $\mathcal{C}(\xi_k) \rightarrow \sum_{i=1}^n c(x_i, y_{i+1})$. This finishes the (core of the) proof. The set Γ is built exactly as in the proof of Proposition 2, so we do not repeat the argument here. \square

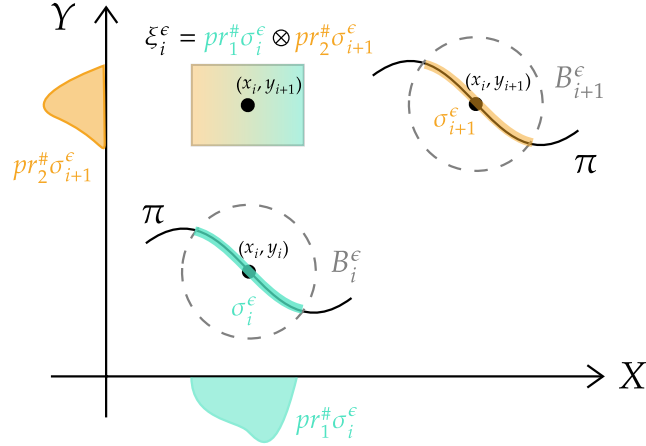


FIGURE 1. Schematic explanation of the construction of the measures σ_i^ϵ and ξ_i^ϵ ($i = 1, \dots, n$) in the proof of Proposition 2.

2.5. c -cyclical monotonicity implies optimality (\Leftarrow). We will now discuss the reverse implication, namely that c -cyclical monotonicity is a sufficient criterion to yield optimality under suitable assumptions on the cost c . The idea is as follows: provided the c -cyclical monotonicity property holds, that is $\mathcal{C}(\eta) \geq 0$ for all $\eta \in \mathcal{C}_\pi$ — or equivalently for all $\eta \in \mathcal{F}_\pi$ — the game is to show that this implies that $\mathcal{C}(\eta) \geq 0$ for all $\eta \in \mathcal{R}_\pi$, thus implying that π is optimal according Proposition 1. In fact, we will first prove the following statement:

Proposition 3. *For all $\pi \in \Pi(\mu, \nu)$, any element of the radial cone \mathcal{R}_π is an accumulation point of the set of finite perturbations \mathcal{F}_π .*

Under the assumption that \mathcal{C} is continuous — i.e. when $c \in C_0(X \times Y)$ — then it follows immediately from the above proposition that c -cyclical monotonicity implies optimality:

Proposition 4. *Let $c : X \times Y \rightarrow \mathbb{R}$ be in $C_0(X \times Y)$, i.e. continuous and vanishing at infinity (or yet that c is continuous and the marginals μ and ν are compactly-supported). If $\pi \in \Pi(\mu, \nu)$ is c -cyclically monotone and $\mathcal{C}(\pi) < \infty$, then π is optimal.*

Proof of Proposition 4. We need that show that $\mathcal{C}(\eta) \geq 0$ for all $\eta \in \mathcal{R}_\pi$. But for any admissible perturbation $\eta \in \mathcal{R}_\pi$, according to Proposition 3, we can find $(\chi_n)_n \subset \mathcal{F}_\pi$ such that $\chi_n \rightarrow \eta$ in the limit $n \rightarrow \infty$. By c -cyclical monotonicity, we have $\mathcal{C}(\chi_n) \geq 0$ for all $n \in \mathbb{N}$. By the assumptions on the cost of transport c , the transport cost functional $\mathcal{C} : \mathcal{P}(X \times Y) \rightarrow \mathbb{R}$ is vaguely continuous, which entails that $\mathcal{C}(\eta) = \lim_{n \rightarrow \infty} \mathcal{C}(\chi_n) \geq 0$ so that the thesis is proved. \square

It is interesting to remark that the result of Proposition 3 is in a sense completely disconnected from the notion of c -cyclical monotonicity *per se*, since it has nothing to do with the cost of transport c . The c -cyclical monotonicity is certainly implied by it under continuity assumption on the cost of transport as shown above in Proposition 4. We will show afterward how to obtain sufficiency of the c -cyclical monotonicity property for optimality without the stringent continuity assumptions on the cost of transport of Proposition 4. First, we prove Proposition 3 by constructing explicitly for each $\eta \in \mathcal{R}_\pi$ a sequence of finite perturbations vaguely converging to η .

2.5.1. Proof of Proposition 3. We fix an admissible perturbation $\eta \in \mathcal{R}_\pi$. We remark that we may (and will) assume without loss of generality that η^+ and η^- are normalised so that they are probability measures, that is $\int_{X \times Y} \eta^+ = \int_{X \times Y} \eta^- = 1$. By definition η^+ and η^- share common marginals, that is $\text{pr}_i^\# \eta^+ = \text{pr}_i^\# \eta^- := m_i$ for $i = 1, 2$. We can then write $\eta^- = \eta_y^-(x) \otimes m_2(y)$ the disintegration of η^- onto its second marginal m_2 , and likewise $\eta^+ = m_1(x) \otimes \eta_x^+(y)$ the disintegration of η^+ onto its first marginal m_1 . We are going to build samples Z_0^-, \dots, Z_{n-1}^- — resp. Z_0^+, \dots, Z_{n-1}^+ — from the probability measure η^- — resp. η^+ — such that **(i)** the associated empirical measures

$$\eta_n^- := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{Z_i^-} \quad \text{and} \quad \eta_n^+ := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{Z_i^+}$$

vaguely converge to respectively η^- and η^+ and **(ii)** such that these empirical measures share common marginals. Then, $\eta_n := \eta_n^+ - \eta_n^-$ will be a rightful finite perturbation at π , *i.e.* an element of \mathcal{F}_π , and $\eta_n \rightarrow \eta$.

Let us now explain how to build such samples. First, assume that $Z_0^- := (x_0, y_0)$ is drawn according to the probability measure η^- . Now, draw $y_1 \in Y$ according to $\eta_{x_0}^+$ and let $Z_0^+ := (x_0, y_1)$. Then, to build Z_1^- , we draw $x_1 \in X$ according to $\eta_{y_1}^-$ and let $Z_1^- := (x_1, y_1)$. This procedure can be further continued, yielding two sequences of points $(Z_n^+)_{n \in \mathbb{N}}$ and $(Z_n^-)_{n \in \mathbb{N}}$ both inside the Cartesian product space $X \times Y$. The associated empirical measures η_n^- and η_n^+ are such that their first marginals always coincide, *i.e.* $\text{pr}_1^\# \eta_n^- = \text{pr}_1^\# \eta_n^+$ — or otherwise stated $\text{pr}_1^\# \eta_n = 0$. Nevertheless, we have $\text{pr}_2^\# \eta_n^+ = \text{pr}_2^\# \eta_n^- - \frac{1}{n} \delta_{y_0} + \frac{1}{n} \delta_{y_n}$ so that these two marginals do not *a priori* coincide except in the eventuality that $y_0 = y_n$. We therefore need to correct this sampling procedure so as to ensure that the second marginals also coincide.

Before that, in all due respect, we need to put some rigour on the above construction. Let $\Omega := \Omega^- \times \Omega^+$ where $\Omega^- = \Omega^+ := X \times Y$, each of these spaces being equipped with the corresponding product σ -algebras. We let $\mathbf{Z}_n := (Z_n^-, Z_n^+) \in \Omega_- \times \Omega_+$. Then, the sequence $(\mathbf{Z}_n)_n$ is a time-homogeneous Markov chain with canonical paths space $\Omega^\mathbb{N}$ — equipped again with its standard product σ -algebra. Let us remark that for all $n \in \mathbb{N}$ we have $\text{Law}(Z_n^-) = \eta^-$ and $\text{Law}(Z_n^+) = \eta^+$. Indeed, let $f : X \times Y \rightarrow \mathbb{R}$ be

a test function, say continuous and bounded. We have

$$\mathbb{E}[f(Z_1^-)|Z_0^- = (x_0, y_0)] = \int_{X \times Y} f(x, y) \eta_{x_0}^+(dy) \otimes \eta_y^-(dx).$$

Therefore

$$\mathbb{E}[f(Z_1^-)] = \int_{X \times Y} \left(\int_{X \times Y} f(x, y) \eta_{x_0}^+(dy) \otimes \eta_y^-(dx) \right) \eta^-(dx_0, dy_0)$$

where we used that by construction $\mathbf{Law}(Z_0^-) = \eta^-$. Let us then integrate over the variable $y_0 \in Y$, so that

$$\mathbb{E}[f(Z_1^+)] = \int_X \left(\int_{X \times Y} f(x, y) \eta_{x_0}^+(dy) \otimes \eta_y^-(dx) \right) m_1(dx_0)$$

Then, let us integrate over $x_0 \in X$ and then over $y \in Y$

$$\begin{aligned} \mathbb{E}[f(Z_1^-)] &= \int_{X \times Y} f(x, y) \eta_y^-(dx) \otimes \left(\int_X \eta_{x_0}^+ m_1(dx_0) \right) (dy) \\ (9) \quad &= \int_{X \times Y} f(x, y) \eta_y^-(dx) \otimes m_2(dy) \end{aligned}$$

$$(10) \quad = \int_{X \times Y} f(x, y) \eta^-(dx, dy)$$

where we used disintegration of η^+ — resp. reintegration of η^- — in (9) and (10). Therefore, $\mathbf{Law}(Z_1^-) = \eta^-$, and by an immediate induction $\mathbf{Law}(Z_n^-) = \eta^-$ for all $n \in \mathbb{N}$. Using the same rationale, one shows that $\mathbf{Law}(Z_n^+) = \eta^+$ for all $n \in \mathbb{N}$.

We stress out that the empirical measures η_n^- and η_n^+ are **random measures** $\eta_n^- = \eta_n^-(\omega)$ and $\eta_n^+ = \eta_n^+(\omega)$ depending on the path $\omega \in \Omega^{\mathbb{N}}$. As noted above, we need to correct this sampling procedure so that the second marginals of these empirical measures coincide. Let us define

$$\chi_n^- := \eta_n^- + \frac{1}{n} \delta_{(\bar{x}, \bar{y})} + \frac{1}{n} \delta_{(x_n, y_n)}, \quad \chi_n^+ := \eta_n^+ + \frac{1}{n} \delta_{(\bar{x}, y_0)} + \frac{1}{n} \delta_{(x_n, \bar{y})}$$

where we choose any $(\bar{x}, \bar{y}) \in \text{supp}(\eta^-)$ here. What we have done here is to add *ad hoc* corrections to the empirical measures η_n^- and η_n^+ so that now the resulting measure χ_n^- and χ_n^+ verify that both of their marginals coincide, i.e. $\text{pr}_i^\# \chi_n^- = \text{pr}_i^\# \chi_n^+$ for all $n \in \mathbb{N}$ and $i = 1, 2$. Furthermore, we duly emphasise that $\text{supp}(\chi_n^-) = \{(x_i, y_i) : i = 0, \dots, n\} \cup \{(\bar{x}, \bar{y})\}$ is contained in $\text{supp}(\eta^-)$, which is itself contained in the support of π since $\eta \in \mathcal{R}_\pi$. Therefore, letting $\chi_n := \chi_n^+ - \chi_n^-$, we readily see that χ_n is a rightful finite perturbation at π , i.e. $\chi_n \in \mathcal{F}_\pi$. Now, let us rewrite $\chi_n^- = \eta_n^- + \varepsilon_n^-$ where $\varepsilon_n^- := \frac{1}{n} \delta_{(\bar{x}, \bar{y})} + \frac{1}{n} \delta_{(x_n, y_n)}$ and likewise $\chi_n^+ = \eta_n^+ + \varepsilon_n^+$ where $\varepsilon_n^+ := \frac{1}{n} \delta_{(\bar{x}, y_0)} + \frac{1}{n} \delta_{(x_n, \bar{y})}$. Since both ε_n^- and ε_n^+ converge vaguely to zero as $n \rightarrow \infty$, if we can prove that the empirical measure η_n^- (resp. η_n^+) converge vaguely to η^- (resp. η^+) then the corrected measure χ_n^- (resp. χ_n^+) will converge to η^- (resp. η^+), henceforth χ_n will converge vaguely to η . This would imply the thesis of Proposition 3 since $(\chi_n)_n$ would be a sequence of

finite perturbations approaching η , thus proving that every element of the radial cone is an accumulation point of finite perturbations.

Nevertheless, some care should be given here, since it need *not* be true that the empirical measures η_n^- (resp. η_n^+) samples consistently the probability measures η^- (resp. η^+) as $n \rightarrow \infty$. This is easily foreseen from the fact that the underlying Markov chain need not be *ergodic*, as shown by the next example.

Example 1. Assume that $X = Y = [0, 1]$. Let $S_\alpha(x) := x + \alpha \pmod{1}$ be a shift, where $\alpha \in \mathbb{R}$. Let η^- be the uniform distribution over the diagonal $\Delta := \{(x, x) : x \in [0, 1]\}$ of the square $[0, 1] \times [0, 1]$ and η^+ be the uniform distribution over the shifted diagonal $\Delta_\alpha := \{(x, S_\alpha(x)) : x \in [0, 1]\}$. Otherwise stated $\eta^- := (Id, S_0)^\# \lambda$ and $\eta^+ := (Id, S_\alpha)^\# \lambda$ where λ is the Lebesgue measure on $[0, 1]$. First, we remark that the initial condition $Z_0^- = (x_0, x_0)$ determines the values of the entire chain. More precisely, we have $Z_n^- = (S_\alpha^{(n)} x_0, S_\alpha^{(n)} x_0)$ and $Z_n^+ = (S_\alpha^{(n-1)} x_0, S_\alpha^{(n)} x_0)$. If $\alpha = \frac{n}{m} \in \mathbb{Q}$ is rational, then $T_\alpha^{(m)} = Id$ and therefore a single chain will never sample entirely neither η^+ nor η^- . On the contrary, it is well-known that if α is irrational the underlying Markov chain is ergodic, and incidently η^+ and η^- will be completely sampled from using a single chain. We note passing by that this example is closely related to a well-known counterexample in the literature to sufficiency of c -cyclical monotonicity to yield optimality, see *e.g.* [7, Example 1.2].

But the obstacle raised by the previous example can easily be bypassed for what matters to us — where all is fair game as long as we manage to sample from η^+ and η^- . Indeed, the idea to swerve away from this problem is simply to run *multiple* and *independent* copy of our constructions. That is, we run several independent Markov chains $(\mathbf{Z}_n^j)_n$ for $j = 1, \dots, M$ where $\mathbf{Z}_n^j := (Z_{n,j}^-, Z_{n,j}^+)$ each chain starting from independently drawn initial conditions $Z_{0,j}^- \sim \eta^-$ for all $j = 1, \dots, M$. For each chain, let $\eta_{n,j}^-$ and $\eta_{n,j}^+$ be the empirical measures, $\chi_{n,j}^-$ and $\chi_{n,j}^+$ be the corrected empirical measures and $\varepsilon_{n,j}^-$ (resp. $\varepsilon_{n,j}^+$) the said corrections. We then consider the *pooled* estimators

$$\underline{\eta}_{n,M}^- := \frac{1}{M} \sum_{j=1}^M \eta_{n,j}^-, \quad \underline{\chi}_{n,M}^- := \frac{1}{M} \sum_{j=1}^M \chi_{n,j}^-, \quad \underline{\varepsilon}_{n,M}^- := \frac{1}{M} \sum_{j=1}^M \varepsilon_{n,j}^-$$

and their obvious counterparts $\underline{\eta}_{n,M}^+$, $\underline{\chi}_{n,M}^+$ and $\underline{\varepsilon}_{n,M}^+$. We then let $\underline{\chi}_{n,M} := \underline{\chi}_{n,M}^+ - \underline{\chi}_{n,M}^-$. That is still an element of \mathcal{F}_π . Now, for *any* fixed $n \in \mathbb{N}$, it holds that the empirical measure $\underline{\eta}_{n,M}^-$ (resp. $\underline{\eta}_{n,M}^+$) vaguely converge to η^- (resp. η^+) from sheer independence of the chains in the limit $M \rightarrow \infty$ and since $\mathbf{Law}(Z_{n,j}^-) = \eta^-$ and $\mathbf{Law}(Z_{n,j}^+) = \eta^+$ for all $n \in \mathbb{N}$ and all $j \in \mathbb{N}$. On this matter, we yet remind the reader that both $\underline{\eta}_{n,M}^-$ and $\underline{\eta}_{n,M}^+$ are *random* measures depending on $\omega \in \Omega^{\mathbb{N}}$, so that this convergence actually holds for almost-surely all ω (with respect to the underlying path

measure). This is a classical result, see *Varadarajan's theorem* [15, Thm. 11.4.1]. We emphasise here that this result has essentially *nothing* to do with the underlying Markovian structure of our construction, and is in essence just the law of large numbers. Moreover, by the exact same rationale, it holds that $\underline{\varepsilon}_{n,M}^- \rightharpoonup e_n^-$ and $\underline{\varepsilon}_{n,M}^+ \rightharpoonup e_n^+$ as $M \rightarrow \infty$ where $e_n^- = \frac{1}{n}\delta_{(\bar{x},\bar{y})} + \frac{1}{n}\eta^-$ and $e_n^+ = \frac{1}{n}\delta_{\bar{x}} \otimes m_2 + \frac{1}{n}m_1 \otimes \delta_{\bar{y}}$.

To finish, let us rewrite $\underline{\chi}_{n,M} = \underline{\eta}_{n,M} - \underline{\varepsilon}_{n,M}$ where we let $\underline{\eta}_{n,M} := \underline{\eta}_{n,M}^+ - \underline{\eta}_{n,M}^-$ and likewise $\underline{\varepsilon}_{n,M} := \underline{\varepsilon}_{n,M}^+ - \underline{\varepsilon}_{n,M}^-$. The double-entry sequence $(\underline{\chi}_{n,M})_{n,M}$ verifies that $\underline{\chi}_{n,M} \rightharpoonup \eta^+ - \eta^- - e_n^+ + e_n^-$ as $M \rightarrow \infty$ and, since both e_n^- and e_n^+ vanish vaguely in the limit $n \rightarrow \infty$, we obtain that the double (vague) limit $\lim_{n \rightarrow \infty} \lim_{M \rightarrow \infty} \underline{\chi}_{n,M}$ is equal to $\eta^+ - \eta^-$, that is η . Finally, we remark that $\|\underline{\chi}_{n,M}\|_{TV} \leq 2(1 + 2/n) \leq 6$ for all $n, M \in \mathbb{N}$, where we recall that $\|\cdot\|_{TV}$ denotes the standard total variation norm. We also recall that finite radii balls of $\mathcal{M}(X \times Y)$ are metrisable for the vague topology. As in any metrisable space, we can perform a diagonal argument on the double-entry sequence $(\underline{\chi}_{n,M})_{n,M}$, so that there exists a sequence $(M_n)_n$ such that $\underline{\chi}_{n,M_n} \rightharpoonup \eta$. This sequence being in \mathcal{F}_π , we have proved that any admissible perturbation $\eta \in \mathcal{R}_\pi$ can be approximated by finite perturbations, that is elements of \mathcal{F}_π , thus yielding the thesis of Proposition 3.

Remark 3. *A small technical remark is to made here. We cannot work here with the narrow topology, that is in duality with continuous and bounded functions $C_b(X \times Y)$. Indeed, neither the space $\mathcal{M}(Z)$ of finite signed Radon measures is metrisable (except in the degenerate case where the base space Z is finite) nor its unit ball (except when the base space Z is compact, but in which case the narrow and vague topologies coincide). On this matter, see e.g. [19].*

2.5.2. *Sufficiency of c -cyclical monotonicity without continuity.* It is shown above in Proposition 4 that c -cyclical monotonicity is a sufficient criterion for optimality under continuity assumption on the cost of transport. This was true in virtue of Proposition 3, in which we proved that every admissible perturbation $\eta \in \mathcal{R}_\pi$ can be approximated, in the vague topology, by finite perturbations, *i.e.* elements of the set \mathcal{F}_π . In this section, we would like to show how to remove these assumptions.

Theorem 2. *Let $c : X \times Y \rightarrow [0, \infty]$ be Borel-measurable and assume that c is finite $\mu \otimes \nu$ -almost everywhere. Then, if $\pi \in \Pi(\mu, \nu)$ is c -cyclically monotone and $\mathcal{C}(\pi) < \infty$, then π is optimal for (OT).*

Remark 4. *We remark that the weakest theorem in this direction is that of [7], see Theorem 1.b there. It states sufficiency of cyclical monotonicity under the assumption that $\{c = \infty\} = F \cup N$ where F is any closed set and $\mu \otimes \nu(N) = 0$. Our theorem above does not recover a part of this assumption, namely the one embodied by the closed set F . It may be possible to do that, although we have not attempted it.*

Let us scrutinise more closely the construction built in the proof of Proposition 3. For the sake of completeness, we recall some elementary fact here on the ergodic theorem of Birkhoff and how to apply it in the context of Markov chains, see *e.g.* [17]. Let $T : \Omega^{\mathbb{N}} \rightarrow \Omega^{\mathbb{N}}$ be the shift operator $T(z_0, z_1, \dots) = (z_1, z_2, \dots)$, where we recall that $\Omega := \Omega^- \times \Omega^+$ where $\Omega^- = \Omega^+ = X \times Y$. According to what precedes, the Markov chain $(\mathbf{Z}_n)_n$ admits as a stationary measure the probability measure $\Xi \in \mathcal{P}(\Omega)$ defined as

$$\Xi(B) := \int_B \eta^-(dx^-, dy^-) \otimes (\delta_{x^-}(dx^+) \otimes \eta_{x^-}^+(dy^+))$$

for all Borel sets $B \in \Omega$. We let $\mathbb{P}_{\Xi} \in \mathcal{P}(\Omega^{\mathbb{N}})$ be the path probability measure of the chain with initial distribution $\Xi \in \mathcal{P}(\Omega)$. By stationarity of Ξ in the Markovian sense, the probability measure \mathbb{P}_{Ξ} is left invariant by the shift operator T , *i.e.* $T^{\#}\mathbb{P}_{\Xi} = \mathbb{P}_{\Xi}$. The celebrated theorem due to Birkhoff states that for any measurable function $f : \Omega^{\mathbb{N}} \rightarrow \mathbb{R}$ which is integrable with respect to \mathbb{P}_{Ξ} , it holds that

$$\frac{1}{n} \sum_{k=1}^n f \circ T^k(\omega) \rightarrow \mathbb{E}_{\mathbb{P}_{\Xi}}[f|\mathcal{J}](\omega)$$

for \mathbb{P}_{Ξ} almost-surely all $\omega \in \Omega^{\mathbb{N}}$ where $\mathbb{E}_{\mathbb{P}_{\Xi}}[f|\mathcal{J}]$ denotes the conditional expectation under \mathbb{P}_{Ξ} of the function f with respect to the σ -algebra consisting of all Borel sets left invariant by the shift, *i.e.* $\mathcal{J} := \{B \in \Omega^{\mathbb{N}} : T^{-1}(B) = B\}$. For what matters to us, we should also state that the same result applies in the case where f is not necessarily integrable with respect to \mathbb{P}_{Ξ} but where it is nevertheless (measurable and) *non-negative*, in which case the convergence almost-surely still holds but the conditional expectation may assume $+\infty$ as value. In particular, we can apply Birkhoff theorem to any function f of the form $g - h$ where g is measurable and non-negative and $h \in L^1(\mathbb{P}_{\Xi})$. In particular, let us consider the function $f_c : \Omega^{\mathbb{N}} \rightarrow \mathbb{R}$ defined as $f_c(\omega) := c(x_0^+, y_0^+) - c(x_0^-, y_0^-)$ where we write $\omega = (z_0, \dots) \in \Omega^{\mathbb{N}}$ with $z_0 = (z_0^-, z_0^+) \in \Omega$ and where we further write $z_0^- = (x_0^-, y_0^-) \in \Omega^-$ and $z_0^+ = (x_0^+, y_0^+) \in \Omega^+$. Here, $g(\omega) = c(z_0^+)$ which is non-negative and $h(\omega) = c(z_0^-)$ which is integrable with respect to \mathbb{P}_{Ξ} , since

$$\int_{\Omega^{\mathbb{N}}} h(\omega) \mathbb{P}_{\Xi}(d\omega) = \int_{\Omega} c(z_0^-) \Xi(dz_0^-, dz_0^+) = \int_{\Omega^-} c(z_0^-) \eta^-(dz_0^-) < \infty$$

where we successively used that the marginal (at “time” $n = 0$) of the path measure \mathbb{P}_{Ξ} is (by definition) Ξ and that the marginal (in space) of Ξ with respect to Ω^- is η^- — and evidently that $\mathcal{C}(\eta^-) < \infty$ is finite by the assumption that $\mathcal{C}(\pi)$ is finite and the fact that $\eta^- \ll \pi$ and $\frac{d\eta^-}{d\pi} \in L^\infty(\pi)$. Let us then come back to the construction of the previous section and show how Birkhoff’s theorem can be applied rightfully in order to conclude.

Let us first consider a single Markov chain $(\mathbf{Z}_n)_n \in \Omega^{\mathbb{N}}$ where $\mathbf{Z}_n = (Z_n^-, Z_n^+)$ and where $Z_n^- = (x_n, y_n)$ and $Z_n^+ = (x_n, y_{n+1})$. Beware *again*

that all these quantities are *random* and depend intrinsically on a path $\omega \in \Omega^{\mathbb{N}}$, *i.e.* $x_n = x_n(\omega)$ and $y_n = y_n(\omega)$. Applying what precedes, we have

$$\frac{1}{n} \sum_{k=0}^{n-1} [c(x_k, y_{k+1}) - c(x_k, y_k)] \rightarrow \mathbb{E}_{\mathbb{P}_{\Xi}}[f_c | \mathcal{J}](\omega)$$

for \mathbb{P}_{Ξ} almost-surely all $\omega \in \Omega^{\mathbb{N}}$. Now, thanks to the c -cyclical monotonicity property which is assumed to hold, we have

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^{n-1} [c(x_k, y_{k+1}) - c(x_k, y_k)] \\ + \frac{1}{n} c(\bar{x}, y_0) + \frac{1}{n} c(x_n, \bar{y}) - \frac{1}{n} c(\bar{x}, \bar{y}) - \frac{1}{n} c(x_n, y_n) \geq 0 \end{aligned}$$

— again for \mathbb{P}_{Ξ} almost-surely all $\omega \in \Omega^{\mathbb{N}}$. By non-negativity² of the cost of transport c , this implies that

$$(11) \quad \frac{1}{n} \sum_{k=0}^{n-1} [c(x_k, y_{k+1}) - c(x_k, y_k)] \geq -\frac{1}{n} c(\bar{x}, y_0) - \frac{1}{n} c(x_n, \bar{y}).$$

Let us assume for the moment that the cost of transport $c : X \times Y \rightarrow \mathbb{R}_+$ is uniformly bounded. Then, the right-hand side in the above inequality vanishes in the limit $n \rightarrow \infty$. This entails that $\mathbb{E}_{\mathbb{P}_{\Xi}}[f_c | \mathcal{J}] \geq 0$. But, taking again the expectation with respect to \mathbb{P}_{Ξ} , this implies that $\mathbb{E}_{\mathbb{P}_{\Xi}}[f_c] \geq 0$. But

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_{\Xi}}[f_c] &:= \int_{\Omega^{\mathbb{N}}} f_c(\omega) \mathbb{P}_{\Xi}(d\omega) \\ &= \int_{\Omega} [c(z_0^+) - c(z_0^-)] \Xi(dz_0^-, dz_0^+) \\ &= \int_{\Omega^+} c(z_0^+) \eta^+(dz_0^+) - \int_{\Omega^-} c(z_0^-) \eta^-(dz_0^-) = \mathcal{C}(\eta) \end{aligned}$$

where we used that the marginal of Ξ with respect to Ω^- (resp. Ω^+) is η^- (resp. η^+). Thus, we obtain $\mathcal{C}(\eta) \geq 0$ as sought-for in the case where the cost of transport is uniformly bounded, leading to the sufficiency of c -cyclical monotonicity to yield optimality in this case. Let us now prove Theorem 2 which states this result for less stringent assumption on the cost of transport.

Proof of Theorem 2. The cost of transport c is only assumed here to be finite $\mu \otimes \nu$ almost-surely. In this case, the right-hand side (11) does not necessarily vanish in the limit $n \rightarrow \infty$ and we cannot conclude as easily as in the case where c is uniformly bounded. Here, the game is to select the point $(\bar{x}, \bar{y}) \in \text{supp}(\pi)$ so that the two terms on the right-hand side vanish in the limit $n \rightarrow \infty$. Since c is finite $\mu \otimes \nu$ -almost everywhere, and thus

² Of course, we could certainly work with a cost that is bounded from below using the known trick consisting in adding and removing this constant, or more generally with costs of transport which are bounded from below by a sum of functions which are integrable with respect to the marginals, see *e.g.* [20, Theorem 4.1]

also $m_1 \otimes m_2$ -almost everywhere since $m_1 \ll \mu$ and $m_2 \ll \nu$, and since $\mathbf{Law}(y_0) = m_2$, there exists a Borel set $A \subset X$ so that $m_1(A) = 1$ and such that for all $\bar{x} \in A$ we have \mathbb{P}_Ξ almost-surely that $c(\bar{x}, y_0) < \infty$ and thus $\frac{1}{n}c(\bar{x}, y_0) \rightarrow 0$ as $n \rightarrow \infty$. Now, as for the second term, we remark that $\mathbf{Law}(x_n) = \mathbf{Law}(x_0) = m_1$ for all $n \in \mathbb{N}$. Therefore, we claim that there exists a set $B \subset Y$ so that $m_2(B) = 1$ and so that $\frac{1}{n}c(x_n, \bar{y}) \rightarrow 0$ in probability for $\bar{y} \in B$ — and thus \mathbb{P}_Ξ almost-surely up to a subsequence. Indeed,

$$\mathbb{P}_\Xi(\frac{1}{n}c(x_{n+1}, \bar{y}) > \varepsilon) = \mathbb{P}_\Xi(c(x_0, \bar{y}) > \varepsilon n) \xrightarrow{n \rightarrow \infty} \mathbb{P}_\Xi(c(x_0, \bar{y}) = \infty)$$

Therefore — again by the finiteness assumption $\mu \otimes \nu$ -almost everywhere of the cost of transport — there exists a set B as claimed. Now, for both terms to vanish *simultaneously*, it suffices to take any $(\bar{x}, \bar{y}) \in \text{supp}(\eta^-) \cap (A \times B)$, which is non-empty since $\eta^-(A \times B) = 1$. \square

3. EXTENSION OF CYCLICAL MONOTONICITY TO THE WEAK OPTIMAL TRANSPORT

3.1. Setting and notations. We use the strategy introduced above to extend the cyclical monotonicity to the weak optimal transport problem, namely:

$$(WOT) \quad \inf_{\pi \in \Pi(\mu, \nu)} \int_X C(x, \pi_x) \mu(dx)$$

where $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ are the two marginals and $C : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ is throughout assumed to be convex with respect to its second variable, namely for every $x \in X$, the map $\rho \mapsto C(x, \rho)$ is convex. Here $(\pi_x)_{x \in X}$ denotes the disintegration of the transport plan π onto its first marginal μ so that $\pi(dx, dy) = \mu(dx) \otimes \pi_x(dy)$. A *sort of* cyclical monotonicity property for (WOT) has been introduced and coined as C -monotonicity in [2]. It reads as follows:

Definition 4. A transport plan $\pi \in \Pi(\mu, \nu)$ is said to be C -monotone if there exists a set $\Gamma \subset X$ with $\mu(\Gamma) = 1$ such that for any finite number of points $x_1, \dots, x_n \in \text{supp}(\Gamma)$ and measures $m_1, \dots, m_n \in \mathcal{P}(Y)$ such that $\sum_{i=1}^n m_i = \sum_{i=1}^n \pi_{x_i}$ we have

$$(12) \quad \sum_{i=1}^n C(x_i, \pi_{x_i}) \leq \sum_{i=1}^n C(x_i, m_i)$$

It is known that this notion does *not* extend rightfully the classical notion of cyclical monotonicity, since one does *not* (always) recover it in the special where the weak optimal transport problem is actually the classical one, that is for the choice $C(x, \rho) := \int_Y c(x, y) \rho(dy)$. The following counterexample is to be found in [3, Example 5.3]:

Example 2. We assume that the underlying marginal spaces are $X = Y = [0, 1]$ and that $C(x, \rho) = \int_{[0,1]} c(x, y) \rho(dy)$ where $c(x, y) := 1 - \mathbf{1}_{\{x\}}(y)$. Let

λ be the Lebesgue measure on $[0, 1]$. Then, the transport plan $\pi := \lambda \otimes \lambda \in \Pi(\lambda, \lambda)$ is easily seen to be C -monotone in the sense of Definition 4, but it is certainly not optimal for the optimal transport problem where $\mu = \nu = \lambda$ and where the cost of transport is precisely given by c .

In contrast, we introduce in Definition 5 below a more consistent extension of the c -cyclical monotonicity property for the weak optimal transport, in the sense that it is equivalent to the classical cyclical monotonicity as in Definition 1 when the weak transport is actually the classical one. Before that, we discuss an *ad hoc* notion of differentiability that will be used throughout this section.

3.1.1. *A notion of differentiability.* Let us explain the notion of differentiability that will be used throughout this section. Here, for each fixed $x \in X$, the convex functional $\mathcal{P}(Y) \ni \rho \rightarrow C(x, \rho)$ is canonically extended to the set $\mathcal{M}(Y)$ of all signed and finite Radon measures by setting $C(x, \rho) = +\infty$ if $\rho \notin \mathcal{P}(Y)$. We first let $\partial_\rho C(x, \bar{\rho})$ be the subdifferential at the point $\bar{\rho} \in \mathcal{P}(Y)$ defined as

$$(13) \quad \partial_\rho C(x, \bar{\rho}) := \left\{ g \in C_0(Y) : C(x, \rho) \geq C(x, \bar{\rho}) + \int_Y g(y)[\rho - \bar{\rho}](dy), \forall \rho \in \mathcal{M}(Y) \right\}$$

where we recall that $C_0(Y)$ denotes the set of real-valued continuous functions on Y which vanish at infinity. The elements of the subdifferential are the so-called subgradients. For what matters to us, this definition is slightly too strong since we want to cope with subgradients which are *not* necessarily continuous. To this end, we define an *extended* subdifferential as follows.

Let us write $\mathfrak{M}(Y)$ the set of Borel measurable functions $g : Y \rightarrow \mathbb{R}$. We say that $g \in \mathfrak{M}(Y)$ is an *extended subgradient* of the function $\mathcal{P}(Y) \ni \rho \rightarrow C(x, \rho)$ at the point $\bar{\rho} \in \mathcal{P}(Y)$ if

$$C(x, \rho) \geq C(x, \bar{\rho}) + \int_Y g(y)[\rho - \bar{\rho}](dy)$$

for all $\rho \in \mathcal{P}(Y)$ **provided that** $\int_Y |g| \bar{\rho} < \infty$ **and that** $\int_Y |g| \rho < \infty$ **as soon as** $C(x, \rho) < \infty$. The set of all g verifying the above inequality and assumptions will be denoted $\mathfrak{D}_\rho C(x, \bar{\rho})$ and will be referred to as the *extended subdifferential* and its elements as *extended subgradients*. Obviously, this is a superset of the standard set of subgradients, that is $\partial_\rho C(x, \bar{\rho}) \subset \mathfrak{D}_\rho C(x, \bar{\rho})$.

Now, we will say that $\rho \mapsto C(x, \rho)$ is *differentiable* at $\bar{\rho} \in \mathcal{P}(Y)$ if the extended subdifferential $\mathfrak{D}_\rho C(x, \bar{\rho})$ consists in a single element (up to additive constants). In this case, we will write (a canonical representation of) this element $\nabla_\rho C(x, \bar{\rho})$ and simply call it the *gradient*. It is a well-known fact of

convex analysis that, assuming that $C(x, \bar{\rho}) < \infty$, it holds

$$\lim_{t \downarrow 0^+} \frac{C(x, (1-t)\bar{\rho} + t\xi) - C(x, \bar{\rho})}{t} = \sup_{g \in \partial_\rho C(x, \bar{\rho})} \int_Y g(y)[\xi - \bar{\rho}](dy)$$

for all $\xi \in \mathcal{P}(Y)$ such that $C(x, \xi) < \infty$. Now, evidently, by definition of extended subgradients, we also have:

$$\lim_{t \downarrow 0^+} \frac{C(x, (1-t)\bar{\rho} + t\xi) - C(x, \bar{\rho})}{t} \geq \sup_{g \in \mathfrak{D}_\rho C(x, \bar{\rho})} \int_Y g(y)[\xi - \bar{\rho}](dy)$$

This entails that

$$\begin{aligned} \sup_{g \in \partial_\rho C(x, \bar{\rho})} \int_Y g(y)[\xi - \bar{\rho}](dy) &\geq \\ &\sup_{g \in \mathfrak{D}_\rho C(x, \bar{\rho})} \int_Y g(y)[\xi - \bar{\rho}](dy) \\ &\geq \sup_{g \in \partial_\rho C(x, \bar{\rho})} \int_Y g(y)[\xi - \bar{\rho}](dy) \end{aligned}$$

where, in the last inequality, we used that $\partial_\rho C(x, \rho) \subset \mathfrak{D}_\rho C(x, \rho)$. Therefore, we have equality of both *suprema*. In particular, if $\mathcal{P}(Y) \ni \rho \mapsto C(x, \rho)$ is differentiable in the aforementioned sense at the point $\bar{\rho} \in \mathcal{P}(Y)$, then we have

$$(14) \quad \lim_{t \downarrow 0^+} \frac{C(x, (1-t)\bar{\rho} + t\xi) - C(x, \bar{\rho})}{t} = \int_Y \nabla_\rho C(x, \rho)(y)[\xi - \bar{\rho}](dy).$$

provided that $C(x, \xi) < \infty$, where the gradient $\nabla_\rho C(x, \rho)$ is understood in the extended sense, *i.e.* it is only considered measurable and thus may not be necessarily continuous.

3.1.2. The C -cyclical monotonicity property. We are now ready to introduce the C -cyclical monotonicity property, which will serve as the analogue of the c -cyclical monotonicity property of the classical optimal transport setting. This property reads as follows:

Definition 5. A transport plan $\pi \in \Pi(\mu, \nu)$ is said to be C -cyclically monotone if it is concentrated on a set $\Gamma \subset X \times Y$ so that for all $n \in \mathbb{N}$ and all $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$, it holds that

$$(15) \quad \sum_{i=1}^n \nabla_\rho C(x_i, \pi_{x_i})(y_i) \leq \sum_{i=1}^n \nabla_\rho C(x_i, \pi_{x_i})(y_{i+1})$$

where we let $y_{n+1} := y_1$ and where $Y \ni y \mapsto \nabla_\rho C(x, \pi_x)(y)$ is the gradient (see *supra*) of the cost C (provided it exists) with respect to its second argument evaluated at the point $(x, \pi_x) \in X \times \mathcal{P}(Y)$.

Example 3. (1) In the special case where $C(x, \rho) = \int_Y c(x, y)\rho(dy)$, that is when the weak optimal transport problem boils down to the classical one, then the gradient of the cost of transport C is simply

the function $y \mapsto c(x, y)$ for all $x \in X$, and (15) is therefore seen to be the usual, standard c -cyclical monotonicity as of Definition 1.

- (2) In the case of the *barycentric quadratic cost*, namely $C(x, \rho) := \frac{1}{2} \|x - \int_Y z \rho(dz)\|^2$, then $\nabla_\rho C(x, \rho)(y) = y^\top (\int_Y z \rho(dz) - x)$ for all $x \in X$ and $\rho \in \mathcal{P}(Y)$ and the C -cyclical monotonicity property (15) reads

$$\sum_{i=1}^n y_i^\top \left(\int_z z \pi_{x_i}(dz) - x_i \right) \leq \sum_{i=1}^n y_{i+1}^\top \left(\int_z z \pi_{x_i}(dz) - x_i \right).$$

Our main result in this section is that (under suitable assumptions) the C -cyclical monotonicity property as defined above in Definition 5 is equivalent to optimality for the weak optimal transport problem (WOT). The necessary part will be stated and proved below in Theorem 3 and the sufficient part in Theorem 4.

3.2. The first-order optimality conditions for (WOT). In this section, we follow the same strategy as in Section 2 where we derived the zeroth-order optimality condition for the classical OT problem. The main difference here is that the objective of the weak optimal transport problem

$$\mathcal{E}(\pi) := \int_X C(x, \pi_x) \mu(dx)$$

is no longer linear. Nevertheless, under the assumption that C be convex with respect to its second variable, the function \mathcal{E} is convex, so that its minimum can be characterised using a *first-order* optimality condition. We will assume from now on that \mathcal{E} is extended from $\Pi(\mu, \nu)$ to the entire space of (finite and signed) Radon measures $\mathcal{M}(X \times Y)$ by letting $\mathcal{E} = +\infty$ outside of the feasible set of transport plans. We start with the following lemma:

Lemma 3. *Let $C : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ be convex and differentiable with respect to its second variable. Then, the objective function $\mathcal{E} : \mathcal{M}(X \times Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ of the weak optimal transport problem is convex. Furthermore, let $\pi \in \Pi(\mu, \nu)$ be such that $\mathcal{E}(\pi) < \infty$ and such that the following property holds*

$$(A_\pi) \quad \mathcal{E}(\gamma) < \infty \implies \int_{X \times Y} |\nabla_\rho C(x, \pi_x)(y)| \gamma(dx, dy) < \infty,$$

for all $\gamma \in \Pi(\mu, \nu)$. It is implicitly assumed in (A_π) that $X \times Y \ni (x, y) \mapsto \nabla_\rho C(x, \pi_x)(y)$ is jointly measurable. Then, for all $\gamma \in \Pi(\mu, \nu)$ such that $\mathcal{E}(\gamma) < \infty$, it holds that

$$(16) \quad \mathcal{E}(\gamma) \geq \mathcal{E}(\pi) + \iint_{X \times Y} \nabla_\rho C(x, \pi_x)(y) [\gamma - \pi](dx, dy).$$

Furthermore, again for all $\gamma \in \Pi(\mu, \nu)$ such that $\mathcal{E}(\gamma) < \infty$, we have

$$(17) \quad \lim_{t \downarrow 0^+} \frac{\mathcal{E}((1-t)\pi + t\gamma) - \mathcal{E}(\pi)}{t} = \int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) [\gamma - \pi](dx, dy).$$

Let us point out that the assumption (A_π) (and the assumption (A'_π) below) is only used to guarantee that the linearised functional is well-defined along all finite-cost competitors.

The inequality (16) of Lemma 3 would read that the function g is a *sub-gradient* (in the usual sense, that is in duality with $C_0(X \times Y)$) of the cost functional \mathcal{C} at the point π if it was furthermore known that g were *at least* continuous. Nevertheless, there is no reason for this to be true, since the measurable map $X \ni x \mapsto \pi_x \in \mathcal{P}(Y)$ is *not* necessarily (vaguely) continuous. Instead, g should be thought as an “algebraic” sort of subgradient.

The technical condition (A_π) ensures that the integral on the right-hand side of (16) makes sense as soon as the left-hand side is finite. We will make use of this assumption throughout our result. We remark that, in the case where $C(x, \rho) = \int_Y c(x, y)\rho(dy)$, in which case $\nabla_\rho C(x, \rho)(y) = c(x, y)$, then this condition simply says that π as finite transport cost, and is therefore always true since it is precisely assumed that π has finite cost, that is $\int_{X \times Y} c d\pi < \infty$. It is also met for the barycentric squared cost if $\int_Y \|y\|^2 \nu(dy) < \infty$. Indeed, in this case $\nabla_\rho C(x, \rho)(y) = y^\top (\int_Y z \rho(dz) - x)$ and thus

$$\int_{X \times Y} |\nabla_\rho C(x, \pi_x)(y)| \gamma(dx, dy) \leq \sqrt{\int_{X \times Y} \|y\|^2 \gamma(dx, dy)} \sqrt{\int_{X \times Y} \left\| x - \int_Y z \pi_x(dz) \right\|^2 \gamma(dx, dy)}$$

by the standard Cauchy-Schwarz inequality. Therefore, marginalising the integrals in the above inequality, we obtain

$$\begin{aligned} \int_{X \times Y} \left\| x - \int_Y z \pi_x(dz) \right\|^2 \gamma(dx, dy) \\ = \int_X \left\| x - \int_Y z \pi_x(dz) \right\|^2 \mu(dx) = \mathcal{C}(\pi) < \infty \end{aligned}$$

and

$$\int_{X \times Y} \|y\|^2 \gamma(dx, dy) = \int_Y \|y\|^2 \nu(dy) < \infty$$

Thus $\int_{X \times Y} |\nabla_\rho C(x, \pi_x)(y)| \gamma(dx, dy) < \infty$ is finite for all $\gamma \in \Pi(\mu, \nu)$ provided that ν has bounded second-moment. We note that this holds irrespective of whether or not $\mathcal{C}(\gamma)$ is finite, so that the assumption (A_π) is always verified. The same rationale can be applied to weak costs of transport of the form $C(x, \rho) := \theta(x - \int_Y z \rho(dz))$ where $\theta : \mathbb{R}^d \rightarrow \mathbb{R}$ is convex and differentiable function that furthermore verifies $\|\nabla \theta\|^2 \leq \theta$ (up to a multiplicative constant). The assumption (A_π) is also seen to hold for the *Marton cost*, i.e. $C(x, \rho) = (\int_Y \mathbf{1}_{\{x\}}(z) \rho(dz))^2$ — with underlying spaces $X = Y = [0, 1]$. Altogether, this assumption is general enough to encompass important cases of applications of the weak optimal transport.

Proof. Let $\pi, \gamma \in \Pi(\mu, \nu)$ with disintegrations $(\pi_x)_{x \in X}$ and $(\gamma_x)_{x \in X}$ with respect to the marginal μ . Then, the disintegration of the convex combination $(1-t)\pi + t\gamma$ with respect to μ is simply given by the convex combination of the disintegrations, to wit $((1-t)\pi_x + t\gamma_x)_{x \in X}$. Therefore,

$$\begin{aligned} \mathcal{E}((1-t)\pi + t\gamma) &= \int_X C(x, (1-t)\pi_x + t\gamma_x) \mu(dx) \\ &\leq (1-t) \underbrace{\int_X C(x, \pi_x) \mu(dx)}_{\mathcal{E}(\pi)} + t \underbrace{\int_X C(x, \gamma_x) \mu(dx)}_{\mathcal{E}(\gamma)} \end{aligned}$$

where we used solely the convexity of C with respect to its second argument. This proves that \mathcal{E} is convex. Now, we also have that

$$\mathcal{E}(\gamma) - \mathcal{E}(\pi) = \int_X (C(x, \gamma_x) - C(x, \pi_x)) \mu(dx)$$

for all $\gamma, \pi \in \Pi(\mu, \nu)$. Here, we note that we used that $\mathcal{E}(\pi) < \infty$, so that the above difference makes sense. For all $\gamma \in \Pi(\mu, \nu)$, using the convexity and differentiability of C with respect to its second variable, we have

$$(18) \quad C(x, \gamma_x) \geq C(x, \pi_x) + \int_Y \nabla_\rho C(x, \pi_x)(y) [\gamma_x - \pi_x](dy)$$

We emphasise here that by the technical assumption (A_π) the integrals on the right-hand side make sense. Then, integrating with respect to μ , we obtain

$$(19) \quad \mathcal{E}(\gamma) \geq \mathcal{E}(\pi) + \int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) \underbrace{[\gamma_x - \pi_x](dy) \otimes \mu(dx)}_{=[\gamma - \pi](dx, dy)}$$

Here, we relied implicitly (again) on the technical assumption (A_π) so as to appeal to Fubini-Lebesgue theorem — as we did also above in (18). This proves (16). Now, let us show (17). First, applying the inequality (16) that we have just proved to $\gamma \leftarrow (1-t)\pi + t\gamma$, we have immediately that

$$\liminf_{t \downarrow 0} \frac{\mathcal{E}((1-t)\pi + t\gamma) - \mathcal{E}(\pi)}{t} \geq \int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) [\gamma - \pi](dx, dy).$$

Let us now show that the *supremum* limit is smaller or equal to the right-hand side, so that (17) will follow. We have

$$\frac{\mathcal{E}((1-t)\pi + t\gamma) - \mathcal{E}(\pi)}{t} = \int_X \phi_t(x) \mu(dx)$$

where we define $\phi_t : X \rightarrow \mathbb{R}$ as

$$\phi_t(x) := \frac{C(x, \pi_x + t(\gamma_x - \pi_x)) - C(x, \pi_x)}{t}.$$

The function ϕ_t is measurable. Furthermore, for each fixed $x \in X$, thanks to the convexity of C with respect to its second variable, it is monotone in the variable $t \in [0, 1]$. In particular, we have $\phi_t(x) \leq \phi_1(x)$ for all $x \in X$. We claim that ϕ_1 is integrable with respect to μ . Indeed, $\phi_1(x) = C(x, \gamma_x) - C(x, \pi_x)$, and both of these terms once integrating against μ yield

respectively the weak costs $\mathcal{C}(\gamma)$ and $\mathcal{C}(\pi)$, which are assumed to be finite. Therefore, we can appeal to Fatou's lemma to obtain

$$\begin{aligned} \limsup_{t \downarrow 0^+} \frac{\mathcal{C}((1-t)\pi + t\gamma) - \mathcal{C}(\pi)}{t} &\leq \int_X \limsup_{t \downarrow 0} \phi_t(x) \mu(dx) \\ &= \int_X \lim_{t \downarrow 0} \phi_t(x) \mu(dx) \end{aligned}$$

Here, the *supremum limit* is a limit by convexity of $\mathcal{P}(Y) \ni \rho \mapsto C(x, \rho)$. But now, we recall by differentiability of C with respect to its second variable that

$$\lim_{t \rightarrow 0} \phi_t(x) = \int_Y \nabla_\rho C(x, \pi_x)(y) [\gamma_x - \pi_x](dy)$$

as shown in (14). We emphasise that the above limit holds for μ almost-surely all $x \in X$ since the assumption $\mathcal{C}(\gamma) < \infty$ implies that $C(x, \gamma_x)$ is finite for μ almost-surely all $x \in X$, which ensures that (14) holds. This terminates the proof of (17). \square

We will show the following first-order optimality conditions for the weak OT problem:

Lemma 4. *Let $C : X \times \mathcal{P}(Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ be such that for all $x \in X$ the map $\mathcal{P}(Y) \ni \rho \mapsto C(x, \rho)$ is convex and differentiable. Assume $\pi \in \Pi(\mu, \nu)$ has finite cost $\mathcal{C}(\pi)$ and the property (A_π) holds. Then π is optimal for the weak optimal transport problem if and only if*

$$(20) \quad \boxed{\int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) \eta^+(dx, dy) \geq \int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) \eta^-(dx, dy)}$$

for all $\eta \in \mathcal{R}_\pi$ provided that $\mathcal{C}(\pi + t\eta) < \infty$ for all $t > 0$ small enough.

We recall that \mathcal{R}_π is the radial cone to feasible set $\Pi(\mu, \nu)$ at the point $\pi \in \Pi(\mu, \nu)$ as defined in (4). We also recall that η^+ (resp. η^-) denotes the positive (resp. negative) part of the signed and finite Radon measure $\eta \in \mathcal{M}(X \times Y)$. The inequality (20) is the first-order optimality condition analogue of the zeroth-order optimality condition of Proposition 1. Here, we need to go to the first order because the objective of the WOT problem is no longer linear — see also Remark 5 below for a more general picture of the situation.

Proof of Lemma 4. We start with the sufficient part. Beware that throughout, we work under the technical assumption (A_π) . Let $\gamma \in \Pi(\mu, \nu)$ be such that $\mathcal{C}(\gamma) < \infty$. By Lemma 3, we have

$$\mathcal{C}(\gamma) \geq \mathcal{C}(\pi) + \int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) [\gamma - \pi](dx, dy).$$

We now prove that the last integral is non-negative. Let us set $\eta := \gamma - \pi$, so that η is an admissible perturbation at π , that is $\eta \in \mathcal{R}_\pi$. Indeed, we have

$$\pi + t\eta = (1-t)\pi + t\gamma \in \Pi(\mu, \nu) \quad \text{for all } t \in [0, 1].$$

Moreover, by convexity of \mathcal{C} ,

$$\mathcal{C}(\pi + t\eta) = \mathcal{C}((1-t)\pi + t\gamma) \leq (1-t)\mathcal{C}(\pi) + t\mathcal{C}(\gamma) < \infty$$

for all $t \in [0, 1]$. Hence, we can rightfully appeal to the condition (20). Writing $\eta = \eta^+ - \eta^-$ for the Jordan decomposition of the signed measure η , the condition (20) yields

$$\int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) \eta^+(dx, dy) \geq \int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) \eta^-(dx, dy).$$

Consequently,

$$\int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) [\gamma - \pi](dx, dy) = \int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) \eta(dx, dy) \geq 0.$$

Therefore $\mathcal{C}(\gamma) \geq \mathcal{C}(\pi)$. Since this holds for every $\gamma \in \Pi(\mu, \nu)$ with $\mathcal{C}(\gamma) < \infty$, the transport plan π is optimal.

Let us now deal with the necessary part of the lemma. Let $\pi \in \Pi(\mu, \nu)$ be an optimal transport plan for the weak optimal transport problem that verifies the assumptions of the lemma. Given a perturbation $\eta \in \mathcal{R}_\pi$, let us show that (20) holds. By optimality of π – and the fact that $\mathcal{C}(\pi) < \infty$ – we have $\mathcal{C}(\pi + t\eta) - \mathcal{C}(\pi) \geq 0$ for all $t > 0$ small enough. Dividing by t this inequality, letting $t \rightarrow 0$ and appealing to (17), we obtain that

$$\int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) \eta(dx, dy) \geq 0.$$

We then obtain the claim that (20) holds by decomposing η into its positive and negative parts. \square

Remark 5. *As in Remark 1, let us indicate that is again a much general situation to which the weak optimal transport problem is only a particular case. Let X be a normed vector space, $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a (differentiable) convex function on X and $C \subset X$ be a convex set. Let us consider the convex minimisation problem*

$$(21) \quad \min_{x \in C} f(x).$$

Then, $x \in C$ is optimal if and only if $\nabla f(x)[h] \geq 0$ for all $h \in R_C(x)$ – where $R_C(x)$ is the radial cone defined in (3). Indeed, let us assume x is optimal, and let $h \in R_C(x)$. Then, $f(x + th) - f(x) \geq 0$ for $t > 0$ small enough, so that standardly dividing by t and letting it tend to zero yields the claim. The other direction goes as follows: since f is convex, we have $f(x) + \nabla f(x)[y - x] \leq f(y)$ for all $x, y \in C$. Thus, if $\nabla f(x)[h] \geq 0$ for all $h \in R_C(x)$, it is then immediate that $f(x) \leq f(y)$ for all y and thus x is optimal. In the case where f is not differentiable, then we can use subgradients. Then, if for all $h \in R_C(x)$ there exists at least one $g \in \partial f(x)$ – where $\partial f(x)$ denotes the (non-empty) subdifferential of f at x – such that $\langle g, h \rangle \geq 0$, then x is optimal. Indeed, for all $y \in C$, letting $h := y - x$, then $f(x) \leq f(x) + \langle g, h \rangle \leq f(x + h) = f(y)$ such x is optimal. In the other direction, assume that x is optimal. Then, for all $h \in R_C(x)$ and

t small enough $f(x + th) \geq f(x)$. This entails that $f'(x; h) \geq 0$ for all $h \in R_C(x)$ where we let $f'(x; h)$ be the directional derivative at f in the direction h (since f is convex, it exists for all h in $\mathbb{R} \cup \{+\infty\}$). Then either $f'(x; h) = 0$ for all h , then f is actually differentiable at x and $\nabla f(x) = 0$, so the conclusion is trivial. Otherwise, there exists a direction $h_0 \in R_C(x)$ such that $f'(x; h_0) > 0$. Using that $f'(x, h_0) = \sup_{g \in \partial f(x)} \langle g, h_0 \rangle$ this means that there exists a subgradient $g_0 \in \partial f(x)$ such that $\langle g_0, h_0 \rangle \geq 0$. Therefore, the exact statement is: if x is optimal, then for all $h \in R_C(x)$ there exists $g \in \partial f(x)$ such that $\langle g, h \rangle \geq 0$.

3.3. C -cyclical monotonicity characterises optimality. We have shown, in the case of the classical OT problem, that one can easily go from the validity of the zeroth-order optimality condition for the classical OT problem onto \mathcal{R}_π from its validity onto \mathcal{F}_π — yielding the *necessary part* of c -cyclical monotonicity for optimality — either by mere density if the cost functional $\pi \mapsto \int_{X \times Y} c d\pi$ was continuous, see Proposition 4, or under less stringent assumptions on the cost of transport c using the ergodic theorem, cf. Theorem 2. *Vice-versa*, we went from the validity of the zeroth-order optimality condition from \mathcal{F}_π onto \mathcal{R}_π — yielding the *sufficient part* of c -cyclical monotonicity for optimality — either by mere density again, appealing to Proposition 3, if again the cost functional is continuous, cf. Proposition 4.

Now, to prove that C -cyclical monotonicity as defined in Definition 5 is equivalent to optimality for the weak optimal transport problem (WOT) is in some sense completely “free”, thanks to the work done previously for the classical optimal transport problem.

First, let us remark that C -cyclical monotonicity as in Definition 5 is trivially equivalent to the veracity of the *first-order optimality condition* (20) when evaluated at *cyclical perturbations* $\eta \in \mathcal{C}_\pi$. Just as in the classical optimal transport case, we can define *finite optimality* here, which simply consists in the validity of first-order optimality condition (20) when evaluated at *finite perturbation* $\eta \in \mathcal{F}_\pi$. We emphasise that finite optimality and C -cyclical monotonicity are again equivalent for the weak optimal transport. Indeed, it suffices to apply Remark 2 with the linear map $L : \mathcal{M}(X \times Y) \rightarrow \mathbb{R} \cup \{+\infty\}$ defined as

$$L(\eta) := \int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) \eta(dx, dy).$$

— provided that the map $X \times Y \ni (x, y) \mapsto \nabla_\rho C(x, \pi_x)(y)$ is Borel measurable, as will be assumed in our theorems.

Second, we also know by Lemma 4 above that optimality in (WOT) is completely equivalent to the first-order optimality condition (20) evaluated for perturbations $\eta \in \mathcal{R}_\pi$ in the radial cone. So, it suffices to prove that (under suitable assumptions) we can go from these two properties back-and-forth, just like we did in the case of the classical optimal transport

problem. But whereas previously, we worked with the cost functional

$$\pi \mapsto \int_{X \times Y} c(x, y) \pi(dx, dy)$$

to be non-negative of the radial cone \mathcal{R}_π , now instead of the cost functional, we work with

$$\eta \mapsto \int_{X \times Y} \nabla_\rho C(x, \pi_x)(y) \eta(dx, dy)$$

which is to be thought as the linearisation of the weak cost functional \mathcal{C} at π . To prove the necessary and sufficient parts of C -cyclical for optimality, we just need to re-enact the former proofs of Theorem 1 and Theorem 2 with the “cost” function $c(x, y) := \nabla_\rho C(x, \pi_x)(y)$.

The necessary part reads as follows:

Theorem 3. *Assume that the map*

$$X \times \mathcal{P}(Y) \times Y \ni (x, \rho, y) \mapsto \nabla_\rho C(x, \rho)(y) \in \mathbb{R} \cup \{+\infty\}$$

is Borel-measurable and (locally) integrable with respect to $\mu \otimes \nu$. Then, if an optimal transport plan $\pi \in \Pi(\mu, \nu)$ has finite cost, that is $\mathcal{C}(\pi) < \infty$, and that it furthermore verifies the property (A_π) holds, then π is C -cyclically monotone.

Proof of Theorem 3. This is direct from the (proof of) Theorem 1. Indeed, we first remark that the map $x \mapsto \pi_x$ is measurable³ by definition of the disintegration. Therefore, the map $X \times Y \ni (x, y) \mapsto \nabla_\rho C(x, \pi_x)(y)$ is also Borel-measurable. According to Lemma 4, by optimality of π – and by the assumptions that $\mathcal{C}(\pi) < \infty$ and that the technical assumption (A_π) holds — we have

$$\int_{X \times Y} c(x, y) \eta(dx, dy) \geq 0$$

for all $\eta \in \mathcal{R}_\pi$ where we let $c(x, y) := \nabla_\rho C(x, \pi_x)(y)$. To prove that π is C -cyclically monotone, we want to prove that $\int_{X \times Y} c(x, y) \eta(dx, dy) \geq 0$ for all $\eta \in \mathcal{C}_\pi$, that is for all cyclical perturbations. To do so, it suffices to reiterate again the construction of the proof of Theorem 1 *as-is*, where this time the “cost” function $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ is the map defined above. Indeed, under the assumption of the above theorem, the function c valued in the extended real line is Borel-measurable, which was the sole assumption of Theorem 1. \square

Remark 6. *This theorem should be compared to [2, Theorem 5.3], where it is proved that optimality implies C -monotonicity in under the assumptions that C be jointly measurable, and convex and lower semicontinuous with respect to its second variable. Here, the sole underlying assumption on C*

³This is understood in the sense that for all Borel set $B \in \mathcal{P}(Y)$, the map $x \mapsto \pi_x(B) \in \mathbb{R}$ is Borel measurable. This implies by the monotone class theorem that $x \in X \mapsto \int_{X \times Y} f(x, y) \mu_x(dy)$ is Borel measurable for all (bounded from below) Borel functions f , see [1, Section 5.3]

itself is that it be convex with respect to its second variable — but evidently assumptions on its gradient are needed. Of course, here, we remind the reader that we work with a different notion of cyclical monotonicity.

As for the sufficient part, it is also as straightforward. First note that we do not have a clear counterpart of Proposition 4. This is because there are no adequate assumptions making the function $(x, y) \mapsto \nabla_\rho C(x, \pi_x)(y)$ continuous (and vanishing at infinity). Indeed, the map $x \mapsto \pi_x$ need not be continuous under further structural assumption on C . For instance, for the classical OT problem, assuming that the optimal transport plan is of a Monge-type, meaning that $\pi_x = \delta_{T(x)}$ where T is the associated transport map, then this requires that it be continuous, which need not be true in full generality. So, we move on directly to the counterpart of Theorem 2:

Theorem 4. *Assume the map*

$$X \times \mathcal{P}(Y) \times Y \ni (x, \rho, y) \mapsto \nabla_\rho C(x, \rho)(y)$$

is Borel-measurable, bounded from below, and finite $\mu \otimes \nu$ -almost everywhere. Then, if $\pi \in \Pi(\mu, \nu)$ is C -cyclically monotone and has finite cost $\mathcal{C}(\pi) < \infty$, and provided property (A_π) holds, then it is optimal for the weak optimal transport problem (WOT).

Proof. Again, as in the above proof for the necessary part, this is a direct application of the preceding proofs, more precisely that of Theorem 2. According to Lemma 4, we need to show that

$$\int_{X \times Y} c(x, y) \eta(dx, dy) \geq 0$$

for all $\eta \in \mathcal{R}_\pi$ such that $\mathcal{C}(\pi + t\eta)$ is finite for all $t > 0$ small enough, where we let $c(x, y) := \nabla_\rho C(x, \pi_x)(y)$, provided the above inequality holds for finite perturbations, that this for all $\eta \in \mathcal{F}_\pi$. Here, and for the same reasons as put forward in the proof of Theorem 3, the function $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ is Borel-measurable and bounded from below under the assumptions of the theorem. It suffices to use the Markovian construction of the proof of Theorem 2 with our specific cost c . \square

3.4. Removing differentiability: the case of the barycentric cost.

In this section, we explore briefly if what precedes can be extended to the case where the weak cost of transport C is not differentiable with respect to its second variable (but is nonetheless convex). In Section 3.4.1, we explain how one could extend the previous results without going into the details. In particular, we do not state any theorem. In Section 3.4.2, we rather specify more in detail what can be said in the case of the barycentric cost, which is not differentiable everywhere, as an illustration of the general arguments put forward in the first section.

3.4.1. *Removing differentiability.* A natural question is if whether or not what precedes can be extended to the case where the weak cost of transport C is not differentiable with respect to its second variable. Here, we recall that differentiability is understood in the sense introduced in Section 3.1.1. Of course, because for all $x \in X$ the function $\mathcal{P}(Y) \ni \rho \mapsto C(x, \rho)$ is convex, we can still work with (extended) subgradients. *What is then the equivalent of Lemma 4 ?* To answer that, we first need to ask ourselves what is the equivalent of Lemma 3. In its proof, we remark that (18) shall be replaced with

$$C(x, \gamma_x) \geq C(x, \pi_x) + \int_Y g_x(y)[\gamma_x - \pi_x](dy)$$

where $g_x : Y \rightarrow \mathbb{R}$ is a subgradient of $C(x, \cdot)$ at the point π . Of course, if g_x is continuous and vanishing at infinity (or in fact simply bounded), then the integral $\int_Y g_x(y)\gamma_x(dy)$ always makes sense. This is not *a priori* the case if g_x is understood in the *extended* sense, in which case g_x is only assumed to be measurable. Furthermore, in the extended setting, for the above inequality to make any sort of sense, we have to assume that $\int_Y |g_x(y)|\rho(dy) < \infty$ **as soon as** $C(x, \rho) < \infty$ — at least for μ almost-surely all $x \in X$. Now, to obtain the equivalent of (19), we have to integrate against μ the above inequality, yielding (formally)

$$\mathcal{C}(\gamma) \geq \mathcal{C}(\pi) + \int_{X \times Y} g_x(y)[\gamma - \pi](dx, dy).$$

Of course, here, we used formally the Fubini-Lebesgue theorem. For this to be readily justified, we need to add the following assumption, which is the analogue of the assumption (A_π) : for all $\pi \in \Pi(\mu, \nu)$ so that $\mathcal{C}(\pi) < \infty$, we have the property

$$(A'_\pi) \quad \mathcal{C}(\gamma) < \infty \implies \int_{X \times Y} |g_x(y)|\gamma(dx, dy) < \infty,$$

for all $\gamma \in \Pi(\mu, \nu)$. Here, $g_x \in \mathfrak{D}_\rho C(x, \pi_x)$ is an extended subgradient at π_x for all $x \in X$. We emphasise that here **it is implicitly assumed that the family of subgradients $(g_x)_{x \in X}$ under consideration in the statement of the assumption (A'_π) verifies that $X \times Y \ni (x, y) \mapsto g_x(y)$ is Borel-measurable.** We then obtain the following claim. *If $\pi \in \Pi(\mu, \nu)$ has finite cost $\mathcal{C}(\pi) < \infty$ and if the property (A'_π) holds — which implicitly implies that for all transport plan $\gamma \in \Pi(\mu, \nu)$ such that $\mathcal{C}(\gamma) < \infty$ there exists a family $(g_x)_{x \in X}$ of extended subgradients $g_x \in \mathfrak{D}_\rho C(x, \pi_x)$ such that $(x, y) \mapsto g_x(y)$ is jointly measurable — and if*

$$(22) \quad \int_{X \times Y} g_x(y)[\gamma - \pi](dx, dy) \geq 0,$$

then π must be optimal for the weak optimal transport problem (WOT). Beware that $g_x \equiv g_x^\gamma$ a priori depends on γ . Conversely, let us assume that π is an optimal transport plan for (WOT). Then, provided that $\mathcal{C}(\gamma) < \infty$,

we have that

$$\lim_{t \downarrow 0} \frac{\mathcal{C}(\pi + t(\gamma - \pi)) - \mathcal{C}(\pi)}{t} \geq 0$$

Here, the limit exists as an element of the extended real-line $\mathbb{R} \cup \{+\infty\}$ since \mathcal{C} is convex and by assumption $\mathcal{C}(\pi) < \infty$. By the same rationale as in the proof of Lemma 3, this entails that

$$\int_X \lim_{t \downarrow 0} \frac{C(x, \pi_x + t(\gamma_x - \pi_x)) - C(x, \pi_x)}{t} \mu(dx) \geq 0.$$

Here, note that the limit of the integrand does exist as an element of the extended real-line $\mathbb{R} \cup \{+\infty\}$ since C is convex with respect to its second variable *and* that, from the fact that $\mathcal{C}(\pi) < \infty$, then $C(x, \pi_x) < \infty$ for μ almost-surely all $x \in X$. Therefore, using what has been devised in Section 3.1.1, we have

$$(23) \quad \int_X \left(\sup_{g \in \mathfrak{D}_\rho C(x, \pi_x)} \int_Y g_x(y) [\gamma_x - \pi_x](y) \right) \mu(dx) \geq 0.$$

3.4.2. The case of the barycentric cost. We will not further explore the extension of what precedes in the case of *any* (non-differentiable) convex weak cost of transport C . Rather, we now specify more in detail what happens in the case of the plain *barycentric cost*. That is, we have $X = Y = \mathbb{R}^d$ and

$$C(x, \rho) = \left\| x - \int_{\mathbb{R}^d} z \rho(dz) \right\|.$$

We have that the barycentric cost C is differentiable with respect to its second variable at the point $(x, \rho) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ if (and only if) $x \neq \int_{\mathbb{R}^d} z \rho(dz)$, and in this case

$$\nabla_\rho C(x, \rho)(y) = - \frac{y^\top (x - \int_{\mathbb{R}^d} z \rho(dz))}{\|x - \int_{\mathbb{R}^d} z \rho(dz)\|}$$

up to additive constants. Otherwise, if $x = \int_{\mathbb{R}^d} z \rho(dz)$, the subdifferential is not reduced to a single element. It is obvious that functions of the form $g(y) = u^\top y$ where $u \in \mathbb{R}^d$ is such that $\|u\| \leq 1$ are subgradients, although we stress out that there may exist subgradients which are not of this form⁴. The supremum in (23) enjoys a totally explicit form. Indeed, it is attained for $g_x(y) := u_x^\top y$ where the vector $u_x \in \mathbb{R}^d$ is given by

$$u_x = - \frac{x - \int_{\mathbb{R}^d} z \pi_x(dz)}{\|x - \int_{\mathbb{R}^d} z \pi_x(dz)\|}$$

if $x \neq \int_{\mathbb{R}^d} z \pi_x(dz)$. Indeed, in this case, the extended subdifferential $\mathfrak{D}_\rho C(x, \pi_x)$ is reduced to a single function precisely stemming from the aforementioned vector. Otherwise, if $x = \int_{\mathbb{R}^d} z \pi_x(dz)$ and if $\int_{\mathbb{R}^d} z \gamma_x(dz) \neq x$, then we claim that

$$(24) \quad u_x = \frac{\int_{\mathbb{R}^d} z \gamma_x(dz) - \int_{\mathbb{R}^d} z \pi_x(dz)}{\|\int_{\mathbb{R}^d} z \gamma_x(dz) - \int_{\mathbb{R}^d} z \pi_x(dz)\|}.$$

⁴For instance, the function $g(y) = -\|y\|$ is an element of $\mathfrak{D}_\rho C(0, \delta_0)$.

Indeed, by definition of a subgradient $g : \mathbb{R}^d \rightarrow \mathbb{R}$, we have

$$(25) \quad \int_{\mathbb{R}^d} g(z)[\gamma_x - \pi_x](dz) \leq \left\| x - \int_{\mathbb{R}^d} z\gamma_x(dz) \right\|$$

where it is used here that $C(x, \pi_x) = 0$ since we are in the case where $x = \int_{\mathbb{R}^d} z\pi_x(dz)$. Now, the choice (24) precisely saturates the bound (25), and thus is the optimiser in (23) in this context. Finally, the third and last case arises when $\int_{\mathbb{R}^d} z\gamma_x(dz) = \int_{\mathbb{R}^d} z\pi_x(dz) = x$. In this case, a subgradient g must verify that

$$\int_{\mathbb{R}^d} g(z)[\gamma_x - \pi_x](dz) \leq 0.$$

To saturate this bound, it suffices to take $u_x = \bar{u}$ where $\bar{u} \in \mathbb{R}^d$ is any vector with $\|\bar{u}\| \leq 1$. Let us define

$$A := \left\{ x \in \mathbb{R}^d : x \neq \int_{\mathbb{R}^d} z\pi_x(dz) \right\}$$

and

$$B_\gamma := \left\{ x \in \mathbb{R}^d : \int_{\mathbb{R}^d} z\pi_x(dz) = x \neq \int_{\mathbb{R}^d} z\gamma_x(dz) \right\}.$$

We remark that A and B_γ are Borel measurable sets since $x \mapsto \pi_x$ and $x \mapsto \gamma_x$ are measurable maps. Then, the inequality (23) reads

$$(26) \quad - \int_A \left(\frac{x - \int_{\mathbb{R}^d} z\pi_x(dz)}{\|x - \int_{\mathbb{R}^d} z\pi_x(dz)\|} \right)^\top \left(\int_{\mathbb{R}^d} y[\gamma_x - \pi_x](dy) \right) \mu(dx) \\ + \int_{B_\gamma} \left\| \int_{\mathbb{R}^d} y[\gamma_x - \pi_x](dy) \right\| \mu(dx) \geq 0$$

or yet

$$(27) \quad - \int_{A \times \mathbb{R}^d} \left(\frac{x - \int_{\mathbb{R}^d} z\pi_x(dz)}{\|x - \int_{\mathbb{R}^d} z\pi_x(dz)\|} \right)^\top y[\gamma - \pi](dx, dy) \geq \\ - \int_{B_\gamma} \left\| \int_{\mathbb{R}^d} z\gamma_x(dz) - x \right\| \mu(dx)$$

We have the following theorem:

Theorem 5. *Let $\pi \in \Pi(\mu, \nu)$ be an optimal transport plan for the weak optimal transport (WOT) with barycentric cost, i.e. $C(x, \rho) = \|x - \int_{\mathbb{R}^d} z\rho(dz)\|$. Let us also assume that $\int_{\mathbb{R}^d} \|y\|\nu(dy) < \infty$. Then, there exists a measurable set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ such that $\pi(\Gamma) = 1$ and so that, for all $n \in \mathbb{N}$ and all points $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$, we have*

$$(28) \quad \boxed{\sum_{i|x_i \neq b_i} \frac{(y_{i+1} - y_i)^\top (b_i - x_i)}{\|b_i - x_i\|} \geq - \sum_{i|x_i = b_i} \|y_{i+1} - y_i\|}$$

where we let $b_i := \int_{\mathbb{R}^d} z\pi_{x_i}(dz)$ and where we let $y_{n+1} := y_1$.

This statement and the inequality (28) should be thought as an analogue of c -cyclical monotonicity for the weak optimal transport problem (WOT) in the specific case of the barycentric cost. The proof is in essence similar to that of Theorem 1, although with some technical differences:

Proof of Theorem 5. Let $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$ where $\Gamma \subset X \times Y$ is a Borel set so that $\pi(\Gamma) = 1$. This set is to be defined at the end of the proof by successive refinements, similar to the proof of Theorem 1. For each $i = 1, \dots, n$, let $B_i^\varepsilon \subset \mathbb{R}^d \times \mathbb{R}^d$ the “ball” of radius $\varepsilon > 0$ centred at $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}^d$. More precisely, to simplify a bit the presentation, assume that $B_i^\varepsilon := B_\varepsilon(x_i) \times B_\varepsilon(y_i)$ where $B_\varepsilon(x) := \{z \in \mathbb{R}^d : \|z - x\| < \varepsilon\}$. We may (and will) assume without loss of generality — up to refining Γ — that $\pi(B_i^\varepsilon) > 0$ for all $\varepsilon > 0$ and all $i = 1, \dots, n$. Let σ_i^ε be the restriction of π on B_i^ε normalised to a probability measure, so that $\sigma_i^\varepsilon(A) := \pi(B_i^\varepsilon)^{-1} \pi(A \cap B_i^\varepsilon)$ for all Borel set $A \subset \mathbb{R}^d \times \mathbb{R}^d$, and let $\sigma^\varepsilon := \sum_{i=1}^n \sigma_i^\varepsilon$. We then define $\xi_i^\varepsilon := \text{pr}_1^\# \sigma_i^\varepsilon \otimes \text{pr}_2^\# \sigma_{i+1}^\varepsilon$ and $\xi^\varepsilon := \sum_{i=1}^n \xi_i^\varepsilon$. Letting $\gamma^\varepsilon := \pi + t(\xi^\varepsilon - \sigma^\varepsilon)$, then $\gamma^\varepsilon \in \Pi(\mu, \nu)$ for all $\varepsilon > 0$ and any $0 < t < t_\varepsilon$ where $t_\varepsilon := \|\frac{d\sigma^\varepsilon}{d\pi}\|_{L^\infty(\pi)}$. We apply the inequality (27) to $\gamma \leftarrow \gamma^\varepsilon$, which gives

$$(29) \quad - \int_{A \times \mathbb{R}^d} \frac{(x - \int_{\mathbb{R}^d} z \pi_x(dz))^\top y}{\|x - \int_{\mathbb{R}^d} z \pi_x(dz)\|} [\gamma^\varepsilon - \pi](dx, dy) \geq \\ - \int_{B_{\gamma^\varepsilon}} \left\| \int_{\mathbb{R}^d} z \gamma_x^\varepsilon(dz) - x \right\| \mu(dx)$$

where we recall that the measurable sets A and B_{γ^ε} are defined by

$$A = \left\{ x \in \mathbb{R}^d : x \neq \int_{\mathbb{R}^d} z \pi_x(dz) \right\}$$

and

$$(30) \quad B_{\gamma^\varepsilon} = \left\{ x \in A^c : x \neq \int_{\mathbb{R}^d} z \gamma^\varepsilon(dz) \right\}.$$

The above inequality (29) further develops to

$$(31) \quad - \int_{A \times \mathbb{R}^d} \frac{(x - \int_{\mathbb{R}^d} z \pi_x(dz))^\top y}{\|x - \int_{\mathbb{R}^d} z \pi_x(dz)\|} [\xi^\varepsilon - \pi^\varepsilon](dx, dy) \geq \\ - \int_{B_{\gamma^\varepsilon}} \left\| \int_{\mathbb{R}^d} z [\xi_x^\varepsilon - \pi_x^\varepsilon](dz) \right\| \mu(dx)$$

Here, $(\xi_x^\varepsilon)_{x \in \mathbb{R}^d}$ and $(\pi_x^\varepsilon)_{x \in \mathbb{R}^d}$ are the disintegrations of ξ^ε and π^ε with respect to μ . We have $\pi_x^\varepsilon = \sum_{i=1}^n \sigma_{i,x}^\varepsilon$ where for all $i = 1, \dots, n$ the disintegration $(\sigma_{i,x}^\varepsilon)_{x \in X}$ of σ_i^ε with respect to μ reads for $x \in B_\varepsilon(x_i)$

$$(32) \quad \sigma_{i,x}^\varepsilon(dy) = \frac{1}{\pi(B_i^\varepsilon)} \pi_x \upharpoonright_{B_\varepsilon(y_i)}(dy)$$

— where $\pi_x \upharpoonright_{B_\varepsilon(y_i)}$ denotes the restriction of π_x to the ball $B_\varepsilon(y_i)$ — and $\sigma_{i,x}^\varepsilon = 0$ for $x \notin B_\varepsilon(x_i)$ otherwise. Likewise, $\xi^\varepsilon = \sum_{i=1}^n \xi_{i,x}^\varepsilon$ where the

disintegration $(\xi_{i,x}^\varepsilon)_{x \in X}$ of ξ_i^ε with respect to μ reads for $x \in B_\varepsilon(x_i)$

$$(33) \quad \xi_{i,x}^\varepsilon(dy) = \frac{1}{\pi(B_i^\varepsilon)} \frac{\pi_x(B_\varepsilon(y_i))}{\pi(B_{i+1}^\varepsilon)} \int_{B_\varepsilon(x_{i+1})} \pi \upharpoonright_{\mathbb{R}^d \times B_\varepsilon(y_{i+1})}(dx', y)$$

and $\xi_{i,x}^\varepsilon = 0$ for $x \notin B_\varepsilon(x_i)$. To show that (28) holds, we are going to take the limit $\varepsilon \rightarrow 0$ in the inequality (31).

First, we note that, by the standard *Lebesgue differentiation theorem* and just like in the proof of Theorem 1, it holds that — up to refining the set Γ :

$$(34) \quad - \int_{A \times \mathbb{R}^d} \frac{(x - \int_{\mathbb{R}^d} z \pi_x(dz))^\top y}{\|x - \int_{\mathbb{R}^d} z \pi_x(dz)\|} [\xi^\varepsilon - \pi^\varepsilon](dx, dy) \\ \xrightarrow{\varepsilon \rightarrow 0} \sum_{i|x_i \neq b_i} \frac{(y_{i+1} - y_i)^\top (b_i - x_i)}{\|b_i - x_i\|}$$

where we recall the notation $b_i := \int_{\mathbb{R}^d} z \pi_{x_i}(dz)$ for all $i = 1, \dots, n$. This is exactly the left-hand side of the sought-for inequality (28). Now, it remains to prove that we can recover the right-hand side of (28) from the right-hand side of (31) in the limit $\varepsilon \rightarrow 0$.

Let us first remark that the right-hand side rewrites as

$$(35) \quad \int_{B_{\gamma^\varepsilon}} \left\| \int_{\mathbb{R}^d} z [\xi_x^\varepsilon - \pi_x^\varepsilon](dz) \right\| \mu(dx) \\ = \sum_{i=1}^n \int_{B_{\gamma^\varepsilon} \cap B_\varepsilon(x_i)} \left\| \int_{\mathbb{R}^d} z [\xi_x^\varepsilon - \pi_x^\varepsilon](dz) \right\| \mu(dx)$$

Indeed, here we used that $x \mapsto \xi_x^\varepsilon$ and $x \mapsto \pi_x^\varepsilon$ are supported onto the balls $B_\varepsilon(x_i)$'s. Using the explicit formula for the disintegrations given above, we have

$$(36) \quad \int_{B_{\gamma^\varepsilon} \cap B_\varepsilon(x_i)} \left\| \int_{\mathbb{R}^d} z [\xi_x^\varepsilon - \pi_x^\varepsilon](dz) \right\| \mu(dx) = \\ \frac{1}{\pi(B_i^\varepsilon)} \int_{B_i^\varepsilon} \mathbf{1}_{B_{\gamma^\varepsilon}}(x) \left\| \frac{1}{\pi_x(B_\varepsilon(y_i))} \int_{B_\varepsilon(y_i)} z \pi_x(dz) \right. \\ \left. - \frac{1}{\pi(B_{i+1}^\varepsilon)} \int_{B_{i+1}^\varepsilon} y \pi(dx', dy) \right\| \pi(dx, dy)$$

For all $\varepsilon > 0$ (small enough) we have

$$(37) \quad \|y_{i+1} - y_i\| - 2\varepsilon \\ \leq \left\| \frac{1}{\pi_x(B_\varepsilon(y_i))} \int_{B_\varepsilon(y_i)} z \pi_x(dz) - \frac{1}{\pi(B_{i+1}^\varepsilon)} \int_{B_{i+1}^\varepsilon} y \pi(dx', dy) \right\| \\ \leq \|y_{i+1} - y_i\| + 2\varepsilon$$

and therefore

$$\begin{aligned}
(38) \quad & (\|y_{i+1} - y_i\| - 2\varepsilon) \frac{1}{\pi(B_i^\varepsilon)} \int_{B_i^\varepsilon} \mathbb{1}_{B_{\gamma\varepsilon}}(x) \pi(dx, dy) \\
& \leq \int_{B_{\gamma\varepsilon} \cap B_\varepsilon(x_i)} \left\| \int_{\mathbb{R}^d} z [\xi_x^\varepsilon - \pi_x^\varepsilon](dz) \right\| \mu(dx) \\
& \leq (\|y_{i+1} - y_i\| + 2\varepsilon) \frac{1}{\pi(B_i^\varepsilon)} \int_{B_i^\varepsilon} \mathbb{1}_{B_{\gamma\varepsilon}}(x) \pi(dx, dy)
\end{aligned}$$

We now claim that there exists $\bar{\varepsilon} > 0$ small enough so that we have the set inclusion $B_{\gamma\varepsilon} \subset B_{\gamma\bar{\varepsilon}}$ for all $\varepsilon < \bar{\varepsilon}$. Indeed, let us remark that the set $B_{\gamma\varepsilon}$ rewrites as the set of points $x \in X$ so that $x = \int_{\mathbb{R}^d} z \pi_x(dz)$ and so that $\int_{\mathbb{R}^d} z \xi_x^\varepsilon(dz) \neq \int_{\mathbb{R}^d} z \sigma_x^\varepsilon(dz)$. This last condition rewrites, using the disintegrations, as

$$(39) \quad \frac{1}{\pi_x(B_\varepsilon(y_i))} \int_{B_\varepsilon(y_i)} z \pi_x(dz) \neq \frac{1}{\pi(B_{i+1}^\varepsilon)} \int_{B_{i+1}^\varepsilon} y \pi(dx', dy)$$

We note that, as $\varepsilon \rightarrow 0$, this condition reads (formally for the moment) $y_i \neq y_{i+1}$. Of course, we may assume without loss of generality that $y_i \neq y_{i+1}$. Indeed, if it were that $y_i = y_{i+1}$, then

$$(40) \quad \int_{B_{\gamma\varepsilon} \cap B_\varepsilon(x_i)} \left\| \int_{\mathbb{R}^d} z [\xi_x^\varepsilon - \pi_x^\varepsilon](dz) \right\| \mu(dx) \leq \|y_{i+1} - y_i\| + 2\varepsilon = 2\varepsilon$$

and thus, in the limit $\varepsilon \rightarrow 0$, this integral does not contribute to the sum since it vanishes to zero. Therefore, only the case where $y_{i+1} \neq y_i$ is of importance. We note passing by that the previous result implies trivially that, in the limit $\varepsilon \rightarrow 0$, the inequality (31) implies

$$(41) \quad \sum_{i|x_i \neq b_i} \frac{(y_{i+1} - y_i)^\top (b_i - x_i)}{\|b_i - x_i\|} \geq - \sum_{i=1}^n \|y_{i+1} - y_i\|.$$

But the sought-for inequality (28) is finer since only the indices $i \in \{1, \dots, n\}$ where $x_i = b_i$ are considered in the sum on the right-hand side. Let us therefore assume that $y_i \neq y_{i+1}$ for all $i = 1, \dots, n$. If $x \in B_{\gamma\varepsilon}$, and using that $y_{i+1} \neq y_i$, then there exists a small enough $\bar{\varepsilon} > 0$ such that (39) holds for all $0 < \varepsilon < \bar{\varepsilon}$. Indeed, this follows from the obvious facts that

$$\left\| \frac{1}{\pi_x(B_\varepsilon(y_i))} \int_{B_\varepsilon(y_i)} z \pi_x(dz) - y_i \right\| \leq \varepsilon$$

and

$$\left\| \frac{1}{\pi(B_{i+1}^\varepsilon)} \int_{B_{i+1}^\varepsilon} y \pi(dx', dy) - y_{i+1} \right\| \leq \varepsilon$$

for all $\varepsilon > 0$. Therefore, letting for instance $\bar{\varepsilon} < \frac{\|y_{i+1} - y_i\|}{2}$, we have the set inclusion $B_{\gamma\varepsilon} \subset B_{\gamma\bar{\varepsilon}}$ for all $0 < \varepsilon < \bar{\varepsilon}$.

Using this fact, we obtain

$$(42) \quad \limsup_{\varepsilon \rightarrow 0} \int_{B_{\gamma\varepsilon} \cap B_\varepsilon(x_i)} \left\| \int_{\mathbb{R}^d} z[\xi_x^\varepsilon - \pi_x^\varepsilon](dz) \right\| \mu(dx) \\ \leq \limsup_{\varepsilon \rightarrow 0} \frac{\|y_{i+1} - y_i\| + 2\varepsilon}{\pi(B_i^\varepsilon)} \int_{B_i^\varepsilon} \mathbb{1}_{B_{\gamma\bar{\varepsilon}}}(x) \pi(dx, dy)$$

Now, according to the Lebesgue differentiation theorem, we know that there exists a set $E \subset \mathbb{R}^d$ so that $\mu(E) = 1$ and so that

$$\frac{1}{B_\varepsilon(x)} \int_{B_\varepsilon(x)} \mathbb{1}_{B_{\gamma\bar{\varepsilon}}}(x') \mu(dx') \xrightarrow{\varepsilon \rightarrow 0} \mathbb{1}_{B_{\gamma\bar{\varepsilon}}}(x)$$

for all $x \in E$. **If we could ensure that $x_i \in E$ for all $i = 1, \dots, n$** , then we would obtain

$$\limsup_{\varepsilon \rightarrow 0} \int_{B_{\gamma\varepsilon} \cap B_\varepsilon(x_i)} \left\| \int_{\mathbb{R}^d} z[\xi_x^\varepsilon - \pi_x^\varepsilon](dz) \right\| \mu(dx) \leq \|y_{i+1} - y_i\| \mathbb{1}_{B_{\gamma\bar{\varepsilon}}}(x_i)$$

and thus the claimed inequality (28) would immediately follow. Let us discuss how to ensure that this happens.

First, beware that the measurable set $E \equiv E[n, \mathcal{X}, \bar{\varepsilon}]$ actually depends on the function $\mathbb{1}_{B_{\gamma\bar{\varepsilon}}}$ which itself depends on n , the family of points $\mathcal{X} := \{(x_i, y_i) : i = 1, \dots, n\}$ under consideration, and $\bar{\varepsilon} > 0$. To simplify the presentation, let us write $B_{\gamma\bar{\varepsilon}} = B[n, \mathcal{X}, \bar{\varepsilon}]$ as well. The idea is then: since any subspace of a metrisable separable space is itself separable, the support $\text{supp}(\pi)$ of the transport plan π is separable. Let us then consider $S \subset \text{supp}(\pi)$ to be a dense and countable subset of $\text{supp}(\pi)$. For each $n \in \mathbb{N}$, the set $\mathbb{X}_n = \{\mathcal{X}\}$ of all families $\mathcal{X} = \{(x_1, y_1), \dots, (x_n, y_n)\} \subset S$ of n elements inside S is therefore itself a countable set as a finite product of countable sets. Then, for each $\bar{\varepsilon} > 0$, consider the set

$$E^{\bar{\varepsilon}} := \bigcap_{n \in \mathbb{N}} \bigcap_{\mathcal{X} \in \mathbb{X}_n} E[n, \mathcal{X}, \bar{\varepsilon}]$$

The set $E^{\bar{\varepsilon}}$ is measurable and $\mu(E^{\bar{\varepsilon}}) = 1$. Then, we finally consider

$$E := \bigcap_{m \in \mathbb{N}} E^{1/m}$$

The set E is again measurable and $\mu(E) = 1$. We then consider $\Gamma := (E \times \mathbb{R}^d) \cap \Gamma_0$ where Γ_0 is the π -full measure set over which the convergence ensured by the Lebesgue differentiation theorem holds in (34). We claim that the sought-for inequality (28) holds over Γ . This follows from the fact, for any $n \in \mathbb{N}$ and any family

$$\mathcal{X} = \{(x_1, y_1), \dots, (x_n, y_n)\} \subset \Gamma \subset (E \times \mathbb{R}^d),$$

(where *a priori* $\mathcal{X} \not\subset S$) and for all $\eta > 0$, there exists a family $\mathcal{X}' \in \mathbb{X}_n$ which is η -close to \mathcal{X} , in the sense that, letting

$$\mathcal{X}' = \{(x'_1, y'_1), \dots, (x'_n, y'_n)\},$$

we have $\|(x_i, y_i) - (x'_i, y'_i)\| < \eta$ for all $i = 1, \dots, n$. Now, let us consider $\eta = \varepsilon/2$. Then, by similar arguments to that of preceding, we have that $B[n, \mathcal{X}, \varepsilon] \subset B[n, \mathcal{X}', \varepsilon/2]$. In particular, selecting $\bar{\varepsilon} = 1/m$ for $m \in \mathbb{N}$ big enough, we obtain

$$(43) \quad \frac{1}{\pi(B_i^\varepsilon)} \int_{B_i^\varepsilon} \mathbb{1}_{B[n, \mathcal{X}, \bar{\varepsilon}]}(x) \pi(dx, dy) \leq \frac{1}{\pi(B_i^\varepsilon)} \int_{B_i^\varepsilon} \mathbb{1}_{B[n, \mathcal{X}', \bar{\varepsilon}/2]}(x) \pi(dx, dy).$$

Finally, the right-hand side converges to $\mathbb{1}_{B[n, \mathcal{X}', \bar{\varepsilon}/2]}(x_i) \|y_{i+1} - y_i\|$ since $x_i \in E$. This terminates the proof. \square

The inequality (28) is to be thought as the analogue to the c -cyclical monotonicity property in the special case of the barycentric cost. Theorem 5 says if π is optimal, then it must verify this analogous property. A natural question is to go the other way around, that is to consider the following question: *Is this true that, if π verifies (28), then π is optimal for the weak optimal transport with barycentric cost?* It is rather straightforward that:

Proposition 5. *Let $\pi \in \Pi(\mu, \nu)$ has finite, i.e. $\mathcal{C}(\pi) < \infty$. Assume that μ almost-surely all $x \in \mathbb{R}^d$, we have $x \neq \int_{\mathbb{R}^d} z \pi_x(dz)$. Also assume that $\int_{\mathbb{R}^d} \|y\| \nu(dy) < \infty$. Then, if there exists $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$ so that $\pi(\Gamma) = 1$ and so that for all $n \geq 1$ and all $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$, we have*

$$(44) \quad \sum_{i=1}^n \frac{(y_{i+1} - y_i)^\top (b_i - x_i)}{\|b_i - x_i\|} \geq 0$$

where $b_i = \int_{\mathbb{R}^d} z \pi_{x_i}(dz)$, then π is optimal for the weak optimal transport problem (WOT) with barycentric cost.

Proof. Because for μ almost-surely all $x \in \mathbb{R}^d$, we have $x \neq \int_{\mathbb{R}^d} z \pi_x(dz)$, then the weak cost of transport C is differentiable for μ almost-surely all x with respect to its second variable and its gradient reads

$$\nabla_\rho C(x, \pi_x)(y) = \frac{(\int_{\mathbb{R}^d} z \pi_x(dz) - x)^\top y}{\|\int_{\mathbb{R}^d} z \pi_x(dz) - x\|} =: c(x, y)$$

To show that π is optimal, we must show that the first-order optimality condition holds, that is

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} c(x, y) [\gamma - \pi](dx, dy) \geq 0$$

for all $\gamma \in \Pi(\mu, \nu)$ so that $\mathcal{C}(\gamma) < \infty$, **provided it holds for cyclical perturbations** $\eta \in \mathcal{C}_\pi$ — **as is the essence of (44)**. This will follow by applying (the proof of) Theorem 4. We therefore only need to check that the “cost” c defined above verifies the assumptions of this theorem. This function is real-valued and clearly Borel measurable — by measurability of $x \mapsto \pi_x$, as already discussed. A bit of discussion is to be made regarding the range of values that the function c above takes. We stated Theorem 4 under the assumption that c be non-negative. But this was mainly for convenience,

and the theorem accommodates to slightly weaker assumptions, as already discussed above — see the *footnote* at p. 19. In particular, we have

$$c(x, y) \geq -\|y\| := g(y)$$

and g is integrable with respect to ν since ν is assumed to have finite first-moment. This terminates the proof. \square

In the case where there are points $x \in X$ so that $x = \int_{\mathbb{R}^d} z \pi_x(dz)$ (with positive μ -measure) the situation is slightly more involved, since to obtain optimality we need to be able to select a family of (extended) subgradients so that (22) holds. Altogether, there is *a priori* no reasons to obtain (28) by some limiting arguments as a sufficient criterion for optimality (which was obtained using a specific kind of subgradients, *i.e.* the one that realises the *supremum* in (23)). We could certainly explore more in detail this approach, but since this section is mainly intended to explain briefly what could be said in the non-differentiable case without going to much in details, we do not comment further on this point.

3.5. A comment on the entropy-regularised problem. When the weak cost of transport $C : X \times \mathcal{P}(Y) \rightarrow \mathbb{R}_+$ is (strictly) convex with respect to its second-argument, this may push the (unique) solution of (WOT) inside the relative interior of $\Pi(\mu, \nu)$ compared to the (OT) problem whose solution lives on the boundary of the constraint set (in fact, on extreme points). *In this case, the defining inequality in the C -cyclical monotonicity introduced in Definition 5 becomes an equality.* Indeed, the relative interior of the set of transport plan is made of those plans with full support, *i.e.* $\text{supp}(\pi) = X \times Y$. This means that if $\eta \in \mathcal{R}_\pi$ is an admissible perturbation, then so is its opposite $-\eta$. This is for instance the case for the *entropy-regularised optimal transport* problem, which we recall to be an instance of the weak optimal transport problem with

$$C(x, \rho) := \varepsilon \int_Y \ln \rho d\rho + \int_Y c(x, y) \rho(dy)$$

where $\varepsilon > 0$ is the regularisation parameter. The weak cost of transport defined as such is differentiable in its second variable and its gradient (in the sense of Section 3.1.1) reads

$$\nabla_\rho C(x, \rho)(y) = c(x, y) + \varepsilon \ln \rho(y) + \varepsilon.$$

As it is well-known, the unique solution to the entropic optimal transport lives in the relative interior of $\Pi(\mu, \nu)$. But then, we obtain that for all $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$ where $\Gamma \subset X \times Y$ is some set over which π concentrates, we have

$$(45) \quad \boxed{\sum_{i=1}^n c(x_i, y_i) + \varepsilon \sum_{i=1}^n \ln \pi_{x_i}(y_i) = \sum_{i=1}^n c(x_i, y_{i+1}) + \varepsilon \sum_{i=1}^n \ln \pi_{x_i}(y_{i+1})}$$

where as usual $y_{n+1} := y_1$. **The equality (45) is then the analogue of the cyclical monotonicity property for the entropic optimal transport problem.** This equality is not at all shocking, since using the well-known duality theory for the entropic optimal transport problem, we know that the optimiser π^* is unique and given by the Gibbs measure

$$\pi^*(x, y) = \exp\left(-\frac{1}{\varepsilon}(c(x, y) - f(x) - g(y))\right)$$

for some entropic Kantorovich potentials $f : X \rightarrow \mathbb{R}$ and $g : Y \rightarrow \mathbb{R}$. In particular, the equality (45) is in fact immediately verified.

APPENDIX A. PROOF OF LEMMA 2

In this appendix, we prove Lemma 2 whose statement is repeated below for convenience of the readers. It is not unlikely that this result is known and common knowledge to some audiences, but as it is we were not able to find any reference for it, so that we prove it here.

Lemma 5. *Let $\gamma \in \mathcal{P}(Z)$ where Z is a Polish space (or in fact only second countable) equipped with its Borel σ -algebra. Then, for all $z \in Z$, there exists a sequence $(E_k(z))_{k \in \mathbb{N}}$ of Borel sets — all containing z — so that $\gamma(E_k(z)) > 0$ for γ almost-surely all $z \in Z$ and all $k \in \mathbb{N}$ and so that for any measurable function $\phi : Z \rightarrow \mathbb{R}$ which is integrable with respect to γ , it holds that*

$$(46) \quad \frac{1}{\gamma(E_k(z))} \int_{E_k(z)} \phi(w) \gamma(dw) \xrightarrow{k \rightarrow \infty} \phi(z)$$

for γ -a.s. all point $z \in Z$.

Proof of Lemma 2. As a Polish space, Z is second countable, meaning that its topology has a countable base $(U_k)_{k \in \mathbb{N}}$. Let us define $\mathcal{F}_k := \sigma(U_1, \dots, U_k)$ the σ -algebra generated by U_1, \dots, U_k , so that $\mathcal{B}(Z) = \sigma(\bigcup_{k \geq 1} \mathcal{F}_k)$ where $\mathcal{B}(Z)$ is the Borel σ -algebra on Z . Since \mathcal{F}_k is finitely generated by k elements, it has at most 2^k atoms $e_1, e_2, \dots, e_{n_k} \in \mathcal{F}_k$. Then, let us define $E_k(z)$ for all $z \in Z$ to be the atom in \mathcal{F}_k that contains z . We remark that for γ almost-surely all $z \in Z$, we have $\gamma(E_k(z)) > 0$. Indeed, if $A_k := \{z \in Z : \gamma(E_k(z)) = 0\}$, then $A_k = \bigcup_{i \in I_k} e_i$ where $I_k \subset \{1, \dots, n_k\}$ is the set of indices so that $i \in I_k$ if $\gamma(e_i) = 0$. Since the atoms are disjoint by definition, we have $\gamma(A_k) = \sum_{i \in I_k} \gamma(e_i) = 0$. By considering the measurable set $A = \bigcup_{k \in \mathbb{N}} A_k \in \mathcal{B}(Z)$, we obtain that for γ almost-surely all $z \in Z$ and for all $k \in \mathbb{N}$ we have $\gamma(E_k(z)) > 0$. Then, for any function $\phi : Z \rightarrow \mathbb{R} \cup \{+\infty\}$ which is integrable with respect to γ , we have that $\mathbb{E}[\phi | \mathcal{F}_k]$ is constant and precisely given by

$$(47) \quad \mathbb{E}[\phi | \mathcal{F}_k] = \frac{1}{\gamma(E_k(z))} \int_{E_k(z)} \phi(w) \gamma(dw)$$

Now, it is a known fact that $\mathbb{E}[\phi | \mathcal{F}_k](z) \rightarrow \phi(z)$ as $k \rightarrow \infty$ for γ -a.s. all point $z \in Z$. Indeed, this follows from *Levy upwards theorem* — which is

a peculiar case of the celebrated Doob's martingale convergence theorems, see *e.g.* [11]. \square

Remark 7. We remark that the exact same construction yields that for all $\Phi \in L^1(\gamma \otimes \gamma)$, it holds that

$$(48) \quad \frac{1}{\gamma \otimes \gamma(E_k(z) \times E_k(t))} \int_{E_k(z) \times E_k(t)} \Phi d\gamma \otimes \gamma \xrightarrow[k \rightarrow \infty]{} \Phi(z, t)$$

for $\gamma \otimes \gamma$ -a.s. all couples $(z, t) \in Z \times Z$. Here, we emphasise that we can consider the same E_k 's are constructed above. Indeed, one replaces in the above proof \mathcal{F}_k by the finitely-generated σ -algebra

$$\mathcal{G}_k := \sigma(\{U_i \times U_j \mid i, j = 1, \dots, k\})$$

(which in fact is the product σ -algebra $\mathcal{F}_k \otimes \mathcal{F}_k$). The atoms of \mathcal{G}_k are the product of the atoms of \mathcal{F}_k , and the claim follows.

Acknowledgements. We acknowledge the financial support of European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme – Grant Agreement n°101077204 High-LEAP. L.N. benefited from the support of the FMJH Program PGMO and from the ANR project GOTA (ANR-23-CE46-0001).

REFERENCES

- [1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Basel: Birkhäuser, 2008.
- [2] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. “Existence, Duality, and Cyclical Monotonicity for Weak Transport Costs”. In: *Calculus of Variations and Partial Differential Equations* 58.6 (2019), p. 203.
- [3] J. Backhoff-Veraguas and G. Pammer. “Stability of Martingale Optimal Transport and Weak Optimal Transport”. In: *The Annals of Applied Probability* 32.1 (2022), pp. 721–752.
- [4] Mathias Beiglböck. “Cyclical Monotonicity and the Ergodic Theorem”. In: *Ergodic Theory and Dynamical Systems* 35.3 (2015), pp. 710–713.
- [5] Mathias Beiglböck and Claus Griessler. “A Land of Monotone Plenty”. In: *Annali Scuola Normale Superiore - Classe di Scienze* (2019), pp. 109–127.
- [6] Mathias Beiglböck and Nicolas Juillet. “On a Problem of Optimal Transport under Marginal Martingale Constraints”. In: *The Annals of Probability* 44.1 (2016), pp. 42–106.
- [7] Mathias Beiglböck et al. “Optimal and Better Transport Plans”. In: *Journal of Functional Analysis* 256.6 (2009), pp. 1907–1927.
- [8] John J. Benedetto and Wojciech Czaja. *Integration and Modern Analysis*. Boston: Birkhäuser, 2009.

- [9] Stefano Bianchini and Laura Caravenna. “On Optimality of C-Cyclically Monotone Transference Plans”. In: *Comptes Rendus. Mathématique* 348.11–12 (2010), pp. 613–618.
- [10] Stefano Bianchini and Laura Caravenna. “On the Extremality, Uniqueness and Optimality of Transference Plans”. In: *Bulletin of the Institute of Mathematics Academia Sinica* 4.4 (2009), pp. 353–454.
- [11] Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, 1995.
- [12] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. New York: Springer, 2011.
- [13] Hans Crauel. *Random Probability Measures on Polish Spaces*. CRC Press, 2002. 138 pp.
- [14] L. De Pascale, A. Kausamo, and K. Wyczesany. “60 Years of Cyclic Monotonicity: A Survey”. In: *Journal of Convex Analysis* 32.2 (2025), pp. 399–430.
- [15] R. M. Dudley. *Real Analysis and Probability*. 2nd ed. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2002.
- [16] Herbert Federer. *Geometric Measure Theory*. Ed. by B. Eckmann and B. L. Van Der Waerden. Classics in Mathematics. Berlin, Heidelberg: Springer, 1996.
- [17] Olav Kallenberg. *Foundations of Modern Probability*. Vol. 99. Probability Theory and Stochastic Modelling. Cham: Springer International Publishing, 2021.
- [18] Luigi De Pascale and Anna Kausamo. “Sufficiency of c-cyclical monotonicity in a class of multi-marginal optimal transport problems”. In: *European Journal of Applied Mathematics* 36.1 (2025), pp. 68–81.
- [19] V. S. Varadarajan. “Weak Convergence of Measures on Separable Metric Spaces”. In: *Sankhyā: The Indian Journal of Statistics (1933-1960)* 19.1/2 (1958), pp. 15–22.
- [20] Cédric Villani. *Optimal Transport*. Grundlehren Der Mathematischen Wissenschaften. Berlin, Heidelberg: Springer, 2009.
- [21] D. A. Zaev. “On the Monge–Kantorovich Problem with Additional Linear Constraints”. In: *Mathematical Notes* 98.5 (2015), pp. 725–741.

CERMICS, ECOLE NATIONALE DES PONTS ET CHAUSSÉES, IPP, CNRS & INRIA
Email address: virginie.ehrlacher@enpc.fr

SAMM, UNIVERSITÉ PARIS 1 PANTHÉON-SORBONNE, PARIS, FRANCE.
Email address: rodrigue.lelotte@univ-paris1.fr

UNIVERSITÉ PARIS-SACLAY, CNRS, LABORATOIRE DE MATHÉMATIQUES D’ORSAY, PARMA,
 INRIA SACLAY, 91405, ORSAY, FRANCE & INSTITUT UNIVERSITAIRE DE FRANCE (I.U.F.)
Email address: luca.nenna@universite-paris-saclay.fr