



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

# FLORE

## Repository istituzionale dell'Università degli Studi di Firenze

### **60 years of cyclic monotonicity: a survey**

Questa è la versione Preprint (Submitted version) della seguente pubblicazione:

*Original Citation:*

60 years of cyclic monotonicity: a survey / A. Kausamo; L. De Pascale; K. Wyczesany. - ELETTRONICO. - (2023), pp. 1-30.

*Availability:*

This version is available at: 2158/1342271 since: 2023-11-01T13:54:53Z

*Terms of use:*

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (<https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf>)

*Publisher copyright claim:*

(Article begins on next page)

# 60 years of cyclic monotonicity: a survey

A. Kausamo, L. De Pascale, K. Wyczesany

*Dedicated to Prof. R. T. Rockafellar.*

## 1 Aims and scopes

The primary purpose of this note is to provide an instructional summary of the state of the art regarding cyclic monotonicity and related notions. We will also present how these notions are tied to optimality in the optimal transport (or Monge-Kantorovich) problem, even though cyclic monotonicity is simply a property of a set and can be studied on its own merit.

Cyclic monotonicity arose from the problem of characterising the subgradient of a convex function. In [26], Minty proved monotonicity, which can also be called 2-cyclic monotonicity, of the subgradient. To the best of our knowledge, the definition of cyclic monotonicity was first introduced in 1966 by Rockafellar in [31], where he proved that it characterises the subgradient.

At the beginning of the '90s, the study of cyclic monotonicity met the Monge-Kantorovich mass transport problem, and it is the connection between the two problems that brought out numerous developments and generalisations as well as many new problems. We will highlight these links and regard cyclic monotonicity as a geometric notion of optimality of a transport plan (see Section 2 for a review of the relevant notions).

The optimal transport or Monge-Kantorovich problem is widely known in the literature, and there are several extensive books that allow readers of any level to learn the basic as well as the advanced theory (e.g. see [2, 37, 35]). Let  $X$  and  $Y$  be topological spaces and let  $\mu$  and  $\nu$  be Borel probability measures on  $X$  and  $Y$ , respectively. We denote by  $p_Z$  the projection onto a subspace  $Z$ , and we set

$$\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(X \times Y) : p_X(\gamma) = \mu, p_Y(\gamma) = \nu\}.$$

This is the set of all *transport plans* between  $\mu$  and  $\nu$ , i.e. all probability measures  $\gamma$  on the product space  $X \times Y$  such that for any Borel sets  $A \subset X$ ,  $B \subset Y$  we have  $\gamma(A \times Y) = \mu(A)$  and  $\gamma(X \times B) = \nu(B)$ . Given a Borel cost function  $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  the Kantorovich formulation of the optimal transportation problem is to find a plan that minimises the total cost, that is

$$\min_{\gamma \in \Pi(\mu, \nu)} \int c(x, y) d\gamma. \quad (1)$$

It is intuitively clear that if a plan  $\gamma$  is optimal with respect to a cost  $c$ , then for any number  $m \in \mathbb{N}$  of points  $\{(x_i, y_i)\}_{i=1}^m \in \text{supp}(\gamma)$  their total cost, which in this case can be written as  $\sum_{i=1}^m c(x_i, y_i)$ , must be less than the total cost of any other pairing. This is exactly the concept behind *c-cyclic monotonicity* that will be studied in detail in the following section.

## 2 Cyclic monotonicity, potentials and generalizations

### 2.1 Monotonicity and cyclic monotonicity

Let  $X = Y = \mathbb{R}^n$ . Recall that a function  $f : X \rightarrow Y$  is monotone if for any  $(x_i, y_i), (x_j, y_j) \in \text{graph}(f)$  we have that

$$\langle x_i - x_j, y_i - y_j \rangle \geq 0.$$

This classical notion, also referred to as 2-monotonicity, can be generalised to  $N$ -point-cyclic monotonicity, and further to cyclic monotonicity, which we define below, and which was introduced by Rockafellar in [31]. Let us remark that Rockafellar considered a more general setting, where  $X$  is a topological vector space over  $\mathbb{R}$  and  $Y = X^*$  is its topological dual.

By  $\mathcal{S}_k$  we denote the set of all permutations on  $k$  elements.

**Definition 1.** A subset  $\Gamma \subset X \times Y$  is *cyclically monotone* if for all  $k \in \mathbb{N}$  and for all  $\{(x_i, y_i)\}_{i=1}^k \subset \Gamma$

$$\sum_{i=1}^k \langle x_i - x_{\sigma(i)}, y_i \rangle \geq 0,$$

holds for any permutation  $\sigma \in \mathcal{S}_k$ , where we set  $x_{k+1} := x_1$ .

By a basic theorem on permutations of  $k$  objects the above condition is equivalent to simply requiring that  $\sum_{i=1}^k \langle x_i - x_{i+1}, y_i \rangle \geq 0$  holds for all  $k \in \mathbb{N}$  and any  $k$ -tuple of points  $\{(x_i, y_i)\}_{i=1}^k \subset \Gamma$ . Moreover, note that cyclic monotonicity is a property of a set, and that if  $\Gamma$  is cyclically monotone then any subset of  $\Gamma$  has the same property. An important fact is that if  $\varphi : X \rightarrow [-\infty, +\infty]$  is a convex function then the subgradient, given by  $\partial\varphi = \{(x, y) : \varphi(z) - \varphi(x) \geq \langle y, z - x \rangle, \forall z\} \subset X \times Y$ , is cyclically monotone.

The celebrated Rockafellar's theorem states that the inverse is also true:

**Theorem 2** ([31, Theorem 1], see also [32, 33, Theorem 24.8]). *Let  $\Gamma \subset X \times Y$  be a cyclically monotone set. Then there exists a convex function  $\varphi : X \rightarrow \mathbb{R}$  such that  $\Gamma \subset \partial\varphi$ .*

We say that a function  $\varphi$  is a *potential* for  $\Gamma$  if its subgradient contains the set  $\Gamma$ . The proof of this theorem is constructive and we present it in the more general setting of Theorem 7.

Rockafellar's result may be further refined proving that  $\Gamma = \partial\varphi$  (so it is not a proper subset) if it is maximal (*i.e.* not properly contained in any cyclically monotone set). Maximal 2-cyclic monotone sets were studied, in particular, by Minty [26].

It turns out that cyclic monotonicity can be equivalently described using the square of the Euclidean norm, denoted by  $\|\cdot\|$ . Indeed, a set  $\Gamma \subset X \times Y$  is cyclically monotone if for all  $k \in \mathbb{N}$  and for all  $\{(x_i, y_i)\}_{i=1}^k \subset \Gamma$

$$\sum_{i=1}^k \|x_i - y_i\|^2 \leq \sum_{i=1}^k \|x_i - y_{i+1}\|^2.$$

Clearly, this can be seen by expanding the above expression (see also Remark 3). Further, this observation is also true if  $X$  is a Hilbert space  $H$  and  $Y = X^*$  with the canonical identification  $X^* = H$ , where we denote the norm in  $H$  by  $\|\cdot\|$ .

## 2.2 Extensions of cyclic monotonicity

Some natural extensions of the notion above will be of interest. The first one involves a more general cost function.

### 2.2.1 $c$ -cyclic monotonicity

Let  $X$  and  $Y$  be two sets and let

$$c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$$

be a *cost function*. We say that  $\Gamma \subset X \times Y$  is  *$c$ -cyclically monotone* if for all  $k \in \mathbb{N}$  and  $\{(x_i, y_i)\}_{i=1}^k \subset \Gamma$  and for any permutation  $\sigma \in \mathcal{S}_k$

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

Clearly, cyclic monotonicity corresponds to the choice  $c(x, y) = -\langle x, y \rangle$ , or equivalently  $c(x, y) = \|x - y\|^2/2$ , which we will refer to as the *quadratic cost*.

*Remark 3.* Note that, as in the case of the quadratic cost, adding 'non-mixing' terms to the cost does not influence  $c$ -cyclic monotonicity. More precisely, let  $c(x, y)$  be a cost function and define  $\tilde{c}(x, y) = c(x, y) + p(x) + q(y)$  for some functions  $p, q$ . Then a set is  $c$ -cyclically monotone if and only if it is  $\tilde{c}$ -cyclically monotone (see e.g. [24]).

It was shown in [30, 34] that if the cost  $c$  attains only finite values then, as in the case of cyclic monotonicity, a set is  $c$ -cyclically monotone if and only if it is contained in the graph of the  $c$ -subdifferential of a  $c$ -class function. Let us recall the relevant definitions.

**Definition 4.** The  $c$ -transform of a function  $\varphi : X \rightarrow [-\infty, +\infty)$  is a function  $\varphi^c : Y \rightarrow [-\infty, +\infty]$  given by

$$\varphi^c(y) = \inf_{x \in X} (c(x, y) - \varphi(x)).$$

Similarly, when  $\psi : Y \rightarrow [-\infty, +\infty)$ , we use the same notation to define

$$\psi^c(x) = \inf_{y \in Y} (c(x, y) - \psi(y)).$$

The  $c$ -transform is order-reversing (with respect to a pointwise order) and it is an involution on its “ $X$ -image” and “ $Y$ -image”, respectively. In many applications  $X = Y$  and  $c$  is symmetric, i.e.  $c(x, y) = c(y, x)$ , in which case the image

$$\mathcal{C} := \{\varphi : X \rightarrow [-\infty, +\infty) : \exists \psi : X \rightarrow [-\infty, +\infty) \text{ such that } \psi^c = \varphi\}$$

is called the  $c$ -class.

**Definition 5.** Given a function  $\varphi : X \rightarrow [-\infty, +\infty)$  its  $c$ -subgradient is given by

$$\partial^c \varphi = \{(x, y) \in X \times Y : c(x, y) < \infty \text{ and } c(x, y) - \varphi(x) \leq c(z, y) - \varphi(z) \text{ for all } z \in X\}.$$

*Remark 6.* There is an alternative way of defining the  $c$ -subgradient of a function  $\varphi : X \rightarrow [-\infty, +\infty)$ , that is

$$\tilde{\partial}^c \varphi = \{(x, y) \in X \times Y : \varphi(x) + \varphi^c(y) = c(x, y) < +\infty\}.$$

Since it is possible that  $\varphi(x) = -\infty$  and  $\varphi^c(y) = +\infty$ , some assumptions must be added to ensure that this definition is well-given. For example, if we know that

$$\text{There exists } x' \in X \text{ such that } c(x', y) \in \mathbb{R} \text{ and } \varphi(x') \in \mathbb{R},$$

then  $\varphi^c(y) \in \mathbb{R} \cup \{-\infty\}$  and  $(x, y) \in \partial^c \varphi$  if and only if  $(x, y) \in \tilde{\partial}^c \varphi$ .

Clearly, for a symmetric cost,  $(x, y) \in \partial^c \varphi$  if and only if  $(y, x) \in \partial^c \varphi^c$ .

The following theorem is the generalisation of Rockafellar's theorem due to Rochet and Rüschendorf.

**Theorem 7** ([30, Theorem 1], [34, Lemma 2.1]). *Let  $X, Y$  be any sets and let  $c : X \times Y \rightarrow \mathbb{R}$  be a cost function. A set  $\Gamma \subset X \times Y$  is  $c$ -cyclically monotone if and only if there exists a  $c$ -class function  $\varphi$  such that  $\Gamma \subset \partial^c \varphi$ , i.e.  $\varphi$  is a potential for  $\Gamma$ .*

*Proof.* Let  $\Gamma$  be a  $c$ -cyclically monotone set. Let  $(x_0, y_0) \in \Gamma$  and define  $\varphi : X \rightarrow \mathbb{R}$  as follows

$$\begin{aligned} \varphi(x) = \inf \{ & c(x, y_n) - c(x_n, y_n) + c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1}) + \cdots + \\ & c(x_1, y_0) - c(x_0, y_0) : (x_i, y_i) \in \Gamma \text{ for } i = 1, \dots, n, n \in \mathbb{N}. \} \end{aligned} \quad (2)$$

Since  $\Gamma$  is  $c$ -cyclically monotone,  $\varphi(x_0) = 0$  (taking  $n = 1$  and  $(x_1, y_1) = (x_0, y_0)$ ) and  $\varphi$  is a  $c$ -class function. To show that  $\Gamma \subset \partial^c \varphi$  consider  $(x', y') \in \Gamma$ , and  $\varepsilon > 0$  and let  $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$  be a chain such that

$$c(x', y_n) - c(x_n, y_n) + c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1}) + \cdots + c(x_1, y_0) - c(x_0, y_0) \leq \varphi(x') + \varepsilon.$$

Let  $z \in X$ , since  $(x', y') \in \Gamma$  we can consider the chain  $(x_1, y_1), \dots, (x_n, y_n), (x', y')$  among the admissible chain for the definition of  $\varphi(z)$  and get

$$\varphi(z) \leq c(z, y') - c(x', y') + \varphi(x') + \varepsilon.$$

Since  $\varepsilon$  is arbitrary this implies  $(x', y') \in \partial^c \varphi$ . Since  $\varphi(x_0) = 0$  and  $c$  assumes only real values, this last inequality, choosing  $(x', y') = (x_0, y_0)$ , also implies  $\varphi(z) < +\infty$ .  $\square$

The above theorem guarantees the existence of a potential for finite-valued cost functions. However, if the cost does attain the value  $+\infty$  this statement is no longer true. See, for instance, [5] where the authors present an example of a set that is  $c$ -cyclically monotone but whose graph is not supported on the  $c$ -subgradient of any function. In the attempt to use formula (2) above in a more general setting, in [9], the notion of  $c$ -connectivity is proposed and is shown to be sufficient to guarantee the existence of a potential.

**Definition 8.** Let  $X, Y$  be sets,  $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  be a cost function and  $\Gamma \subset X \times Y$ . We denote  $(x, y) \lesssim (\tilde{x}, \tilde{y})$  if there exist pairs  $(x_0, y_0), \dots, (x_n, y_n) \in \Gamma$  with  $(x_0, y_0) = (x, y)$  and  $(x_n, y_n) = (\tilde{x}, \tilde{y})$  such that  $c(x_1, y_0), \dots, c(x_n, y_{n-1}) < +\infty$ . If  $(x, y) \lesssim (\tilde{x}, \tilde{y})$  and  $(\tilde{x}, \tilde{y}) \lesssim (x, y)$ , we denote  $(x, y) \approx (\tilde{x}, \tilde{y})$ . We say that  $\Gamma$  is  $c$ -connecting if  $c(x, y)$  is finite for all  $(x, y) \in \Gamma$  and if for all  $(x, y), (\tilde{x}, \tilde{y}) \in \Gamma$  we have  $(x, y) \approx (\tilde{x}, \tilde{y})$ .

Note that  $\approx$  is an equivalence relation on  $X \times Y$ . If the cost function is real-valued then all sets are trivially connecting. Hence,  $c$ -connecting sets do not need to be  $c$ -cyclically monotone and so, on its own,  $c$ -connectivity does not capture optimality in a way that  $c$ -cyclic monotonicity does. However, if a  $c$ -connecting set  $\Gamma$  is also  $c$ -cyclically monotone then, for any cost function, there exists a  $c$ -potential which can be obtained using Rockafellar's approach. Indeed, the  $c$ -connectivity, together with  $c$ -cyclical monotonicity, implies that the function  $\varphi$  defined by the formula (2) is real-valued everywhere on the projection  $p_X(\Gamma)$ , which means that one can proceed with the proof as in the case of a real-valued cost function  $c$ .

However, in general, the  $c$ -connectivity of  $\Gamma$  is not a necessary condition for the existence of a potential, which can be seen on simple discrete examples. In [5], the authors show that for general costs  $c$ -path-boundedness is the notion equivalent to the existence of a potential. We discuss it in the next subsection.

### 2.2.2 $c$ -path-boundedness

When the cost function is allowed to take infinite values, an important case that has been considered on the case-by-case basis (e.g. see [10, 11, 12, 14, 20]), the  $c$ -cyclic monotonicity is not sufficient to guarantee the existence of a potential. It was shown in [5] that a natural condition to consider is  $c$ -path-boundedness.

**Definition 9.** Fix sets  $X, Y$  and  $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ . We say that  $\Gamma \subset X \times Y$  is  $c$ -path-bounded if  $c(x, y) < \infty$  for any  $(x, y) \in \Gamma$ , and for any  $(x, y) \in \Gamma$  and  $(z, w) \in \Gamma$ , there exists a constant  $M = M((x, y), (z, w)) \in \mathbb{R}$  such that for any  $m \in \mathbb{N}$  and any  $\{(x_i, y_i) : 1 \leq i \leq m\} \subset \Gamma$  we have

$$c(x, y) - c(x_1, y) + \sum_{i=1}^{m-1} (c(x_i, y_i) - c(x_{i+1}, y_i)) + c(x_m, y_m) - c(z, y_m) \leq M$$

A  $c$ -path-bounded set must be  $c$ -cyclically monotone; indeed, if there is some cycle for which the sum is positive, then, choosing one point on this cycle to be both the beginning and the end of a path, one can replicate it many times to get paths with arbitrarily large sums. Further, if the cost is finite-valued then  $c$ -path-boundedness and  $c$ -cyclic monotonicity are equivalent as we may choose  $M = c(x, w) - c(z, w) < +\infty$ .

As mentioned before, for costs attaining infinite values,  $c$ -connectivity together with  $c$ -cyclic monotonicity implies  $c$ -path-boundedness. However, a  $c$ -path bounded set is not always  $c$ -connecting. Indeed, consider, for example, any cost function  $c : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ , which is finite on the set  $S := \{(x, y) \in \mathbb{R}^2 : x + y > 0\}$  and  $+\infty$  otherwise. We may easily choose  $\Gamma$  to consist of two points  $(x_1, y_1), (x_2, y_2)$  such that  $(x_i, y_i) \in S$  for  $i = 1, 2$ , and which form a  $c$ -cyclically monotone set, but such that  $(x_1, y_2) \notin S$  which means that  $\Gamma$  is not  $c$ -connecting. But since there are only two points in  $\Gamma$  it is easy to check that it is  $c$ -path-bounded, and one can easily find a potential. For a much more elaborate example but with more details provided see Example 35 in Appendix A.

The following theorem established in [5] states that  $c$ -path-boundedness of a set is equivalent to the existence of a potential.

**Theorem 10** ([5, Theorem 1]). *Let  $X, Y$  be two arbitrary sets and  $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  an arbitrary cost function. For a given subset  $\Gamma \subset X \times Y$  there exists a  $c$ -class function  $\varphi : X \rightarrow [-\infty, +\infty]$  such that  $\Gamma \subset \partial^c \varphi$  if and only if  $\Gamma$  is  $c$ -path-bounded.*

The proof of this theorem differs greatly from the proofs in the case of finite-valued costs. The authors reformulated the problem of finding a  $c$ -potential for a given set as the problem of existence of a solution to a special family of linear inequalities. More precisely, the following theorem was proved in [5].

**Theorem 11** ([5, Theorem 3.1]). *Let  $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  be a cost function and let  $\Gamma \subset X \times Y$ . Then there exists a  $c$ -potential for  $\Gamma$ , namely a  $c$ -class function  $\varphi : X \rightarrow [-\infty, +\infty]$  such that  $\Gamma \subset \partial^c \varphi$  if and only if the following system of inequalities,*

$$c(x, y) - c(z, y) \leq \varphi(x) - \varphi(z), \quad (3)$$

*indexed by  $(x, y), (z, w) \in \Gamma$ , has a solution  $\varphi : p_X \Gamma \rightarrow \mathbb{R}$ , where  $p_X \Gamma = \{x \in X : \exists y \in Y, (x, y) \in \Gamma\}$ .*

Intuitively, the above theorem says the following: since for every  $(x_i, y_i) \in \Gamma$  letting  $\varphi_i(x) = c(x, y_i) - a_i$  we have that  $\partial^c \varphi_i(x_i) = y_i$  for any  $a_i \in \mathbb{R}$ , what one needs to take care of is the “good gluing” of such functions, i.e. the existence of constants  $a_i$  such that the function  $\varphi_a(x) := \min_{(x_i, y_i) \in \Gamma} \varphi_i(x)$  satisfies for every  $x_i \in p_X \Gamma$  that  $\varphi_a(x_i) = \varphi_i(x_i)$ . Therefore, finding a potential amounts to showing that there exists a solution  $a$  to the family of inequalities

$$c(x_i, y_i) - a_i \leq c(x_i, y_j) - a_j.$$

The authors then show that given  $\{\eta_{i,j}\}_{i,j \in I} \in [-\infty, +\infty)$ , where  $I$  is some arbitrary index set and with  $\eta_{i,i} = 0$ , the system of inequalities

$$\eta_{i,j} \leq a_i - a_j, \quad i, j \in I$$

has a solution if and only if for any  $i, j \in I$  there exists some constant  $M(i, j)$  such that for any  $m$  and any  $i_2, \dots, i_{m-1}$ , letting  $i = i_1$  and  $j = i_m$  one has that  $\sum_{k=1}^{m-1} \eta_{i_k, i_{k+1}} \leq M(i, j)$ . This directly translates into  $c$ -path-boundedness.

The proof of the solvability of the system of inequalities uses Zorn’s lemma. The authors also give a new proof of Theorem 7, which can be found in [39].

Having established the existence of a potential one may want to consider its regularity. This is a deep and hugely rich theory, which despite being related, is different to what we consider in this note. Therefore, we will not address these issues. We refer the interested reader to the second book of C. Villani [37], which offers a useful starting point and many further references.

### 2.2.3 Adding more variables

Finally, one may consider a *multi-dimensional* version of the same ideas, which to the best of our knowledge, was first introduced by Kim and Pass in [23]. Let  $X_1, \dots, X_N$  be sets and

$$c : X_1 \times \dots \times X_N \rightarrow \mathbb{R} \cup \{+\infty\}.$$

**Definition 12.** We say that  $\Gamma \subset X_1 \times \dots \times X_N$  is  *$c$ -cyclically monotone* if for all  $k \in \mathbb{N}$  and  $\{(x_1^i, \dots, x_N^i)\}_{i=1}^k \subset \Gamma$  and for all permutations  $\sigma_2, \dots, \sigma_N \in \mathcal{S}_k$ , we have

$$\sum_i c(x_1^i, x_2^i, \dots, x_N^i) \leq \sum_i c(x_1^i, x_2^{\sigma_2(i)} \dots, x_N^{\sigma_N(i)}).$$

In correspondence to the  $c$ -subgradient for two variables, for the multivariable case one considers  $c$ -splitting sets.

**Definition 13.** A set  $\Gamma \subset X_1 \times \dots \times X_N$  is called  $c$ -splitting if there exist  $N$  functions  $\varphi_i : X_i \rightarrow [-\infty, +\infty)$  such that

$$\varphi_1(x_1) + \varphi_2(x_2) + \dots + \varphi_N(x_N) \leq c(x_1, x_2, \dots, x_N)$$

holds for all  $(x_1, x_2, \dots, x_N) \in X_1 \times \dots \times X_N$ , and

$$\varphi_1(x_1) + \varphi_2(x_2) + \dots + \varphi_N(x_N) = c(x_1, x_2, \dots, x_N)$$

holds for all  $(x_1, x_2, \dots, x_N) \in \Gamma$ . We call the functions  $(\varphi_1, \dots, \varphi_N)$  a  $(\Gamma, c)$ -splitting tuple.

Similarly to the two variables case, it is not difficult to check that a  $c$ -splitting set is  $c$ -cyclically monotone. The reverse implication will be discussed in Section 4.1.

The existence of the  $(\Gamma, c)$ -splitting tuple for a given set  $\Gamma$  has been investigated in some special cases. In particular, Bartz, Bauschke and Wang [8, Theorem 2.7] showed that if the cost function  $c$  can be written in the form

$$c(x_1, \dots, x_N) = \sum_{1 \leq i < j \leq N} c_{i,j}(x_i, x_j),$$

where  $c_{i,j} : X_i \times X_j \rightarrow \mathbb{R}$ , then if for each  $1 \leq i < j \leq N$  the projection  $\Gamma_{i,j}$  of  $\Gamma \subset X_1 \times \dots \times X_N$  onto  $X_i \times X_j$  is  $c_{i,j}$ -cyclically monotone, then  $\Gamma$  is  $c$ -cyclically monotone, and further, there exist a  $c$ -splitting tuple  $(\varphi_1, \dots, \varphi_N)$  given by

$$\varphi_i(x_i) = \sum_{i < j \leq N} \psi_{i,j}(x_i) + \sum_{1 \leq j < i} \psi_{j,i}^{c_{j,i}}(x_i),$$

where for every  $1 \leq i < j \leq N$  we have  $\psi_{i,j} : X_i \rightarrow [-\infty, +\infty]$  is the  $c_{i,j}$ -potential for  $\Gamma_{i,j}$ .

*Remark 14.* This statement can be now generalized to allow  $c_{i,j} : X_i \times X_j \rightarrow \mathbb{R} \cup \{+\infty\}$  if we additionally assume that  $\Gamma_{i,j}$  is  $c_{i,j}$ -path-bounded since, as we have seen, this guarantees the existence of the functions  $\psi_{i,j}$ .

Clearly, the condition of  $c_{i,j}$ -cyclically monotone projections is stronger than  $c$ -cyclic monotonicity on the whole set  $\Gamma$ , and the authors present the relevant example in [8]. However, it turns out that if we additionally assume that  $X_1 = \dots = X_N = \mathbb{R}$  are one-dimensional, then the two conditions are equivalent.

Let us remark that the maximal multimarginal  $c$ -monotonicity for the “quadratic” multimarginal cost (in the spirit of Minty [26]) was considered in [7].

### 3 Classical optimal transport

In this section we collect results connecting  $c$ -cyclic monotonicity and its various extensions to optimality of a transport plan in the Monge-Kantorovich problem (1). Prior to the detailed exploration of the deep relationship between  $c$ -cyclic monotonicity and the optimality of transport plans, let us recall the classical setup of two marginal transport problem and justify why in this case one can expect a minimiser of the transport problem to exist.



Let  $X$  and  $Y$  be topological spaces and let  $\mu$  and  $\nu$  be Borel probability measures on  $X$  and  $Y$ , respectively. Let

$$c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$$

be a Borel cost function that may assume the value  $+\infty$ . We say that  $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  is *essentially bounded* with respect to  $\mu$  and  $\nu$  if there exist upper semi-continuous functions  $a : X \rightarrow (-\infty, +\infty]$ ,  $a \in L^1(\mu)$  and  $b : Y \rightarrow (-\infty, +\infty]$ ,  $b \in L^1(\nu)$  such that  $c(x, y) \geq a(x) + b(y)$  for all  $x \in X$ ,  $y \in Y$ . Since we may absorb the functions  $a$  and  $b$  into the cost, instead of writing that the cost is essentially bounded we will sometimes write that  $c : X \times Y \rightarrow [0, +\infty]$ . One needs this assumption to avoid the situation where there are plans with a total cost  $-\infty$  (and hence nothing can be inferred about the optimal plan).

If  $c$  is assumed lower semi-continuous, then also the ‘total cost’ functional

$$C[\gamma] := \int c(x, y) d\gamma$$

is lower semi-continuous. This fact, together with Prokhorov’s theorem [29] (see Th. 5.1.3 in [3]) that assures compactness of the set of all transport plans  $\Pi(\mu, \nu)$ , gives the existence of at least one minimiser of the transport problem (1). This is captured in the following theorem which can be found, for example, in [38].

**Theorem 15** ([38, Theorem 4.1]). *Let  $X, Y$  be two Polish spaces,  $\mu \in \mathcal{P}(X)$  and  $\nu \in \mathcal{P}(Y)$ . Let  $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  be a lower semi-continuous cost function which is essentially bounded with respect to  $\mu$  and  $\nu$ . Then there exists a  $c$ -optimal plan  $\pi \in \Pi(\mu, \nu)$ .*

In the above theorem, the existence of a plan with finite total cost is not assumed (nor guaranteed) and when no finite cost plan exists, any plan (say,  $\mu \otimes \nu$ ) is optimal. Therefore, in the results to follow, it will be assumed that there exists a plan with finite total cost. A transport plan will be called *finite* if its total cost is finite.

It is of interest to understand the structure of an optimal transport plan. It turns out that “often” a transport plan is optimal if and only if it is concentrated on a  $c$ -cyclically monotone set. There is a vast array of this type of results, which we outline in this section.

### 3.1 Optimality implies $c$ -cyclic monotonicity

Analysing the geometric structure of an optimal plan is a problem with a long history. It seems that  $c$ -cyclic monotonicity first appeared in this context in [24] in the following equivalent formulation of the Kantorovich’s problem

$$\min \left\{ \int c(\mathfrak{X}, \mathfrak{Y}) d\omega : \mathfrak{X} \text{ and } \mathfrak{Y} \text{ random variables with } Law(\mathfrak{X}) = \mu, Law(\mathfrak{Y}) = \nu \right\}. \quad (4)$$

The equivalence between problems (4) and (1) can be seen in the following way: If  $\mathfrak{X}$  and  $\mathfrak{Y}$  are admissible for (4) then the joint law  $\gamma = \mathfrak{X} \otimes \mathfrak{Y}_\# \omega$  is a transport plan in  $\Pi(\mu, \nu)$  and

$$\int c(x, y) d\gamma = \int c(\mathfrak{X}, \mathfrak{Y}) d\omega.$$

On the other hand if  $\gamma \in \Pi(\mu, \nu)$  then one can choose  $(\Omega, \omega) = (X \times Y, \gamma)$  and  $\mathfrak{X} = p_X$ ,  $\mathfrak{Y} = p_Y$ . In that context, which was partly motivated by some models appearing in financial mathematics,

authors started to characterize optimal random variables using  $c$ -cyclic monotonicity. In such a model,  $\mathfrak{X}$  and  $\mathfrak{Y}$  could represent the prices at a given moment of two assets or resources. In a robust financial framework the joint probability law of the portfolio  $(\mathfrak{X}, \mathfrak{Y})$  cannot be determined from market information, however, the distribution of each price is known. The cost  $c$  could be interpreted as an option whose pay-off is determined by the vector of prices  $(\mathfrak{X}, \mathfrak{Y})$  whose distribution is then of interest (maximizing or minimizing is, clearly, the same).

The classical structure of cyclic monotonicity of optimal plans was mentioned as a possible alternative tool in Brenier [13] and explicitly exploited in Caffarelli [15]. After that, Gangbo and McCann [19] extended the result to lower semi-continuous cost functions bounded from below. They showed that every finite optimal plan with respect to such a cost lies on a  $c$ -cyclically monotone set.

For more general settings there are, essentially, two arguments to prove that the support of the optimal plan must be  $c$ -cyclically monotone. The first one uses duality and appeared in [24], while the second relies on modifying a transport plan that is not  $c$ -cyclically monotone and showing that its cost can be improved. The latter technique was introduced in [1] and used, for example, in Proposition 2.24 of Villani<sup>1</sup> [38] for the quadratic cost  $\|x - y\|^2$  or in Theorem 2.3 of [19] for a continuous, positive cost.

To the best of our knowledge, the most general result was proved by Beiglböck, Goldstern, Maresch, and Schachermayer [9] who removed regularity assumptions on the cost:

**Theorem 16** ([9, Theorem 1.a]). *Let  $X, Y$  be Polish spaces equipped with Borel probability measures  $\mu, \nu$  and let  $c : X \times Y \rightarrow [0, \infty]$  be a Borel measurable cost function. Then every optimal transport plan with finite total cost is  $c$ -cyclically monotone.*

The reverse implication ( $c$ -cyclic monotonicity implies optimality) is not true in general, and we collect known results and counterexamples in the rest of this section.

### 3.2 When $c$ -cyclic monotonicity implies optimality?

We begin by presenting an example from [4], which shows that one cannot always expect  $c$ -cyclic monotonicity of the support of the plan to imply optimality. In this example the cost is regular, namely lower semi-continuous. However, it assumes not only real values but also the value  $+\infty$ .

**Example 17** ([4, Example 3.1]). *Let  $X = Y = [0, 1]$  and let  $\mu = \nu = \lambda$  be the Lebesgue measure. Letting  $\alpha \in [0, 1)$  be an irrational number define*

$$\Gamma = \{(x, x) : x \in X\}, \quad \Gamma_\alpha = \{(x, x \oplus \alpha) : x \in X\},$$

where  $\oplus$  is the addition modulo 1. Define the cost  $c = a \in [0, \infty)$  on  $\Gamma$ ,  $c = b \in [0, \infty)$  on  $\Gamma_\alpha$  and  $+\infty$  otherwise. Now both sets  $\Gamma$  and  $\Gamma_\alpha$  are  $c$ -cyclically monotone. Define maps  $T : X \rightarrow X \times Y$  to be  $T(x) = (x, x)$  and  $T_\alpha : X \rightarrow X \times Y$  to be  $T_\alpha(x) = (x, x \oplus \alpha)$ . These maps induce transport plans  $\pi = T_\# \lambda$  and  $\pi_\alpha = T_\alpha \# \lambda$  which are supported on  $\Gamma$  and  $\Gamma_\alpha$ , respectively. Both of these plans have finite total cost, equal to  $a$  and  $b$  respectively, which means that the optimality depends on the choice of values  $a, b$ , while, as already mentioned, both supports are  $c$ -cyclically monotone.

---

<sup>1</sup>At the time when the book was written the question of sufficiency of  $c$ -cyclic monotonicity was wide open. In fact, Open Problem 2.25 of the same book (later solved by Schechermeyer and Teichmann in [36]) asks whether sufficiency holds for the quadratic cost.

This example shows that one will have to assume something about the cost or the transported measures. It turns out that a joint property of measures and the cost is needed, which was shown by Beiglböck, Goldstern, Maresch, and Schachermayer in [9].

**Theorem 18** ([9, Theorem 1.b]). *Let  $X, Y$  be Polish spaces equipped with Borel probability measures  $\mu, \nu$  and  $c : X \times Y \rightarrow [0, \infty]$  a Borel measurable cost function. Assume that the set  $\{(x, y) \in X \times Y : c(x, y) = +\infty\}$  is the union of a closed set  $F$  and a  $\mu \otimes \nu$ -null set  $N$ . Let  $\gamma \in \Pi(\mu, \nu)$  be a finite and  $c$ -cyclically monotone plan. Then  $\gamma$  is optimal.*

This result is the state of the art, establishing when the equivalence of optimality and  $c$ -cyclic monotonicity holds. Nevertheless, in what follows, we are going to present some prior developments of this problem as we find it interesting and instructive. Moreover, it will give the reader the idea behind the proof of the above theorem.

### 3.2.1 Finitely-supported measures

For finite spaces  $X$  and  $Y$  or, equivalently, for finitely-supported measures  $\mu$  and  $\nu$  one can prove that

**Theorem 19.** *Assume that  $\mu$  and  $\nu$  are finitely-supported Borel probability measures on Polish spaces  $X$  and  $Y$ , respectively. A transport plan  $\gamma \in \Pi(\mu, \nu)$  is optimal with respect to the optimal transport problem corresponding to a cost function  $c : X \times Y \rightarrow \mathbb{R}$  if and only if it is  $c$ -cyclically monotone.*

The above theorem can be proved in several different ways. The first and, probably, more popular one, uses linear programming but, despite being part of the folklore, we cannot find a complete reference with proof. Several results for finitely supported measures are contained in [17]. In the quadratic case, this result is left as an exercise (Exercise 2.21) in [38]. Further, we remark that it is a particular case of Theorem 20 below. In Appendix B we give another proof that uses a peculiar structure of finitely-supported transport plans with the same marginals. More precisely, given two finitely-supported transport plans with the same marginals, it is possible to obtain one of them by permuting the coordinates of points in the support of the other.

Using Theorem 19 and the density of convex combinations of discrete measures in the space of probability measures one can prove that there exists an optimal transport plan with  $c$ -cyclically monotone support for general marginal measures  $\mu \in \mathcal{P}(X)$ ,  $\nu \in \mathcal{P}(Y)$  (see [25] and Remark 2.26 in [38]).

### 3.2.2 Purely atomic measures

In [28], Pratelli shows that if the measures  $\mu$  and  $\nu$  are purely atomic (that is, concentrated on at most countable sets) then optimality and  $c$ -cyclic monotonicity are indeed equivalent.

**Theorem 20** ([28, Theorem A]). *Let  $X$  and  $Y$  be Polish spaces,  $\mu \in \mathcal{P}(X)$  and  $\nu \in \mathcal{P}(Y)$  purely atomic, and  $c : X \times Y \rightarrow [0, \infty]$  a cost function. Then a finite transport plan  $\gamma \in \Pi(\mu, \nu)$  is optimal if and only if it is concentrated on a  $c$ -cyclically monotone set.*

In the same paper, using the above theorem and an approximation argument, Pratelli obtains a result for general measures under the assumption of continuity of the cost.

### 3.2.3 General measures

As mentioned above, Pratelli [28] shows, for any Borel measures, that if the cost is continuous the equivalence holds.

**Theorem 21** ([28, Theorem B]). *Let  $X$  and  $Y$  be Polish spaces,  $\mu \in \mathcal{P}(X)$  and  $\nu \in \mathcal{P}(Y)$ , and assume that the cost function  $c : X \times Y \rightarrow [0, \infty]$  is continuous. Then a finite transport plan  $\gamma \in \Pi(\mu, \nu)$  is optimal if and only if it is concentrated on a  $c$ -cyclically monotone set.*

Instead of insisting on the cost being continuous, it turns out that the statement holds for any real-valued Borel cost. A posteriori, this can be seen from Theorem 18. Indeed, for any finite-valued cost function the set  $\{(x, y) : c(x, y) = +\infty\}$  is empty and hence satisfies the conditions of the theorem, which in turn means that optimality and  $c$ -cyclic monotonicity of the plan are equivalent. Since Theorem 18 has a complex proof we choose to first show the statement for real-valued costs. We use the fundamental steps from [9] and, for clarity, we add more detail.

**Theorem 22.** *Let  $X, Y$  be Polish spaces equipped with Borel probability measures  $\mu, \nu$ , and let  $c : X \times Y \rightarrow [0, \infty)$  ( $c$  is real-valued) be a Borel measurable cost function. Then every  $c$ -cyclically monotone transport plan  $\gamma$  with finite total cost is optimal.*

*Proof.* By Theorem 7, since  $c$  is real-valued, there exists a function  $\varphi : X \rightarrow \mathbb{R} \cup \{-\infty\}$  such that  $\Gamma \subset \partial^c \varphi$ . In particular, this implies that

$$c(x, y) - \varphi(x) \leq c(x', y) - \varphi(x'), \quad \forall x' \in X, (x, y) \in \Gamma.$$

In fact, Theorem 7 gives a formula for  $\varphi$ , which we use to prove, in line with [9], that  $\varphi$  is universally measurable. This, together with an argument showing integrability, yields the optimality of the plan.

A crucial ingredient of the proof is the fact that the image of a Borel set via a Borel map is universally measurable (we recall the definition and the relevant facts about *universal measurability* in Appendix C). To show this, we analyse the function defined in (2) by splitting the minimisation process in two steps. First we minimise among chains of fixed length  $n$ , and define

$$\varphi_n(x) = \inf \{ c(x, y_n) - c(x_n, y_n) + c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1}) + \cdots + c(x_1, y_0) - c(x_0, y_0) : (x_i, y_i) \in \Gamma \text{ for } i = 1, \dots, n \}.$$

Then we minimise with respect to  $n$ , which gives the potential

$$\varphi(x) = \inf_n \varphi_n(x).$$

Since  $c$  is Borel measurable, we get that for each  $n$  and any  $n$ -tuple of points in  $\Gamma$  the function

$$c(x, y_n) - c(x_n, y_n) + c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1}) + \cdots + c(x_1, y_0) - c(x_0, y_0)$$

is a Borel measurable function on  $X \times \Gamma^n$ . Hence, for each  $n$  and  $\alpha \in \mathbb{R}$  the set

$$L_\alpha^n := \{ (x, y_n, x_n, \dots, y_0, x_0) : c(x, y_n) - c(x_n, y_n) + c(x_n, y_{n-1}) - c(x_{n-1}, y_{n-1}) + \cdots + c(x_1, y_0) - c(x_0, y_0) < \alpha \}$$

is a Borel subset of  $X \times \Gamma^n$ . Since

$$\{x : \varphi(x) < \alpha\} = \cup_n \{x : \varphi_n(x) < \alpha\},$$

$$\{x : \varphi_n(x) < \alpha\} = p_X(L_\alpha^n)$$

and  $p_X$  is a Borel function, we obtain the universal measurability of  $\varphi$ . We now show that there exists a Borel measurable function  $\tilde{\varphi}$  such that

$$\tilde{\varphi} \leq \varphi \quad \text{and} \quad \tilde{\varphi} = \varphi \quad \mu - a.e.$$

Indeed let  $\mathcal{I} = \{I_n\}$  be the family of half-closed intervals  $[a_n, b_n)$  with rational endpoints. Observe that  $\mathcal{I}$  generates the Borel  $\sigma$ -algebra on  $\mathbb{R}$ . Let  $\tilde{\mu}$  be the completion of  $\mu$  (for the definition see (17) in Appendix C). Since  $\varphi$  is universally measurable,  $\varphi^{-1}(I_n)$  is  $\tilde{\mu}$ -measurable and thus can be written as the union of sets  $B_n \cup N_n$  with  $B_n$  Borel measurable and  $\mu(N_n) = 0$ . Let  $N = \cup_n N_n$  we define

$$\tilde{\varphi}(x) = \begin{cases} -\infty & \text{if } x \in N, \\ \varphi(x) & \text{otherwise.} \end{cases}$$

We then consider  $\Gamma_1 := \Gamma \setminus p_X^{-1}(N)$  which is a  $c$ -cyclically monotone Borel set such that  $\gamma(\Gamma_1) = 1$  and

$$c(x, y) - \tilde{\varphi}(x) \leq c(x', y) - \tilde{\varphi}(x'), \quad \forall x' \in X, (x, y) \in \Gamma_1.$$

Moreover  $\tilde{\varphi}(x) \in \mathbb{R}$  on  $p_X(\Gamma_1)$ . We proceed, as usual, passing to a  $c$ -transform of the candidate Kantorovich potential  $\tilde{\varphi}$ . Define

$$\psi(y) = \inf_{x \in p_X(\Gamma_1)} \{c(x, y) - \tilde{\varphi}(x)\}.$$

The function  $\psi$  is universally measurable since

$$\{y : \psi(y) < \alpha\} = p_Y(\{(x, y) \in \Gamma_1 : c(x, y) - \tilde{\varphi}(x) < \alpha\}),$$

and  $\{(x, y) : c(x, y) - \tilde{\varphi}(x) < \alpha\}$  is a Borel set since  $c$  and  $\tilde{\varphi}$  are Borel measurable. Notice, also, that

$$\tilde{\varphi}(x) + \psi(y) = c(x, y), \quad \forall (x, y) \in \Gamma_1,$$

and

$$\tilde{\varphi}(x) + \psi(y) \leq c(x, y), \quad \forall (x, y) \in p_X(\Gamma_1) \times Y,$$

As in the previous case, assigning the value  $-\infty$  on a  $\nu$ -null set  $\tilde{N}$ , we consider a Borel measurable  $\tilde{\psi} : Y \rightarrow [-\infty, +\infty)$  such that

$$\tilde{\psi} \leq \psi \quad \text{and} \quad \tilde{\psi} = \psi \quad \nu - a.e.$$

We define  $\Gamma_2 := \Gamma_1 \setminus p_Y^{-1}(\tilde{N})$ . The set  $\Gamma_2$  is Borel measurable and  $c$ -cyclically monotone, and  $\gamma(\Gamma_2) = 1$ . Moreover

$$\tilde{\varphi}(x) + \tilde{\psi}(y) = c(x, y), \quad \forall (x, y) \in \Gamma_2$$

and

$$\tilde{\varphi}(x) + \tilde{\psi}(y) \leq c(x, y), \quad \forall (x, y) \in p_X(\Gamma_1) \times Y.$$

Since the potential needs to be defined everywhere on  $X \times Y$  we let  $\tilde{\varphi}$  be  $-\infty$  outside  $p_X(\Gamma_2)$ .

As a result we obtain a pair of Borel measurable functions  $(\tilde{\varphi}, \tilde{\psi})$ , which is  $c$ -splitting for the  $c$ -cyclically monotone set  $\Gamma_2$  on which  $\gamma$  is concentrated. The possible lack of integrability of  $\tilde{\varphi}$  and  $\tilde{\psi}$  does not allow us to deduce the optimality of  $\gamma$  in the usual way, that is integrating directly expressions  $\tilde{\varphi} + \tilde{\psi} = c$  on the support of  $\gamma$  and  $\tilde{\varphi} + \tilde{\psi} \leq c$  everywhere. We use truncations instead. We fix an arbitrary transport plan  $\gamma' \in \Pi(\mu, \nu)$ ; we may assume that  $C[\gamma'] < +\infty$ . We consider truncated potentials  $\tilde{\varphi}_n = (n \wedge (\tilde{\varphi} \vee -n))$  and  $\tilde{\psi}_n = (n \wedge (\tilde{\psi} \vee -n))$  and denote  $\xi = \tilde{\varphi} + \tilde{\psi}$ ,  $\xi_n = \tilde{\varphi}_n + \tilde{\psi}_n$ . Now  $\xi^n \uparrow \xi^+$  on the set  $\{\xi \geq 0\}$ ,  $\xi^n \downarrow \xi^-$  on the set  $\{\xi \leq 0\}$  and by monotone convergence, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \int_{\xi \geq 0} \xi_n d\gamma \uparrow \int_{\xi \geq 0} \xi d\gamma < +\infty, \quad \int_{\xi \geq 0} \xi_n d\gamma' \uparrow \int_{\xi \geq 0} \xi d\gamma' < +\infty, \\ \int_{\xi \leq 0} \xi_n d\gamma \downarrow \int_{\xi \leq 0} \xi d\gamma \quad \text{and} \quad \int_{\xi \leq 0} \xi_n d\gamma' \downarrow \int_{\xi \leq 0} \xi d\gamma'. \end{aligned}$$

Therefore the integrals of  $\xi$  with respect to  $\gamma$  and  $\gamma'$  are well-defined, and

$$\lim_{n \rightarrow \infty} \int \xi_n d\gamma = \int \xi d\gamma \quad \text{and} \quad \lim_{n \rightarrow \infty} \int \xi_n d\gamma' = \int \xi d\gamma'.$$

Note that since  $\gamma$  and  $\gamma'$  have the same marginals, we have

$$\begin{aligned} \int \xi d\gamma &= \lim_{n \rightarrow \infty} \int \xi_n d\gamma = \lim_{n \rightarrow \infty} \left( \int \tilde{\varphi}_n d\mu + \int \tilde{\psi}_n d\nu \right) \\ &= \lim_{n \rightarrow \infty} \int \xi_n d\gamma' = \int \xi d\gamma'. \end{aligned}$$

The optimality of  $\gamma$  now readily follows since

$$C[\gamma] = \int c d\gamma = \int \xi d\gamma = \int \xi d\gamma' \geq \int c d\gamma' = C[\gamma']. \quad \square$$

Note that the starting point of the above proof of Theorem 22 is the formula (2) which in Theorem 7 gives the existence of a real-valued potential. However, as we have seen in Section 2, when the cost attains infinite values this approach fails. Hence, some assumptions must be added to give sense to the formula, and a useful notion to consider seems to be  $c$ -connectivity (recall Definition 8).

Another approach is to arrive at the existence of a potential without the use of a formula of the type (2). In fact it was shown in [5] that if the cost attains infinite values then  $c$ -cyclic monotonicity is no longer equivalent to the existence of a potential (see example at the end of Section 2). The authors established that a necessary and sufficient condition for the existence of a potential is that of  $c$ -path-boundedness (see Definition 9). This approach also appears in the multi marginal setting (see Theorem 28).

The issue arising from the latter approach is the possible lack of measurability. This is why we are reverting to the condition of  $c$ -connectivity.

**Theorem 23.** *Let  $X, Y$  be Polish spaces equipped with Borel probability measures  $\mu, \nu$ , and let  $c : X \times Y \rightarrow [0, \infty]$  a Borel measurable cost function. Let  $\gamma$  be a  $c$ -cyclically monotone transport plan with finite total cost and assume that  $\gamma$  is supported on a  $c$ -connecting set  $\Gamma$ . Then  $\gamma$  is optimal.*

*Proof.* Let us recall the Rockafellar-Rochet-Rüschemdorf formula: fix  $(x_0, y_0) \in \Gamma$  and define the function  $\varphi_n : X \times \Gamma^n : (-\infty, +\infty]$  by setting

$$\varphi_n(x, x_1, y_1, \dots, x_n, y_n) = c(x, y_n) - c(x_n, y_n) + \sum_{i=0}^{n-1} (c(x_{i+1}, y_i) - c(x_i, y_i)). \quad (5)$$

Define for all  $x \in X$

$$\varphi(x) := \inf\{\varphi_n(x, x_1, y_1, \dots, x_n, y_n) \mid n \geq 1, (x_i, y_i)_{i=1}^n \in \Gamma^n\}. \quad (6)$$

The key is to prove that also in this case the formula (6) makes sense. From that point on one can proceed like in the proof of Theorem 22. Hence it suffices to show the following:

**Claim.**  $\varphi(x) \in \mathbb{R}$  for all  $x \in p_X(\Gamma)$  and  $(7)$

$$\varphi(x) \leq \varphi(x') + c(x, y) - c(x', y) \quad \text{for all } x \in X \text{ and } (x', y) \in \Gamma. \quad (8)$$

*Proof.* We fix  $x \in p_X(\Gamma)$ . So there exists  $y \in Y$  such that  $(x, y) \in \Gamma$ . Since  $\Gamma$  is  $c$ -connecting and  $(x, y), (x_0, y_0) \in \Gamma$ , we can find  $(x_1, y_1), \dots, (x_{n-1}, y_{n-1}) \in \Gamma$  such that, setting  $(x_n, y_n) = (x, y)$ , we have

$$\varphi_n(x; x_1, y_1, \dots, x_n, y_n) < +\infty.$$

This means that  $\varphi(x) < +\infty$ , and it remains to show that  $\varphi(x) > -\infty$ .

We denote  $a_1 = x$ . Since  $x \in p_X(\Gamma)$ , there exists  $b_1 \in Y$  such that  $(a_1, b_1) \in \Gamma$ . The  $c$ -connectivity of  $\Gamma$  gives us the existence of a chain  $(a_1, b_1), \dots, (a_m, b_m) \in \Gamma$  such that  $c(a_2, b_1), c(a_3, b_2), \dots, c(a_m, b_{m-1}) < +\infty$ . For any ordered collection  $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$  such that

$$\varphi_n(x, x_1, y_1, \dots, x_n, y_n) < +\infty$$

we extend the chain  $(x_1, y_1), \dots, (x_n, y_n)$  by setting for all  $i \in \{1, \dots, m\}$   $x_{n+i} = a_i$  and  $y_{n+i} = b_i$ . Due to the  $c$ -cyclic monotonicity of  $\gamma$  and the finiteness of all the terms involved, we have (denoting  $x_{m+n+1} = x_0$ ) that

$$0 \leq \sum_{i=0}^{n+m} (c(x_{i+1}, y_i) - c(x_i, y_i)) = c(x_0, y_{n+m}) - c(x_{n+m}, y_{n+m}) + \sum_{i=0}^{n+m-1} (c(x_{i+1}, y_i) - c(x_i, y_i)).$$

That is

$$\begin{aligned} \alpha &:= c(a_m, b_m) - c(x_0, b_m) + \sum_{i=1}^{m-1} (c(a_i, b_i) - c(a_{i+1}, b_i)) \\ &\leq c(x, y_n) - c(x_n, y_n) + \sum_{i=0}^{n-1} (c(x_{i+1}, y_i) - c(x_i, y_i)). \end{aligned}$$

This, in turn, means that  $\alpha \leq \varphi_n(x, x_1, y_1, \dots, x_n, y_n)$ , and passing to the infimum we get that  $-\infty < \alpha \leq \varphi(x)$ .

To prove the inequality (8), observe that its right hand side can be written as

$$\begin{aligned}
& \inf_{n \geq 1, (x_i, y_i)_{i=1}^n \in \Gamma^n} \left\{ c(x', y_n) - c(x_n, y_n) + \sum_{i=0}^{n-1} (c(x_{i+1}, y_i) - c(x_i, y_i)) \right\} + c(x, y) - c(x', y) \\
&= \inf_{n \geq 1, (x_i, y_i)_{i=1}^n \in \Gamma^n} \left\{ c(x, y) - c(x', y) + \sum_{i=0}^{n-1} (c(x_{i+1}, y_i) - c(x_i, y_i)) + (c(x', y_n) - c(x_n, y_n)) \right\} \\
&= \inf_{m \geq 1, (x_i, y_i)_{i=1}^m \in \Gamma^m, (x_m, y_m) = (x', y)} \left\{ c(x, y_m) - c(x_m, y_m) + \sum_{i=0}^{m-1} [c(x_{i+1}, y_i) - c(x_i, y_i)] \right\} \\
&= \inf_{m \geq 1, (x_i, y_i)_{i=1}^m \in \Gamma^m, (x_m, y_m) = (x', y)} \varphi_n(x, x_1, y_1, \dots, x_m, y_m).
\end{aligned}$$

Finally, we conclude that on both sides of (8) we take the infimum of the same function but on the right-hand side the set over which we minimize is smaller, which concludes the proof.  $\square$

*Remark 24.* Example 17 shows that the assumption of connectivity in the above theorem is not necessary for optimality.

The previous theorem can be used as a building block to study under which conditions all  $c$ -cyclically monotone transport plans are optimal. This boils down to the study of the structure of the set where the cost is infinite, as it has been done in [9], and we recall the theorem.

**Theorem 18** ([9, Theorem 1.b]). Assume that the set where  $c \equiv +\infty$  is the union of a closed set  $F$  and a  $\mu \otimes \nu$ -null set  $N$ . Let  $\gamma \in \Pi(\mu, \nu)$  be a finite and  $c$ -cyclically monotone plan. Then  $\gamma$  is optimal.

The proof is based on finding a set  $\Gamma \subset X \times Y$  of full  $\gamma$ -measure and a partition  $\{\Gamma \cap (C_i \times D_i)\}_{i \in I}$  of  $\Gamma$  which is countable and such that each element of the partition is  $c$ -connecting. This can be done due to the fact that the set  $X \times Y \setminus F$  is open. The key property of the partition is that not only the starting plan  $\gamma$  but also *any other finite transport plan*  $\tilde{\gamma}$  satisfies  $\tilde{\gamma}(\bigcup_{i \in I} C_i \times D_i) = 1$ .

### 3.3 Existence of plans supported on a $c$ -subgradient

In [6] the authors took a slightly different path to finding a Brenier-type map. Omitting the equivalence between optimality and a plan being supported on a  $c$ -subgradient of a function, they showed that such an optimal plan exists. More precisely, we have the following theorem.

**Theorem 25** ([6, Theorem 1.1]). Let  $X = Y$  be a Polish space, let  $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  be a continuous and symmetric cost function, essentially bounded from below with respect to probability measures  $\mu \in \mathcal{P}(X)$  and  $\nu \in \mathcal{P}(Y)$ . Assume  $\mu$  and  $\nu$  are strongly  $c$ -compatible, namely satisfy that for any measurable  $A \subset X$  we have

$$\mu(A) + \nu(\{y \in Y : \forall x \in A, c(x, y) = \infty\}) \leq 1, \quad (9)$$

and for any measurable  $A \subset X$  with  $\mu(A) \neq 0, 1$  we have

$$\mu(A) + \nu(\{y \in Y : \forall x \in A, c(x, y) = \infty\}) < 1. \quad (10)$$

If there exists some finite plan transporting  $\mu$  to  $\nu$ , then there exists a  $c$ -class function  $\varphi$  and an optimal transport plan  $\pi \in \Pi(\mu, \nu)$  concentrated on  $\partial^c \varphi$ .



*Remark 26.* Note that the condition (9) is a necessary condition for the existence of a plan such that each atom is transported at finite cost. In the special discrete setting this condition corresponds to the Hall's marriage theorem that guarantees the existence of a matching. The second condition (10) turns out to be a sufficient condition. It guarantees that the optimal plan, which is known to be  $c$ -cyclically monotone is *connecting* and therefore  $c$ -path-bounded.

*Remark 27.* The assumption of lower semi-continuity of the cost is standard and as mentioned previously it guarantees the existence of an optimal plan. The authors assume upper semi-continuity to guarantee measurability of the set  $\{y \in Y : \forall x \in A, c(x, y) = \infty\}$ , which in this case is simply closed and hence measurable. However, this assumption is not needed. Indeed, assuming only the lower semi-continuity of the cost we see that the set

$$\{y \in Y : \forall x \in A, c(x, y) = \infty\} = \bigcap_{n \in \mathbb{N}} \{y \in Y : \forall x \in A, c(x, y) > n\}$$

is a countable intersection of open sets and as such is measurable. Moreover, the measurability of the potential follows from the fact that the support of the plan is  $c$ -connecting.

Moreover, in the case when the target measure  $\nu$  has finite support, it was shown ([6, Theorem 1.3]) that it is sufficient to assume measurability of the cost (and a mild condition regarding a joint property of the measures and the cost called  $c$ -regularity, see [6, Definition 5.1]).

## 4 More general transport problems

### 4.1 The multi-marginal, integral case

Let  $N \geq 2$ ,  $X_1, \dots, X_N$  be Polish spaces, denote  $X = \prod_{i=1}^N X_i$  and consider a cost function  $c : X \rightarrow \mathbb{R} \cup \{+\infty\}$ . The natural generalization of the classical optimal transport to the multidimensional case is to consider

$$\min_{\gamma \in \Pi(\mu_1, \dots, \mu_N)} C(\gamma) := \min_{\gamma \in \Pi(\mu_1, \dots, \mu_N)} \int_X c d\gamma,$$

in the set

$$\Pi(\mu_1, \dots, \mu_N) := \{\gamma \in \mathcal{P}(X) \mid p_{X_i}(\gamma) = \mu_i \text{ for all } i = 1, \dots, N\},$$

that is, in the set of *couplings* or *transport plans* between the  $N$  marginals  $\mu_1, \dots, \mu_N$ .

Like in the two-marginals case, to establish the existence of minimizers it is sufficient to assume, for example, that the cost function is lower semi-continuous and essentially bounded from below.

The notion of multidimensional  $c$ -cyclic monotonicity was introduced in Definition 12. It brings the concept of  $c$ -cyclic monotonicity in classical optimal transport to the multimarginal case. The multimarginal definition is more complex but carries the same idea: if we permute the ‘destinations’ of points in the support of an optimal plan, the cost should not improve. In finite spaces for real-valued cost functions,  $c$ -cyclic monotonicity and optimality are equivalent also in the multimarginal case, see Theorem 36 in Appendix B.

The rest of this subsection is dedicated to an overview of what we know about the sufficiency of  $c$ -cyclic monotonicity for optimality in general Polish spaces for arbitrary Borel probability measures as marginals.

Recall that the concept of a  $c$ -splitting set (Definition 13) is the multimarginal counterpart of the  $c$ -subgradient and a useful tool for proving the optimality of a plan. In [23] Kim and Pass proved that the multimarginal  $c$ -cyclic monotonicity is a necessary condition for the  $c$ -splitting property

of a set. In [21], Griessler showed that in Polish spaces, under some boundedness conditions on a real-valued cost function,  $c$ -cyclical monotonicity of the support of a plan implies its optimality. More precisely, we have the following theorem.

**Theorem 28** ([21, Theorem 1.2]). *Let  $X_1, \dots, X_N$  be Polish spaces and  $\mu_1, \dots, \mu_N$  be Borel probability measures on the spaces  $X_1, \dots, X_N$ . Let  $c : X_1 \times \dots \times X_N \rightarrow [0, \infty)$  be a continuous cost function such that for some  $f_i \in L^1(X_i, \mu_i)$ ,  $i = 1, \dots, N$ , one has  $c(x_1, \dots, x_N) \leq \sum_{i=1}^N f_i(x_i)$ .<sup>2</sup> Let  $\gamma \in \Pi(\mu_1, \dots, \mu_N)$  be a transport plan such that  $\text{supp}(\gamma)$  is  $c$ -cyclically monotone. Then  $\gamma$  is optimal.*

Griessler starts by proving that a  $c$ -cyclically monotone transport plan is finitely optimal, i.e. that all its finitely-supported submeasures are optimal with respect to their marginals. He continues by commenting that by a linear programming argument one can show that finite and  $c$ -cyclically monotone sets are  $c$ -splitting. The transition from finite  $c$ -splitting sets to arbitrary  $c$ -splitting sets is done by first fixing a  $c$ -cyclically monotone set  $\Gamma$  and an element  $x^0 \in \Gamma$ , and then proving that the family

$$\mathcal{F} := \bigcup_{G \subset \Gamma \text{ finite}} \mathcal{G}_G,$$

where

$$\mathcal{G}_G = \{(\varphi_1, \dots, \varphi_N) : \varphi \text{ is a } (G \cup \{x^0\}, c) \text{-splitting tuple such that} \\ \varphi_i(x_i) \leq c(x_1^0, \dots, x_{i-1}^0, x_i, x_{i+1}^0, \dots, x_N^0) \text{ for all } i = 1, \dots, N\},$$

has the finite-intersection property. Hence by the compactness of the space  $\overline{\mathbb{R}}^{X_1} \times \overline{\mathbb{R}}^{X_2} \times \dots \times \overline{\mathbb{R}}^{X_N}$  the set

$$\bigcap_{G \subset \Gamma, G \text{ finite}} \mathcal{G}_G$$

is nonempty, proving the existence of a  $(\Gamma, c)$ -splitting tuple.

To prove the existence of a Borel measurable tuple Griessler uses the continuity of the cost function: if  $(\varphi_1, \dots, \varphi_N)$  is a  $(\Gamma, c)$ -splitting tuple, then by first setting

$$\tilde{\varphi}_1(x_1) = \inf \left\{ c(x_1, y_2, \dots, y_N) - \sum_{j=2}^N \varphi_j(y_j) : (y_2, \dots, y_N) \in X_2 \times \dots \times X_N \right\}$$

and then, assuming that  $\tilde{\varphi}_1, \dots, \tilde{\varphi}_{i-1}$  are already defined,

$$\tilde{\varphi}_i(x_i) = \inf \left\{ c(y_1, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_N) - \sum_{j=1}^{i-1} \tilde{\varphi}_j(y_j) - \sum_{j=i+1}^N \varphi_j(y_j) : \right. \\ \left. y_j \in X_j, j \in \{1, \dots, i-1, i+1, \dots, N\} \right\}$$

one obtains a  $(\Gamma, c)$ -splitting tuple that is upper semicontinuous and, in particular, Borel measurable.

---

<sup>2</sup>This condition does not need to hold everywhere. It is enough that there exist  $N$  sets  $N_i \subset X_i$  such that  $\mu_i(N_i) = 0$  and such that the condition holds outside  $\cup_i p_{X_i}^{-1}(N_i)$ .

To pass from measurable potentials to potentials whose integrals are well-defined (and thus to optimality) Griessler shows that if  $(\varphi_1, \dots, \varphi_N)$  is a  $(\Gamma, c)$ -splitting tuple and if  $x^0 \in \Gamma$ , then there exists a  $(\Gamma, c)$ -splitting tuple  $(\tilde{\varphi}_1, \dots, \tilde{\varphi}_N)$  such that

$$\tilde{\varphi}_i(x_i) \leq c(x_1^0, \dots, x_{i-1}^0, x_i, x_{i+1}^0, \dots, x_N^0) \quad \text{for all } i \in \{1, \dots, N\} \text{ and } x_i \in X_i. \quad (11)$$

This is achieved by first noting that, since  $x^0 \in \Gamma$  and since  $(\varphi_1, \dots, \varphi_N)$  is a  $(\Gamma, c)$ -splitting tuple, we have

$$\sum_{i=1}^N \varphi_i(x_i^0) = c(x_1^0, \dots, x_N^0),$$

and, in particular, (since  $c$  assumes only real values)  $\varphi_i(x_i^0) \in \mathbb{R}$  for all  $i \in \{1, \dots, N\}$ . Then one defines

$$\begin{aligned} \tilde{\varphi}_1(x_1) &= \varphi_1(x_1) + \sum_{j=2}^N \varphi_j(x_j^0) \quad \text{and} \\ \tilde{\varphi}_i(x_i) &= \varphi_i(x_i) - \varphi_i(x_i^0) \quad \text{for } i = 2, \dots, N. \end{aligned}$$

Starting from a Borel measurable  $(\Gamma, c)$ -splitting tuple  $(\varphi_1, \dots, \varphi_N)$  one gets the existence of a Borel measurable  $(\Gamma, c)$ -splitting tuple  $(\tilde{\varphi}_1, \dots, \tilde{\varphi}_N)$  satisfying, in addition, (11). The fact that the integrals of the functions  $\tilde{\varphi}_i$  are well-defined follows by the assumption that  $c$  is bounded from above by a sum  $\sum_{i=1}^N f_i$  of  $\mu_i$ -integrable functions: for a fixed  $x^0 \in \Gamma$  that satisfies  $\sum_{i=1}^N f_i(x_i^0) < +\infty$  we have

$$\begin{aligned} \int_{X_i} \tilde{\varphi}_i(x_i) d\mu_i(x_i) &\leq \int_{X_i} c(x_1^0, \dots, x_{i-1}^0, x_i, x_{i+1}^0, \dots, x_N^0) d\mu_i(x_i) \\ &\leq \sum_{j \neq i} f_j(x_j^0) + \int_{X_i} f_i(x_i) d\mu_i(x_i) < +\infty. \end{aligned}$$

If  $\Gamma$  is a  $c$ -cyclically monotone set on which a transport plan  $\gamma \in \Pi(\mu_1, \dots, \mu_N)$  is concentrated, the optimality of the plan  $\gamma$  now follows easily: if  $\gamma'$  is another transport plan, one has (denoting  $X = \prod_{i=1}^N X_i$ ) that

$$\int_X c d\gamma = \int_X \sum_{i=1}^N \tilde{\varphi}_i d\gamma = \int_X \sum_{i=1}^N \varphi_i d\gamma' \leq \int_X c d\gamma'.$$

It is an open problem if the  $c$ -cyclic monotonicity of the support of a plan implies its optimality if some of the assumptions of Griessler are lifted. For example if we consider real-valued, Borel (not necessarily continuous) cost function bounded from above by a sum of integrable functions. For this type of a cost one could prove the existence of a  $(c, \Gamma)$ -splitting tuple as above, but without the continuity assumption on the cost the existence of a Borel measurable splitting tuple is not clear. Note that in the 2-marginal case the measurability was based on the Rockafellar-type formula (2), but in the multimarginal case this kind of formula is not available.

If the cost function assumes the value  $+\infty$  there are results in the opposite direction. In [27], M. Petrache presented an example, for the three-marginal case, of a  $c$ -cyclically monotone transport plan that is not optimal. In the example the spaces  $X_i$  were all equal to  $\mathbb{N}$ , the natural numbers, and the cost function assumes infinite value. Clearly, this discrete example can naturally be embedded

in  $\mathbb{R}^3$  we view the discrete marginal measures as elements of  $\mathcal{P}(\mathbb{N})$ . Note that in this case the cost is not continuous. While if we regard  $\mathbb{N}$  as a Polish space equipped with the discrete metric, the cost is continuous but not bounded from above by a sum of integrable functions of the marginal measure spaces.

**Example 29** ([27, Proposition 2.1]). *We fix  $N = 3$  and consider the three-marginal optimal transportation problem with the spaces  $X_1 = X_2 = X_3 = \mathbb{N}$  and all marginals equal to the measure*

$$\mu = \sum_{k=1}^{\infty} 2^{-k} \delta_k \in \mathcal{P}(\mathbb{N}).$$

To define the cost function we fix an auxiliary function  $f : \mathbb{N} \rightarrow (0, 1]$  such that

$$\sum_{k=1}^{\infty} 4^{-k} (f(2k-1) - f(2k)) > \frac{1}{6}. \quad (12)$$

This condition is satisfied, for example, by all decreasing functions  $f : \mathbb{N} \rightarrow (0, 1]$  such that  $f(1) > f(2) + \frac{2}{3}$ . We define the cost function  $c : (\mathbb{N})^3 \rightarrow [0, \infty]$  as follows:

$$\begin{cases} c(a, b, c) \text{ is symmetric with respect to permuting the coordinates } (a, b, c), \\ c(1, 1, 1) = 1, \\ c(a, a, a+1) = f(a) \text{ for all } a \in \mathbb{N} \text{ and} \\ c = +\infty \text{ for all triples } (a, b, c) \text{ not described by the previous two conditions.} \end{cases}$$

We consider the following transport plans  $\gamma, \bar{\gamma} \in \Pi(\mu, \mu, \mu)$ :

$$\begin{aligned} \gamma &= \sum_{k=1}^{\infty} 4^{-k} (\delta_{(2k-1, 2k-1, 2k)} + \delta_{(2k-1, 2k, 2k-1)} + \delta_{(2k, 2k-1, 2k-1)}) \quad \text{and} \\ \bar{\gamma} &= \frac{1}{2} \delta_{(1, 1, 1)} + \frac{1}{2} \sum_{k=1}^{\infty} 4^{-k} (\delta_{(2k, 2k, 2k+1)} + \delta_{(2k, 2k+1, 2k)} + \delta_{(2k+1, 2k, 2k)}). \end{aligned}$$

We will show that the plan  $\gamma$  is  $c$ -cyclically monotone but not optimal.

The plan  $\gamma$  cannot be optimal since the plan  $\bar{\gamma}$  has a lower cost:

$$\begin{aligned} C[\gamma] - C[\bar{\gamma}] &= 3 \sum_{k=1}^{\infty} 4^{-k} f(2k-1) - \frac{1}{2} - 3 \sum_{k=1}^{\infty} 4^{-k} f(2k) \\ &= 3 \sum_{k=1}^{\infty} 4^{-k} (f(2k-1) - f(2k)) - \frac{1}{2} > 0, \end{aligned}$$

where we have used the property (12) of the function  $f$ .

It remains to show that  $\gamma$  is  $c$ -cyclically monotone. We show that it is finitely-optimal; by Propositions 40 and 42 of Appendix D these two notions are equivalent. We fix a finitely-supported subplan  $\alpha$  of  $\gamma$ , and another transport plan  $\alpha'$  that has the same marginals as  $\alpha$ . It is enough to show that  $C[\alpha] \leq C[\alpha']$ . If  $C[\alpha'] = +\infty$ , the claim holds trivially. We may thus assume that  $C[\alpha'] < +\infty$ .

We show that in fact in this case  $\alpha = \alpha'$ ; the idea is that the set where for given marginals the cost can be finite is very small.

In the following we denote by  $(\delta_{(a,b,c)})^S$  the symmetrisation

$$(\delta_{(a,b,c)})^S = \frac{1}{6}(\delta_{(a,b,c)} + \delta_{(a,c,b)} + \delta_{(b,a,c)} + \delta_{(b,c,a)} + \delta_{(c,a,b)} + \delta_{(c,b,a)}).$$

Due to the symmetries of the problem and the structure of the plan  $\gamma$  we may assume that for some  $M > 0$  and coefficients  $a_k \geq 0$

$$\alpha = \sum_{k=0}^M a_k (\delta_{(2k-1, 2k-1, 2k)})^S.$$

The marginals of this plan are all equal to

$$\mu_\alpha = \sum_{k=1}^M a_k \left( \frac{2}{3} \delta_{2k-1} + \frac{1}{3} \delta_{2k} \right).$$

We may also assume that the measure  $\alpha'$  is symmetric. Since  $C[\alpha'] < +\infty$ , there exists nonnegative constants  $a', a'_1, \dots, a'_M, b'_1, \dots, b'_M$  such that

$$a' + \sum_{k=1}^M (a'_k + b'_k) = 1 \quad \text{and} \quad \alpha' = a' \delta_{(1,1,1)} + \sum_{k=1}^M (a'_k (\delta_{(2k-1, 2k-1, 2k)})^S + b'_k (\delta_{(2k, 2k, 2k+1)})^S).$$

Since the marginals of  $\alpha'$  are all equal to  $\mu_\alpha$ , we have

$$\mu_\alpha(\{1\}) = \frac{2}{3} a_1 = \alpha'(\{1\} \times \mathbb{N} \times \mathbb{N}) = a' + \frac{2}{3} a'_1$$

and thus

$$2a_1 = 3a' + 2a'_1. \tag{13}$$

Similarly we have, for all  $k \geq 2$ :

$$\mu_\alpha(\{2k-1\}) = \frac{2}{3} a_k = \alpha'(\{2k-1\} \times \mathbb{N} \times \mathbb{N}) = \frac{2}{3} a'_k + \frac{1}{3} b'_{k-1}$$

which implies that

$$2a_k = 2a'_k + b'_{k-1} \quad \text{for all } k \geq 2. \tag{14}$$

And finally, for  $k \geq 1$ , we have

$$\mu_\alpha(\{2k\}) = \frac{1}{3} a_k = \alpha'(\{2k\} \times \mathbb{N} \times \mathbb{N}) = \frac{1}{3} a'_k + \frac{2}{3} b'_k$$

giving

$$a_k = a'_k + 2b'_k \quad \text{for all } k \geq 1. \tag{15}$$

Equations (14) and (15) imply that

$$a'_k + 2b'_k = a'_k + \frac{b'_{k-1}}{2} \quad \text{for all } k \geq 2,$$

and thus

$$b'_{k+1} = \frac{b'_k}{4} \quad \text{for all } k \geq 1.$$

Since  $b'_k = 0$  for all  $k > M$  we have, therefore,  $b'_k = 0$  for all  $k = 1$ . Hence, by (15), we have  $a_k = a'_k$  for all  $k \geq 1$ , and this inserted in (13) gives  $a' = 0$ , completing the proof of  $\alpha = \alpha'$ .

The counterexample above shows that neither of the two-marginal Theorems 20 and 21 of Pratelli can be extended to the multi-marginal case.

## 4.2 The $L^\infty$ -problem.

Let  $X_1$  and  $X_2$  be Polish spaces and  $c : X_1 \times X_2 \rightarrow [0, \infty]$  be a lower semi-continuous cost function. The  $L^\infty$  optimal mass transportation problem can be stated as follows:

$$\min_{\Pi(\mu_1, \mu_2)} C_\infty[\gamma] := \min_{\Pi(\mu_1, \mu_2)} \gamma - \text{essup } c(x_1, x_2) \quad (P_\infty)$$

This problem is ‘a limit case’ of the classical optimal transport since, if  $c \geq 0$ , is the variational limit, as a  $p \rightarrow \infty$ , of the optimal transport problems with the cost functional  $C_p[\gamma] := (\int_X c^p d\gamma)^{\frac{1}{p}}$ . In fact, if we take a sequence  $(\gamma_p)_{p=1}^\infty$ , with  $\gamma_p$  minimizer of the corresponding functional  $C_p$  (and thus  $c$ -cyclically monotone), we get, as the weak\*-cluster points  $\gamma = \lim_{p \rightarrow \infty} \gamma_p$ , some solutions of the problem  $(P_\infty)$  corresponding to the cost function  $c$ . If  $c$  is continuous and if for each  $p$  the value  $C_p[\gamma]$  is finite, these solutions are *infinitely cyclically monotone*, in the sense of the following definition:

**Definition 30.** We say that a set  $\Gamma \subset X_1 \times X_2$  is *infinitely  $c$ -cyclically monotone*, if for every  $k$ -tuple of points  $(x_1^i, x_2^i)_{i=1}^k$  of  $\Gamma$  and for every permutation  $\sigma$  of the set  $\{1, \dots, k\}$  we have

$$\max\{c(x_1^i, x_2^i) \mid i \in \{1, \dots, k\}\} \leq \max\{c(x_1^i, x_2^{\sigma(i)}) \mid i \in \{1, \dots, k\}\}.$$

We also say that a coupling  $\gamma \in \Pi(\mu_1, \mu_2)$  is *infinitely cyclically monotone* if it is concentrated on an infinitely cyclically monotone set.

The infinitely cyclically monotone transport plans are a class of ‘well-behaved’ solutions to the problem  $(P_\infty)$ . These plans are well-behaved in the sense that the restrictions of infinitely cyclically monotone plans are still infinitely cyclically monotone and thus optimal (if infinitely cyclically monotone plans are optimal). In general, the problem  $(P_\infty)$  is much more cumbersome than the standard optimal transportation problem. The cost functional is neither linear nor convex, there is no classical duality theory, and there is a high degree of non-uniqueness of the solutions. Loosely speaking, in the problem  $(P_\infty)$  we are only interested in optimizing the worst case, so locally its solutions can be far from optimal. In this last sense, the infinitely cyclically monotone minimizers have better properties, and, at least for continuous cost  $c$  that admits finite  $C_p$  solutions, the infinitely cyclically monotone couplings exist as stated above.

The question we consider is: are optimal plans infinitely cyclically monotone, and are infinitely cyclically monotone plans optimal? The classical two approaches to the necessity of the condition (see Section 3.1) are based on the linearity or on the duality for convex problems, but as was mentioned this formulation of the problem lacks both. In fact, not all plans that are optimal for  $(P_\infty)$  are infinitely cyclically monotone. The next example illustrates the typical behavior. By  $\mathcal{L}^1|_{[0,1]}$  we denote one-dimensional Lebesgue measure on the interval  $[0, 1]$ .

**Example 31.** Let  $X_1 = X_2 = \mathbb{R}$ ,  $c(x, y) = |x - y|$ ,  $\mu = \frac{1}{2}(\mathcal{L}^1|_{[0,1]} + \delta_1)$ , and  $\nu = \frac{1}{2}(\mathcal{L}^1|_{[0,1]} + \delta_{10})$ , then any plan that sends 1 to 10 is optimal, even if the interval  $[0, 1]$  were mapped to itself in a very ‘non-optimal’ way (for example by a map  $x \mapsto 1 - x$ ). The locally ‘best-behaved’ optimal transport plan would be  $\gamma = \frac{1}{2}((id, id)_\# \mathcal{L}^1|_{[0,1]} + \delta_{(1,10)})$ , and this plan is, actually, infinitely cyclically monotone.

Concerning the sufficiency, Champion, De Pascale, and Juutinen in [16] laid the foundations for the two-marginal  $L^\infty$  transport on  $\mathbb{R}^d$  with the Euclidean distance as a cost function. Later Jylhä in [22] proved the corresponding result in Polish spaces for continuous cost functions  $c : X_1 \times X_2 \rightarrow [0, \infty]$ . If the continuity assumption is dropped, the optimality of an infinitely cyclically monotone plan fails even in the case of a lower semi-continuous cost with finite values. The following counterexample is a slightly modified version of the one given by Pratelli and Ambrosio in [4] of a non-optimal  $c$ -cyclically monotone transportation plan. In the original counterexample it was necessary that  $c$  assumes the value  $+\infty$  (since for a lower semi-continuous cost that assumes only finite values, in the integral case,  $c$ -cyclic monotonicity is equivalent to optimality).

**Example 32.** Consider the two-marginal  $L^\infty$ -optimal transportation problem with marginals  $\mu = \nu = \mathcal{L}|_{[0,1]}$  and the cost function

$$c(x, y) = \begin{cases} 1 & \text{if } x = y \\ 2 & \text{otherwise.} \end{cases}$$

Let  $\alpha$  be an irrational number. Set  $T_1 = Id_{[0,1]}$  and  $T_2 : [0, 1] \rightarrow [0, 1]$ ,  $T_2(x) = x + \alpha \pmod{1}$ . The map  $T_1$  is an optimal transportation map for the problem  $(P_\infty)$  with  $C_\infty[T_1] = 1$ . Since the map  $T_2$  has the cost 2, it cannot be optimal. However, it is infinitely cyclically monotone, as we show below. Assume, contrary to the claim, that  $T_2$  is not infinitely cyclically monotone. Then there exists a minimal  $K \in \mathbb{N}$  and a  $K$ -tuple of couples  $\{x_i, y_i\}_{i=1}^K$ , all belonging to the support of the plan given by  $T_2$ , such that

$$\max_{1 \leq i \leq K} c(x_i, y_i) > \max_{1 \leq i \leq K} c(x_{i+1}, y_i),$$

with the convention  $x_{K+1} = x_1$ . By the definition of the map  $T_2$  we have  $y_i = x_i + \alpha \pmod{1}$  for all  $i$ . By the definition of the cost  $c$ , the only form in which this inequality can hold is:  $2 > 1$ . The right-hand side now tells us that  $y_i = x_i + \alpha \pmod{1}$  for all  $i$ , that is,  $x_{i+1} = x_i + \alpha \pmod{1}$  for all  $i$ . Summing up, this gives (keeping in mind that  $x_{K+1} = x_1$ ) that  $x_1 = x_1 + K\alpha \pmod{1}$ , contradicting the irrationality of  $\alpha$ .

#### 4.2.1 The multi-marginal case

The  $L^\infty$  problem can be stated analogously also in the multi-marginal case: For  $N \geq 2$ , let  $(X_1, \mu_1), \dots, (X_N, \mu_N)$  be Polish probability spaces equipped with Borel probability measures  $\mu_i \in \mathcal{P}(X_i)$ . Denoting  $X = \prod_{i=1}^N X_i$ , we consider the cost function  $c : X \rightarrow [0, \infty]$  and the following minimization problem

$$\min_{\gamma \in \Pi(\mu_1, \dots, \mu_N)} C_\infty[\gamma] := \min_{\gamma \in \Pi(\mu_1, \dots, \mu_N)} \gamma - \text{essup}_{(x_1, \dots, x_N) \in X} c, \quad (P_\infty^N)$$

where

$$\Pi(\mu_1, \dots, \mu_N) := \{\gamma \in \mathcal{P}(X) : p_{X_i}(\gamma) = \mu_i \text{ for all } i = 1, \dots, N\}.$$

In this context the  $c$ -cyclic monotonicity takes the following form (compare with Definition 13):

**Definition 33.** We say that a set  $\Gamma \subset \prod_{i=1}^N X_i$  is *infinitely  $c$ -cyclic monotone* if for every  $k$ -tuple of points  $(x^{1,i}, \dots, x^{N,i})_{i=1}^k$  of  $\Gamma$  and every  $(N-1)$ -tuple of permutations  $(\sigma_2, \dots, \sigma_N)$  of the set  $\{1, \dots, k\}$  we have

$$\max\{c(x_1^i, x_2^i, \dots, x_N^i) : i \in \{1, \dots, k\}\} \leq \max\{c(x_1^i, x_2^{\sigma_2(i)}, \dots, x_N^{\sigma_N(i)}) : i \in \{1, \dots, k\}\}.$$

We also say that a coupling  $\gamma \in \Pi(\mu_1, \dots, \mu_N)$  is infinitely cyclically monotone if it is concentrated on an infinitely cyclically monotone set.

For  $N = 2$  this definition coincides with the infinitely cyclically monotone condition presented above for the two-marginal case, and exactly like in the two-marginal  $L^\infty$ -optimal transportation problem, also here an optimal plan is not necessarily infinitely cyclically monotone. De Pascale and Kausamo proved in [18] the sufficiency of the infinite cyclic monotonicity for continuous costs and marginal measures with compact supports:

**Theorem 34** ([18, Theorem 1.3]). *Let  $\mu_i \in \mathcal{P}(X_i)$  with compact support for  $i = 1, \dots, N$ , let  $c : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be continuous. If  $\gamma \in \Pi(\mu_1, \dots, \mu_N)$  is an infinitely cyclically monotone plan for  $c$ , then  $\gamma$  is optimal.*

Again, since in this case there is no duality available, one cannot rely on  $c$ -splitting  $N$ -tuples. Therefore, the optimality proofs have to be based on different techniques. For example, De Pascale and Kausamo first prove that all finitely-supported submeasures of an infinitely cyclically monotone transport plan are optimal with respect to their marginals. Using a suitable finite discretisations of the original infinitely cyclically monotone transport problem, they then apply  $\Gamma$ -convergence techniques to pass from optimality of the discretised transport plan to the optimality of the full transport plan. The key property of the infinitely cyclically monotone condition of a support of a transport plan is that it is inherited by subsets.

## A Additional example

The following example is a version of [4, Example 3.1], also presented in this note in Example 17, but the main focus is shifted to show that  $c$ -path-bounded sets need not be  $c$ -connecting.

**Example 35.** *Let  $X = Y = [0, 1]$  with the identification of the end points 0 and 1, and let  $\mu = \nu$  be the Lebesgue measure on  $[0, 1]$ . Fix an irrational number  $\alpha \in [0, 1)$  and define the cost function  $c : X \times Y \rightarrow [0, \infty]$  as*

$$c(x, y) = \begin{cases} 1 & \text{if } x = y \\ 2 & \text{if } x = y + \alpha \pmod{1} \\ +\infty & \text{otherwise.} \end{cases}$$

Consider the set

$$\Gamma = \{(x, x) : x \in X\}.$$

Then the set  $\Gamma$  is  $c$ -path-bounded but not  $c$ -connecting.

To see that  $\Gamma$  is  $c$ -path-bounded, fix points  $(x, y), (z, w) \in \Gamma$ . For any  $m \in \mathbb{N}$  and any chain  $\{(x_i, y_i)\}_{i=2}^{m-1}$  of points in  $\Gamma$ , letting  $(x_1, y_1) = (x, y)$  and  $(x_m, y_m) = (z, w)$ , we have

$$\sum_{i=1}^{m-1} (c(x_i, y_i) - c(x_{i+1}, y_i)) = \sum_{i=1}^{m-1} (1 - c(x_{i+1}, y_i)) \leq 0,$$

since for all  $i \in \{1, \dots, m-1\}$  we have  $c(x_{i+1}, y_i) \in \{1, 2, +\infty\}$ . Hence, for any start and finish points, we may choose the constant  $M((x, y), (z, w))$  in Definition 9 to be 0.

The set  $\Gamma$  is not connecting: if it were connecting then, for example, for points  $(0, 0), (\frac{1}{2}, \frac{1}{2}) \in \Gamma$  there must exist a chain  $(x_1, y_1), \dots, (x_n, y_n) \in \Gamma$  with  $(x_1, y_1) = (0, 0)$  and  $(x_n, y_n) = (\frac{1}{2}, \frac{1}{2})$  such that  $c(x_1, y_0), \dots, c(x_n, y_{n-1}) < +\infty$ . This would imply the existence of  $0 \leq m \leq n$  such that  $\frac{1}{2} \equiv 0 + m\alpha \pmod{1}$ , which is impossible given the irrationality of  $\alpha$ .



## B Finite spaces

In this appendix we establish the equivalence of  $c$ -cyclic monotonicity and optimality in finite spaces for real-valued cost functions. Equivalently one could consider finitely-supported measures on arbitrary spaces. In the following,  $X_1, \dots, X_N$  are finite spaces and  $\mu_1, \dots, \mu_N$  probability measures on  $X_1, \dots, X_N$ , respectively. We denote  $X = \prod_{i=1}^N X_i$ .

**Theorem 36.** *Let  $\gamma \in \Pi(\mu_1, \dots, \mu_N)$  be an optimal transport plan with respect to a cost function  $c : X \rightarrow \mathbb{R}$  where  $X = \prod_{i=1}^N X_i$  and  $X_i$  are finite spaces. Then  $\gamma$  is  $c$ -cyclically monotone if and only if it is optimal.*

*Proof.* Assume that  $\gamma$  is optimal. Since we work in finite spaces we denote  $X_i = \{x_i^j\}_{j \in \{1, \dots, n_i\}}$  for  $i \in \{1, \dots, N\}$ , where  $n_i$  is the cardinality of  $X_i$ . We also denote  $n = n_1 \cdot \dots \cdot n_N$  and

$$I = \{1, \dots, n_1\} \times \{1, \dots, n_2\} \times \dots \times \{1, \dots, n_N\}.$$

The index vectors  $(i_1, \dots, i_N) \in I$  will be abbreviated by  $\mathbf{i}$  and vectors  $(x_1^{i_1}, \dots, x_N^{i_N}) \in X$  by  $\mathbf{x}^{\mathbf{i}}$ . Using this notation, the plan  $\gamma$  can be expressed as

$$\gamma = \sum_{\mathbf{i} \in I} a^{\mathbf{i}} \delta_{\mathbf{x}^{\mathbf{i}}} \quad \text{for some coefficients } a^{\mathbf{i}} \geq 0.$$

Let us assume, contrary to the claim, that  $\gamma$  is not  $c$ -cyclically monotone. This means that there exist  $k \in \mathbb{N}$ , index set  $\tilde{I} := \{(i_1^j, i_2^j, \dots, i_N^j)\}_{j=1}^k \subset I$ , points  $\{(x_1^{i_1^j}, \dots, x_N^{i_N^j})\}_{j=1}^k \subset \text{supp} \gamma$ , and permutations  $\sigma_2, \dots, \sigma_N$  of the set  $\{1, \dots, k\}$  such that

$$\sum_{1 \leq j \leq k} c(\mathbf{x}^{\sigma(j)}) < \sum_{1 \leq j \leq k} c(\mathbf{x}^{i^j}), \quad (16)$$

where we have abbreviated

$$\mathbf{x}^{i^j} := (x_1^{i_1^j}, \dots, x_N^{i_N^j}) \quad \text{and} \quad \mathbf{x}^{\sigma(j)} := (x_1^{i_1^j}, x_2^{i_2^{\sigma_2(j)}}, \dots, x_N^{i_N^{\sigma_N(j)}}).$$

We denote

$$a := \min\{a^{\mathbf{i}} \mid \mathbf{i} \in \tilde{I}\} > 0$$

and define

$$\tilde{\gamma} := \sum_{\mathbf{i} \in I \setminus \tilde{I}} a^{\mathbf{i}} \delta_{\mathbf{x}^{\mathbf{i}}} + \sum_{\mathbf{i} \in \tilde{I}} (a^{\mathbf{i}} - a) \delta_{\mathbf{x}^{\mathbf{i}}} + a \sum_{1 \leq j \leq k} \delta_{\mathbf{x}^{\sigma(j)}}.$$

The measure  $\tilde{\gamma}$  has the same marginals as  $\gamma$  because  $\sigma_2, \dots, \sigma_N$  are permutations. Using (16) we get

$$\begin{aligned} C[\tilde{\gamma}] &= \sum_{\mathbf{i} \in I \setminus \tilde{I}} a^{\mathbf{i}} c(\mathbf{x}^{\mathbf{i}}) + \sum_{\mathbf{i} \in \tilde{I}} (a^{\mathbf{i}} - a) c(\mathbf{x}^{\mathbf{i}}) + a \sum_{1 \leq j \leq k} c(\mathbf{x}^{\sigma(j)}) \\ &< \sum_{\mathbf{i} \in I \setminus \tilde{I}} a^{\mathbf{i}} c(\mathbf{x}^{\mathbf{i}}) + \sum_{\mathbf{i} \in \tilde{I}} (a^{\mathbf{i}} - a) c(\mathbf{x}^{\mathbf{i}}) + a \sum_{1 \leq j \leq k} c(\mathbf{x}^{i^j}) = C[\gamma], \end{aligned}$$

contradicting the optimality of  $\gamma$ .

The proof of the opposite direction ( $c$ -cyclical monotonicity implies optimality) is essentially the same as the proof of Proposition 40 and we do not repeat it here.  $\square$

## C Universally measurable sets and functions

Let  $X$  be a Polish space. Consider the measure space  $(X, \Sigma, \mu)$  with  $\mu$  a  $\sigma$ -finite Borel measure (so  $\mathcal{B}(X) \subset \Sigma$ ). Let  $\mathcal{N}_\mu = \{N \subset X : \mu(N) = 0\}$ , then  $\tilde{\Sigma}_\mu := \{A \cup N : A \in \Sigma, N \in \mathcal{N}_\mu\}$  is still a  $\sigma$ -algebra. The measure

$$\begin{aligned} \tilde{\mu} : \tilde{\Sigma}_\mu &\rightarrow [0, +\infty] \\ A \cup N &\mapsto \mu(A) \end{aligned} \tag{17}$$

is called the *completion* of  $\mu$ .

A set  $B \subset X$  is called *universally measurable* if it is measurable for the completion of every finite Borel measure  $\mu$  on  $X$ .

**Proposition 37.** *The set of universally measurable sets is a Borel  $\sigma$ -algebra.*

*Proof.* This follows from the fact that the defining properties of a  $\sigma$ -algebra are stable for the intersection.  $\square$

If  $Y$  is another Polish space and  $T : X \rightarrow Y$  is a Borel map, then the image of a Borel set through  $T$  is called an *analytic* set. The following theorem is relevant here but we do not include the proof as it is beyond the scope of this paper.

**Theorem 38.** *Every analytic set is universally measurable.*

## D Finite optimality

In this section we prove that the  $c$ -cyclic monotonicity of a transport plan is equivalent to it being finitely optimal. We present the proofs directly in the multi-marginal case and for both the integral and the  $L^\infty$  optimal transport.

We start with the definition of the finite optimality.

**Definition 39.** Let  $\gamma$  be a positive and finite Borel measure on  $X$ . We say that  $\gamma$  is *finitely optimal* if all its finitely-supported submeasures are optimal with respect to their marginals. By submeasure we mean any probability measure  $\alpha$  satisfying  $\text{supp}(\alpha) \subset \text{supp}(\gamma)$ .

**Proposition 40.** *If  $\gamma \in \Pi(\mu_1, \dots, \mu_N)$  is  $c$ -cyclically monotone or infinitely cyclically monotone, then it is finitely optimal for the corresponding optimal transportation problem (integral or  $L^\infty$ , respectively).*

**Lemma 41.** *Let  $\alpha = \sum_{i=1}^l m_i \delta_{(x^{1,i}, \dots, x^{N,i})}$  and  $\bar{\alpha} = \sum_{i=1}^{\tilde{l}} \bar{m}_i \delta_{(\bar{x}^{1,i}, \dots, \bar{x}^{N,i})}$  be two discrete measures with positive, integer coefficients and the same marginals. Let us denote by  $\tilde{l} = m^1 + \dots + m^l$  the number of rows of the following table*

$$\left. \begin{array}{ccc} x^{1,1} & \dots & x^{N,1} \\ \vdots & & \vdots \\ x^{1,l} & \dots & x^{N,l} \end{array} \right\} m^1\text{-times}$$

... ..

$$\left. \begin{array}{ccc} x^{1,l} & \dots & x^{N,l} \\ \vdots & & \vdots \\ x^{1,l} & \dots & x^{N,l} \end{array} \right\} m^l\text{-times}$$

where the first  $m^1$  rows are equal among themselves, the following  $m^2$  rows are equal among themselves and so on. Let  $\bar{A}$  be the analogous table associated to  $\bar{\alpha}$ . Then  $\bar{A}$  has  $\tilde{l}$  rows and there exist  $(N-1)$  permutations of the set  $\{1, \dots, \tilde{l}\}$  such that  $\bar{A}$  is equal to

$$\begin{array}{ccc} x^{1,1} & \dots & x^{N,\sigma^N(1)} \\ \vdots & & \vdots \\ x^{1,1} & \dots & x^{N,\sigma^N(m_1)} \\ x^{1,2} & \dots & x^{N,\sigma^N(m_1+1)} \\ \vdots & & \vdots \\ x^{1,l} & \dots & x^{N,\sigma^N(m_1+\dots+m_{l-1}+1)} \\ \vdots & & \vdots \\ x^{1,l} & \dots & x^{N,\sigma^N(\tilde{l})}. \end{array}$$

*Proof.* For each  $k \in \{1, \dots, N\}$ , the  $k$ -th marginal of  $\alpha$  is given by the sum of the Dirac masses centered on the points of the  $k$ -th column of the table  $A$  with multiplicity. Analogously, the  $k$ -th marginal of  $\bar{\alpha}$  is given by the sum of the Dirac masses centered at the points of the  $i$ -th column of the table  $\bar{A}$  with multiplicity. Since the marginals of  $\alpha$  and  $\bar{\alpha}$  are the same, each point  $x^{k,i}$  appearing in the  $k$ -th marginal must appear in both matrices the same number of times, proving the existence of the bijections  $\sigma^2, \dots, \sigma^N$  as required. This also implies that  $\bar{A}$  has  $\tilde{l}$  rows.  $\square$

*Proof of Proposition 40.* We fix a finitely-supported submeasure  $\alpha = \sum_{i=1}^l a_i \delta_{X^i}$  of  $\gamma$ . We need to show that  $\alpha$  is an optimal coupling of its marginals. To do this, we fix another coupling,  $\bar{\alpha} = \sum_{i=1}^{\tilde{l}} \bar{a}_i \delta_{\bar{X}^i}$ , with the same marginals as  $\alpha$ . We have to show that

$$\tilde{C}[\alpha] \leq \tilde{C}[\bar{\alpha}], \quad (18)$$

where  $\tilde{C}$  is any of the two total costs under consideration.

Let us first assume that the discrete measures  $\alpha$  and  $\bar{\alpha}$  have rational coefficients. We consider the measures  $M\alpha$  and  $M\bar{\alpha}$ , where  $M$  is the product of the denominators of the coefficients of  $\alpha$  and  $\bar{\alpha}$ . They are discrete measures having positive, integer coefficients and the same marginals, so we can apply Lemma 41 to find permutations  $\sigma^2, \dots, \sigma^N$  such that  $M\alpha$  and  $M\bar{\alpha}$  have representations  $A$  and  $\bar{A}$ , respectively. If  $\tilde{C} = C$  we have, using the  $c$ -cyclical monotonicity of  $\alpha$

$$MC[\alpha] = \sum_{i=1}^{\tilde{l}} c(x^{1,i}, \dots, x^{N,i}) \leq \sum_{i=1}^{\tilde{l}} c(x^{1,i}, x^{2,\sigma^2(i)}, \dots, x^{N,\sigma^N(i)}) = MC[\bar{\alpha}],$$

proving the optimality of  $\alpha$ . If  $\tilde{C} = C_\infty$ , the conclusion is also immediate:

$$C_\infty[\alpha] = \max_{1 \leq i \leq k} c(x^{1,i}, \dots, x^{N,i}) \leq \max_{1 \leq i \leq k} c(x^{1,i}, x^{2,\sigma^2(i)}, \dots, x^{N,\sigma^N(i)}) = C_\infty[\bar{\alpha}].$$

Now, assume that  $\alpha$  and  $\bar{\alpha}$  have real (not necessarily rational) coefficients, and that

$$\alpha := \sum_{i=1}^l a_i \delta_{X^i}, \quad \bar{\alpha} := \sum_{i=1}^{\bar{l}} \bar{a}_i \delta_{\bar{X}^i}.$$

We show that for all  $\varepsilon > 0$  there exist two discrete measures

$$\beta := \sum_{i=1}^l q_i \delta_{X^i} \quad \text{and} \quad \bar{\beta} := \sum_{i=1}^{\bar{l}} \bar{q}_i \delta_{\bar{X}^i},$$

with the same marginals,  $q_i, \bar{q}_i \in \mathbb{Q}$  and

$$|a_i - q_i| < \varepsilon, \quad |\bar{a}_i - \bar{q}_i| < \varepsilon.$$

Being concentrated on  $X^1, \dots, X^l$  and  $\bar{X}^1, \dots, \bar{X}^{\bar{l}}$  is equivalent to the fact that the vector  $\underline{\mathbf{a}} := (a_1, \dots, a_l, \bar{a}_1, \dots, \bar{a}_{\bar{l}})$  is a solution of

$$\mathcal{A}\underline{\mathbf{a}} = 0,$$

where  $\mathcal{A}$  is a matrix with coefficients 1, 0,  $-1$ . Indeed, if we write, for example, the equality between the first two marginals we obtain

$$\sum_{i=1}^l a_i \delta_{x^{1,i}} = \sum_{i=1}^{\bar{l}} \bar{a}_i \delta_{\bar{x}^{1,i}}.$$

Therefore, some of the points  $\bar{x}^{1,i}$  must coincide with, for example,  $x^{1,1}$  which implies that for two sets of indices we have

$$\sum_{i \in I} a_i = \sum_{j \in J} \bar{a}_j.$$

Since the matrix  $\mathcal{A}$  has integer coefficients

$$\overline{Ker_{\mathbb{Q}} \mathcal{A}} = Ker_{\mathbb{R}} \mathcal{A},$$

and this allows to choose  $\beta$  and  $\bar{\beta}$ . Since  $C[\alpha] \approx C[\beta]$ ,  $C[\bar{\alpha}] \approx C[\bar{\beta}]$ ,  $C_{\infty}[\alpha] = C_{\infty}[\beta]$  and  $C_{\infty}[\bar{\alpha}] = C_{\infty}[\bar{\beta}]$ , this finishes the proof.  $\square$

The following proposition gives the opposite implication:

**Proposition 42.** *Let  $\gamma \in \Pi_N(\mu_1, \dots, \mu_N)$  be a finitely optimal transport plan. Then  $\gamma$  is  $c$ -cyclically monotone or infinitely cyclically monotone, depending on the underlying transport problem.*

*Proof.* First, we present the proof in the case of the integral optimal transport problem. Assume on the contrary that there exist points  $\{(x_1^i, \dots, x_N^i)\}_{i=1}^k \subset \text{supp}(\gamma)$  and  $N-1$  permutations  $\sigma_2, \dots, \sigma_N$  of the set  $\{1, \dots, k\}$  such that

$$\sum_{i=1}^k c(x_1^i, x_2^{\sigma_2(i)}, \dots, x_N^{\sigma_N(i)}) < \sum_{i=1}^k c(x_1^i, x_2^i, \dots, x_N^i). \quad (19)$$

Consider the finite submeasure  $\alpha$  of  $\gamma$ , defined by

$$\alpha = \frac{1}{k} \sum_{i=1}^k \delta_{(x_1^i, x_2^i, \dots, x_N^i)},$$

and the measure

$$\alpha' = \frac{1}{k} \sum_{i=1}^k \delta_{(x_1^i, x_2^{\sigma_2(i)}, \dots, x_N^{\sigma_N(i)})}.$$

Since  $\sigma_2, \dots, \sigma_N$  are bijections, the measures  $\alpha$  and  $\alpha'$  have the same marginals. However, by inequality (19) we have  $C[\alpha'] < C[\alpha]$ , contradicting the finite optimality of  $\gamma$ .

The case of the  $L^\infty$  optimal transport is analogous. The only difference is that inequality (19) takes the form

$$\max_{1 \leq i \leq k} c(x_1^i, x_2^{\sigma_2(i)}, \dots, x_N^{\sigma_N(i)}) < \max_{1 \leq i \leq k} c(x_1^i, x_2^i, \dots, x_N^i).$$

□

## References

- [1] T. Abdellaoui and H. Henich. Sur la distance de deux lois dans le cas vectoriel. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, 319(4):397–400, 1994.
- [2] L. Ambrosio and N. Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer, 2013.
- [3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.
- [4] L. Ambrosio and A. Pratelli. Existence and stability results in the  $L^1$  theory of optimal transportation. In *Optimal transportation and applications*, pages 123–160. Springer, 2003.
- [5] S. Artstein-Avidan, S. Sadovskiy, and K. Wyczesany. A Rockafellar-type theorem for non-traditional costs. *Advances in Mathematics*, 395:108157, 2022.
- [6] S. Artstein-Avidan, S. Sadovskiy, and K. Wyczesany. Optimal measure transportation with respect to non-traditional costs. *Calculus of Variations and Partial Differential Equations*, 62(1):35, 2023.
- [7] S. Bartz, H. Bauschke, H. Phan, and X. Wang. Multi-marginal maximal monotonicity and convex analysis. *Mathematical Programming*, 185(1):385–408, 2021.
- [8] S. Bartz, H. Bauschke, and X. Wang. A class of multi-marginal c-cyclically monotone sets with explicit c-splitting potentials. *Journal of Mathematical Analysis and Applications*, 461(1):333–348, 2018.
- [9] M. Beiglböck, M. Goldstern, G. Maresch, and W. Schachermayer. Optimal and better transport plans. *Journal of Functional Analysis*, 256(6):1907–1927, 2009.

- [10] J. Bertrand. Prescription of Gauss curvature using optimal mass transport. *Geometriae Dedicata*, 183(1):81–99, 2016.
- [11] J. Bertrand, A. Pratelli, and M. Puel. Kantorovich potentials and continuity of total cost for relativistic cost functions. *Journal de Mathématiques Pures et Appliquées*, 110:93–122, 2018.
- [12] J. Bertrand and M. Puel. The optimal mass transport problem for relativistic costs. *Calculus of Variations and Partial Differential Equations*, 46(1-2):353–374, 2013.
- [13] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.
- [14] Y. Brenier. Extended Monge-Kantorovich theory. In *Optimal transportation and applications*, pages 91–121. Springer, 2003.
- [15] L. Caffarelli. The regularity of mappings with a convex potential. *Journal of the American Mathematical Society*, 5(1):99–104, 1992.
- [16] T. Champion, L. De Pascale, and P. Juutinen. The  $\infty$ -wasserstein distance: Local solutions and existence of optimal transport maps. *SIAM Journal on Mathematical Analysis*, 40(1):1–20, 2008.
- [17] M. Cuturi and G. Peyré. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
- [18] L. De Pascale and A. Kausamo. Sufficiency of  $c$ -cyclical monotonicity in a class of multi-marginal optimal transport problems. *arXiv preprint arXiv:2212.08375*, 2022.
- [19] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Mathematica*, 177(2):113–161, 1996.
- [20] W. Gangbo and V. Oliker. Existence of optimal maps in the reflector-type problems. *ESAIM: Control, Optimisation and Calculus of Variations*, 13(1):93–106, 2007.
- [21] C. Griessler.  $c$ -cyclical monotonicity as a sufficient criterion for optimality in the multi-marginal Monge–Kantorovich problem. *Proceedings of the American Mathematical Society*, 146(11):4735–4740, 2018.
- [22] H. Jylhä. The  $l^\infty$ -optimal transport: infinite cyclical monotonicity and the existence of optimal transport maps. *Calculus of Variations and Partial Differential Equations*, 52:303–326, 2015.
- [23] Y.-H. Kim and B. Pass. A general condition for monge solutions in the multi-marginal optimal transport problem. *SIAM Journal on Mathematical Analysis*, 46(2):1538–1550, 2014.
- [24] M. Knott and C. Smith. On Hoeffding-Fréchet bounds and cyclic monotone relations. *Journal of multivariate analysis*, 40(2):328–334, 1992.
- [25] R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. 1995.
- [26] G. Minty. On the monotonicity of the gradient of a convex function. *Pacific Journal of Mathematics*, 14(1):243–247, 1964.

- [27] M. Petrache. Cyclically monotone non-optimal  $n$ -marginal transport plans and smirnov-type decompositions for  $n$ -flows. *ESAIM: Control, Optimisation and Calculus of Variations*, 26:120, 2020.
- [28] A. Pratelli. On the sufficiency of  $c$ -cyclical monotonicity for optimality of transport plans. *Mathematische Zeitschrift*, 258(3):677–690, 2008.
- [29] Yu. V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, 1(2):157–214, 1956.
- [30] J.-C. Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of mathematical Economics*, 16(2):191–200, 1987.
- [31] R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17(3):497–510, 1966.
- [32] R. T. Rockafellar. *Convex analysis*, volume 11. Princeton University Press, 1997.
- [33] R. T. Rockafellar. *Convex analysis*. Princeton University Press, 2015.
- [34] L. Rüschendorf. On  $c$ -optimal random variables. *Statistics & probability letters*, 27(3):267–270, 1996.
- [35] F. Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.
- [36] W. Schachermayer and J. Teichmann. Characterization of optimal transport plans for the monge-kantorovich problem. *Proceedings of the American Mathematical Society*, 137(2):519–529, 2009.
- [37] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- [38] C. Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.
- [39] K. Wyczesany. *Topics in high-dimensional geometry and optimal transport*. 2021.