Canad. J. Math. Vol. **67** (2), 2015 pp. 350–368
http://dx.doi.org/10.4153/CJM-2014-011-x

Multimarginal Optimal Transport Maps for One-dimensional Repulsive Costs

Maria Colombo, Luigi De Pascale, and Simone Di Marino

Abstract. We study a multimarginal optimal transportation problem in one dimension. For a symmetric, repulsive cost function, we show that, given a minimizing transport plan, its symmetrization is induced by a cyclical map, and that the symmetric optimal plan is unique. The class of costs that we consider includes, in particular, the Coulomb cost, whose optimal transport problem is strictly related to the strong interaction limit of Density Functional Theory. In this last setting, our result justifies some qualitative properties of the potentials observed in numerical experiments.

1 Introduction

In some recent papers the authors considered a mathematical model for the strong interaction limit of the density functional theory (DFT). In particular, in [3], Buttazzo, De Pascale, and Gori-Giorgi showed that the model for the minimal interaction of N electrons can be formulated in terms of a multimarginal Monge transport problem. At the same time, in [6], Cotar, Friesecke, and Klüppelberg showed that an analogous optimal transportation problem describes the semiclassical limit of DFT in the case of two electrons and provides estimates from below in the general case.

Let $c\colon (\mathbb{R}^d)^N \to \mathbb{R}$ be the Coulomb cost function

$$c(x_1,\dots,x_N) = \sum_{1\le i<j\le N} \frac{1}{|x_i - x_j|}, \qquad (x_1,\dots,x_N) \in (\mathbb{R}^d)^N, \tag{1.1}$$

let $\rho \in \mathcal{P}(\mathbb{R}^d)$ be a given probability measure on $\mathbb{R}^d$, and let $\mathcal{T}(\rho)$ be the set of transport maps

$$\mathcal{T}(\rho) = \{T\colon \mathbb{R}^d \to \mathbb{R}^d \text{ Borel} : T_\sharp\rho = \rho\},$$

where $T_\sharp\rho$ represents the pushforward measure of the measure ρ through the Borel map T. We consider the Monge multimarginal problem

$$(M) = \inf\Big\{ \int_{\mathbb{R}^d} c(x, T_2(x),\dots,T_N(x))\,d\mu(x) : T_2,\dots,T_N \in \mathcal{T}(\rho)\Big\}. \tag{1.2}$$

Following the standard theory of optimal transport, we introduce the set of transport plans

$$\Pi(\rho) = \{\gamma \in \mathcal{P}(\mathbb{R}^{dN}) : \pi^i_\sharp\gamma = \rho,\ i = 1,\dots,N\},$$

Received by the editors July 23, 2013.
Published electronically May 22, 2014.
AMS subject classification: **49Q20**, 49K30.
Keywords: Monge–Kantorovich problem,optimal transport problem, cyclical monotonicity.

where $\pi^i\colon (\mathbb{R}^d)^N \to \mathbb{R}^d$ are the projections on the i-th component for $i = 1, \dots, N$, and the Kantorovich multimarginal problem

$$(1.3) \qquad (K) = \min\Big\{ \int_{(\mathbb{R}^d)^N} c(x_1, \dots, x_N) d\gamma(x_1, \dots, x_N) : \gamma \in \Pi(\rho) \Big\},$$

where, in contrast with (M), we allow the splitting of mass. With every $N-1$-tuple of transport maps $T_2, \dots, T_N \in \mathcal{T}(\rho)$ we associate the transport plan $\gamma = (\mathrm{Id}, T_2, \dots, T_N)_\sharp \rho \in \Pi(\rho)$. We remark that the existence of an optimal transport plan, namely a minimizer of (K), follows from the lower semicontinuity of the cost, from the linearity of the cost of a plan γ with respect to γ, and from the fact that the admissible plans form a tight subset of the set of measures on $(\mathbb{R}^d)^N$. Moreover, since $c\colon (\mathbb{R}^d)^N \to [0, \infty]$ is continuous, one can deduce from the analogous 2-marginal result obtained by Pratelli [18, Theorem B] that $(K) = (M)$. Indeed, given an optimal plan $\gamma \in \Pi(\rho)$ it is enough to apply the 2-marginal result to $X = \mathbb{R}^d$ and $Y = (\mathbb{R}^d)^{N-1}$ with ρ and $(\pi^2, \dots, \pi^N)_\sharp \gamma$ as marginals to obtain the nontrivial inequality $(M) \leq (K)$.

Since the cost function is symmetric, namely

$$c(x_1, \dots, x_N) = c(x_{\sigma(1)}, \dots, x_{\sigma(N)}), \quad \text{for } x_i \in \mathbb{R}^d, \sigma \in \mathcal{S}_N,$$

it is natural to introduce a "symmetric" version (K_{sym}) of the Kantorovich problem (K) and a "cyclical" version (M_{cycl}) of the Monge problem (M). From the physical point of view, this new formulation includes in the model the fact that the N electrons are indistinguishable. To this end, we denote by $\mathcal{S}_N$ the set of permutations of $\{1, \dots, N\}$, and we let

$$\tau_\sigma(x_1, \dots, x_N) = (x_{\sigma(1)}, \dots, x_{\sigma(N)}) \qquad \text{for } x \in (\mathbb{R}^d)^N, \sigma \in \mathcal{S}_N.$$

We introduce the set of symmetric transport plans

$$\Pi_{\mathrm{sym}}(\rho) = \{\gamma \in \Pi(\rho) : (\tau_\sigma)_\sharp \gamma = \gamma \ \forall \sigma \in \mathcal{S}_N\}$$

and the problems

$$(K_{\mathrm{sym}}) = \min\Big\{ \int_{(\mathbb{R}^d)^N} c(x_1, \dots, x_N) d\gamma(x_1, \dots, x_N) : \gamma \in \Pi_{\mathrm{sym}}(\rho) \Big\}.$$

$$(M_{\mathrm{cycl}}) = \inf\Big\{ \int_{\mathbb{R}^d} c\big(x, T(x), \dots, T^{(N-1)}(x)\big)\, d\mu(x) : T \in \mathcal{T}(\rho),\ T^{(N)} = \mathrm{Id} \Big\},$$

where $T^{(i)}$ stands for the i-th composition of T with itself.

Thanks to the symmetry of the cost and to the linearity of the cost of transport plans, we have easily that $(K) = (K_{\mathrm{sym}})$. Indeed, it is enough to associate with every $\gamma \in \Pi(\rho)$ the symmetric transport plan

$$\gamma^* = \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \sigma_\sharp \gamma,$$

which has the same cost as γ. The analogous remark for the Monge problem is not obvious; however, as proved in [5], in the sharp assumption that ρ is non-atomic we

https://doi.org/10.4153/CJM-2014-011-x Published online by Cambridge University Press

have that $(K) = (M_{\text{cycl}})$. Summarizing the last observations, if ρ is a non-atomic probability measure, we have that

$$(K) = (M) = (K_{\text{sym}}) = (M_{\text{cycl}}). \tag{1.4}$$

We remark that, seeing the nontrivial equality between (M) and (M_{cycl}), it is natural to expect as a general principle that an optimal map for (M) exhibits a cyclical structure when it exists. We also remark that the class of transport plans $\Pi_{\text{sym}}(\rho)$ is the natural one in which to hope for uniqueness results, as described later.

The problem of existence of optimal transport maps and the description of their properties is the main aim of optimal transport theory. As regards existence, in the case of two marginals, it holds under the so called "twist condition" (see [4, 9, 21]). In particular for the Coulomb cost (1.1) with $N = 2$ in any dimension, we have existence of optimal maps. In the multimarginal case, existence has been proved under some local assumptions of the cost in [16], but these assumptions are far from being satisfied by our cost. Indeed we will see that the optimal transport plans are absolutely not unique due to the symmetry of the cost, whereas in [16] uniqueness is a natural consequence of the assumptions on the cost and of the method. In [12] N. Ghoussoub and A. Moameni provide existence of optimal transport maps for some particular costs generated by vector fields (see also [11]). In this case, optimal maps are not unique and show a cyclical structure like the one presented in (M_{cycl}). However the structure of the cost function is different from the Coulomb cost, and correspondingly the strategy of the proofs is different.

In this paper we show the existence of optimal transport maps in the problem (1.2) in dimension $d = 1$, providing an explicit construction of the optimal map. The result was conjectured in [19] by M. Seidl from the physical point of view and recalled in [3]. Kohn–Sham DFT is used to reduce the computational cost in the many electrons Schroedinger equation, and it requires us to compute a term called the *exchange-correlation functional*. The multimarginal optimal transportation cost we consider in this paper is an approximation from below of the exchange-correlation functional in Kohn–Sham DFT. The main result of this paper reduces an optimization problem over measures on $\mathbb{R}^{Nd}$ to a problem over measures on $\mathbb{R}^N$ and is useful in deriving numerical methods to evaluate the mentioned approximation. The one-dimensional result is related to the so called quantum wires. The generalization of this result to higher dimensions is open. In dimension $d = 1$, we also show that the optimal map is unique in the class of symmetric optimal transport plans; this is not the case in higher dimension, as shown in [17] with a radial probability measure ρ as marginal.

We state our result in the natural class of repulsive cost functions of the form

$$c(x_1, \dots, x_N) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} f(x_i - x_j), \tag{1.5}$$

with

$$f\colon \mathbb{R} \to [0, +\infty] \text{ even.} \tag{1.6}$$

In particular, when $f(x) = |x|^{-1}$, we obtain the Coulomb cost (1.1). The main result is the following theorem.

Theorem 1.1 *Let $f\colon \mathbb{R} \to [0,+\infty]$ be a function that satisfies (1.6) and let c be the cost (1.5). Let ρ be a non-atomic probability measure on $\mathbb{R}$ such that $(K) < \infty$. Let $-\infty = d_0 < d_1 < \cdots < d_N = +\infty$ be such that*

$$\rho([d_i, d_{i+1}]) = 1/N, \qquad i = 0, \dots, N-1. \tag{1.7}$$

Let $T\colon \mathbb{R} \to \mathbb{R}$ be the unique (up to ρ-null sets) function increasing on each interval $[d_i, d_{i+1}]$, $i = 0, \dots, N-1$, and such that

$$\begin{aligned} T_\sharp(1_{[d_i,d_{i+1}]}\rho) &= 1_{[d_{i+1},d_{i+2}]}\rho, \quad i = 0, \dots, N-2, \\ T_\sharp(1_{[d_{N-1},d_N]}\rho) &= 1_{[d_0,d_1]}\rho. \end{aligned} \tag{1.8}$$

Then T is an admissible map for (M_{cycl}) and

$$(K) = \int_{\mathbb{R}} c(x, T(x), T^{(2)}(x), \dots, T^{(N-1)}(x))\, d\rho. \tag{1.9}$$

Moreover, the only symmetric optimal transport plan is the symmetrization of the plan induced by the map T.

In [3, Section VII C] the particular case when the probability density is uniform on some interval in $\mathbb{R}$ is studied. In this case, indeed, an explicit concave potential can be built.

The result is based on a careful analysis of the structure of c-cyclically monotone sets (see Definition 2.1). Indeed, the analytic property of minimizing the Kantorovich problem (K) implies, through standard results in [15, Lemma 2] (based on the two marginal theory in [20] and recalled in Proposition 2.2), that the support of the optimal plan is a c-monotone set. In turn, the c-monotonicity can be read in terms of a geometric property of the support (see Definition 2.3 and Corollary 2.5). It follows from the structure of c-cyclically monotone sets that any optimal symmetric plan induces a splitting of $\mathbb{R}$ in N parts $(-\infty, d_1], [d_1, d_2], \dots, [d_N, \infty)$ with equal mass. If we look the restriction of the optimal plan to the set $\{x_1 < x_2 \cdots < x_n\}$, the marginals are exactly the restriction of ρ to the segments, and the plan here is obtained by mapping each piece in the following one through a monotone map T according to (1.8).

Remark 1.2 One might wonder if it is possible to give a detailed description of any optimal plan or any optimal map. Apart from the special case $N = 2$, we can construct a wild class of optimal cyclical maps. For simplicity, we build our example with $N = 3$ and for a uniform density ρ in $[0, 1]$, and the Coulomb cost (the construction is similar for a generic N, ρ, and a cost of the form (1.5)). Let A_0 be any measurable subset of $[0, 1/3)$ and let $B_0 = [0, 1/3) \setminus A_0$; then we define $A_1 = A_0 + 1/3$, $A_2 = A_0 + 2/3$, and B_1 and B_2 in a similar way. Then we consider the map

$$T(x) = \begin{cases} x + 1/3 & \text{if } x \in A_0 \cup A_1, \\ x - 2/3 & \text{if } x \in A_2, \\ x + 2/3 & \text{if } x \in B_0, \\ x - 1/3 & \text{if } x \in B_1 \cup B_2. \end{cases}$$

It is easy to see that the pushforward of $\mathcal{L}|_{A_i}$ through T is $\mathcal{L}_{A_{i+1}}$, and the pushforward of $\mathcal{L}|_{B_i}$ is $\mathcal{L}|_{B_{i-1}}$ for $i = 0, 1, 2$ (where the subscripts are considered modulo 3). So it is clear that $T_\sharp\rho = \rho$, $T^{(3)} = \mathrm{Id}$, and that the associated cost is the minimum possible, $15/2$ (see computations in [3, Section VII C]). In conclusion the best uniqueness result can be obtained in the class of symmetric plans.

Remark 1.3 In a recent paper [7], the authors showed that, given a single particle distribution ρ, as the number of particles goes to ∞, the optimal symmetric plan with all marginals ρ converges to the solution of a suitable optimal transport problem. In turn, the solution can be explicitly computed, and it is the independent product measure all of whose factors are given by ρ.

This result can be easily deduced from Theorem 1.1 in our one dimensional setting. Indeed, in our case the optimal symmetric plan γ_n is obtained by splitting the support of ρ in n parts of equal mass and sending each part onto the other $n-1$. As $n \to \infty$, the plans γ_n weakly converge to the product measure with infinity many marginals all equal to ρ.

Theorem 1.1 provides an explanation of the numerical results presented, for example, in [14, fig. 2] or [13]. More precisely, the qualitative behavior of the Kantorovich potential (or of the Kohn–Sham potential which is sometimes very close) numerically observed in the papers above may be explained. These potentials are important elements of the theory. A Kantorovich potential is a maximizer for the dual problem of (1.3) given by

$$(D) = \max\Big\{ \int_{\mathbb{R}} u(x)d\rho(x) : \ u\colon \mathbb{R} \to \mathbb{R} \text{ is continuous, and for every } x_1, \dots, x_N \in \mathbb{R}^d, \quad u(x_1) + \cdots + u(x_N) \le \sum_{i=1}^{N}\sum_{i<j} \frac{1}{|x_i - x_j|}\Big\}.$$

The optimality conditions (complementary slackness) read as follows:

$$u'(x) = -\sum_{i=1}^{N-1} \frac{x - T^{(i)}(x)}{|x - T^{(i)}(x)|^3}.$$

In the following example we consider an atom similar to Beryllium with one nucleus (located in the origin) and 3 electrons. As electron density we consider

$$\rho = \frac{1}{\sqrt{\pi}} e^{-x^2} dx,$$

which has the right qualitative behavior. We will show that a Kantorovich potential is concave near the peak of the density (the nucleus) and convex in a region distant from the nucleus. This is apparently the opposite of the numerical results in the papers above, but the discrepancy is explained by the fact that in those papers the authors compute $-u$ (see [14, equation (2.5)]). Consider $a \in \mathbb{R}$ such that $\rho((-\infty, -a]) = \rho([-a, a]) = \rho([a, +\infty)) = 1/3$ and $b \in \mathbb{R}$ such that $\rho((-\infty, -b]) = \rho([b, +\infty)) = 1/6$ and let u be a Kantorovich potential. Let

$x_1 < x_2 \in [-a, a]$, for both points

(1.10)

$$u'(x_i) = -\frac{x_i - T(x_i)}{|x_i - T(x_i)|^3} - \frac{x_i - T^{(2)}(x_i)}{|x_i - T^{(2)}(x_i)|^3} = \frac{1}{|x_i - T(x_i)|^2} - \frac{1}{|x_i - T^{(2)}(x_i)|^2}.$$

By the structure of T and ρ, we have that

$$x_2 - x_1 < T(x_2) - T(x_1) \quad \text{and} \quad x_2 - x_1 < T^{(2)}(x_2) - T^{(2)}(x_1)$$

and therefore, since

$$|x_1 - T(x_1)| = T(x_1) - x_2 + x_2 - x_1 < T(x_1) - x_2 + T(x_2) - T(x_1) = |T(x_2) - x_2|,$$

we obtain

$$|x_1 - T(x_1)| < |x_2 - T(x_2)| \quad \text{and} \quad |x_1 - T^{(2)}(x_1)| > |x_2 - T^{(2)}(x_2)|.$$

Summarizing the inequalities above we obtain from (1.10),

$$u'(x_2) = \frac{1}{|x_2 - T(x_2)|^2} - \frac{1}{|x_2 - T^{(2)}(x_2)|^2} < \frac{1}{|x_1 - T(x_1)|^2} - \frac{1}{|x_1 - T^2(x_1)|^2}$$
$$= u'(x_1),$$

which proves that u is concave in $[-a, a]$.

We now prove that u is convex in $[b, +\infty)$ and then also in $(-\infty, -b]$. Let $x_1 < x_2 \in [b, +\infty)$, for both points

$$u'(x_i) = -\frac{x_i - T(x_i)}{|x_i - T(x_i)|^3} - \frac{x_i - T^{(2)}(x_i)}{|x_i - T^{(2)}(x_i)|^3} = -\frac{1}{|x_i - T(x_i)|^2} - \frac{1}{|x_i - T^{(2)}(x_i)|^2}.$$

As before, by the structure of T and ρ, we have that

$$x_2 - x_1 > T(x_2) - T(x_1) \quad \text{and} \quad x_2 - x_1 > T^{(2)}(x_2) - T^{(2)}(x_1),$$

and therefore

$$|x_1 - T(x_1)| < |x_2 - T(x_2)| \quad \text{and} \quad |x_1 - T^{(2)}(x_1)| > |x_2 - T^{(2)}(x_2)|.$$

Summarizing the inequalities above we conclude that $u'(x_1) < u'(x_2)$, namely the convexity of u in $[b, \infty)$.

2 A Necessary Condition for Optimality

Let $\gamma \in \Pi(\rho)$ be a transport plan.

Definition 2.1 Let $c\colon \mathbb{R}^N \to [0, +\infty]$ be a cost function. We say that a set $\Gamma \subset \mathbb{R}^N$ is c-monotone with respect to $p \subseteq \{1, \dots, N\}$ if

(2.1) $$c(x) + c(y) \le c(X(x, y, p)) + c(Y(x, y, p)) \qquad \forall x, y \in \Gamma,$$

where $X(x, y, p), Y(x, y, p) \in \mathbb{R}^N$ are obtained from x and y by exchanging their coordinates on the complement of p, namely, for $i \in \{1, \dots, N\}$,

$$X_i(x, y, p) = \begin{cases} x_i & \text{if } i \in p, \\ y_i & \text{if } i \notin p, \end{cases} \qquad Y_i(x, y, p) = \begin{cases} y_i & \text{if } i \in p, \\ x_i & \text{if } i \notin p. \end{cases}$$

We say that $\Gamma \subset \mathbb{R}^N$ is c-monotone if (2.1) holds true for every $p \subseteq \{1, \dots, N\}$.

https://doi.org/10.4153/CJM-2014-011-x Published online by Cambridge University Press

The following Proposition ([15, Lemma 2] based on the two marginal theory in [20]; see also [10]) contains a necessary condition for optimality.

Proposition 2.2 *Let $c\colon \mathbb{R}^N \to [0, +\infty]$ be a continuous cost and let ρ be a probability measure on $\mathbb{R}$. Let $\gamma \in \Pi(\rho)$ be an optimal transport plan for problem (1.3) and assume $(K) < \infty$. Then* $\operatorname{spt}\gamma$ *is c-monotone.*

The idea of the previous proposition is that, if c-monotonicity fails for x and y, it is possible to rearrange the mass of the optimal plan γ close to x and y constructing a new plan with inferior cost.

The rest of this section is dedicated to the study of c-monotone sets that exhibit a particular geometric structure related to our choice of repulsive costs c. It is expressed in terms of the following definition. Given $x \in \mathbb{R}^N$ there exists (at least) a permutation $\sigma \in \mathcal{S}_N$ such that

$$x_{\sigma(1)} \leq x_{\sigma(2)} \leq \cdots \leq x_{\sigma(N)}.$$

We denote by $x^* = (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(N)})$, and x^* will be called the increasing rearrangement of x.

Definition 2.3 We say that two points $x, y \in \mathbb{R}^N$ are well ordered if considering the increasing rearrangements x^* of x and y^* of y, either

$$x_1^* \leq y_1^* \leq x_2^* \leq y_2^* \leq \cdots \leq x_N^* \leq y_N^* \quad \text{or} \quad y_1^* \leq x_1^* \leq y_2^* \leq x_2^* \leq \cdots \leq y_N^* \leq x_N^*.$$

We remark that, since the cost c is symmetric, we have that $c(x) = c(x^*)$ for every $x \in \mathbb{R}^N$. The following proposition relates c-monotonicity to the well ordering property of couples of points. In particular, it provides a geometric characterization of $c-$monotonicity peculiar to the repulsive cost functions.

Proposition 2.4 *Let f and c be defined as in Theorem 1.1. Let $x, y \in \mathbb{R}^N$.*
If x, y are well ordered, then

$$c(x) + c(y) \leq c\big(X(x, y, p)\big) + c\big(Y(x, y, p)\big) \qquad \forall p \subseteq \{1, \ldots, N\}. \tag{2.2}$$

Moreover, if $c(x) + c(y) < \infty$, then x and y are well ordered if and only if

$$c(x) + c(y) = \min\Big\{ c\big(X(x, y, p)\big) + c\big(Y(x, y, p)\big) : p \subseteq \{1, \ldots, N\}\Big\}. \tag{2.3}$$

Proposition 2.4 has two interesting consequences. First, putting together Propositions 2.2 and 2.4 we obtain a necessary geometric condition on the support of an optimal plan for problem (K). Then we deduce that the support of an optimal plan is disjoint from the diagonals. We state and prove these corollaries here, postponing the proof of Proposition 2.4 to Section 3.

Corollary 2.5 *Let f, c be defined as in Theorem 1.1. Let ρ be a probability measure on $\mathbb{R}$ such that $(K) < +\infty$ and let γ be an optimal plan for problem (K).*
Then every $x, y \in \operatorname{spt}\gamma$ is well ordered.

Proof By Proposition 2.2 we have that

$$c(x)+c(y)=\min\big\{c\big(X(x,y,p)\big)+c\big(Y(x,y,p)\big):p\subseteq\{1,\dots,N\}\big\}\quad x,y\in\operatorname{spt}\gamma.$$

By Proposition 2.4 we obtain that

$$x,y \text{ are well ordered}\qquad x,y\in\operatorname{spt}\gamma\cap\{c<\infty\}. \tag{2.4}$$

Now we want to prove the same for every $x,y\in\operatorname{spt}\gamma$. To this end, we remark that the property of being well ordered is a closed condition. Indeed, setting

$$g(x,y)=\big[(x_1^*-y_1^*)^++\cdots+(x_N^*-y_N^*)^+\big]\cdot\big[(y_1^*-x_1^*)^++\cdots+(y_N^*-x_N^*)^+\big],$$

we have that g is continuous thanks to the fact that $(x,y)\mapsto(x^*,y^*)$ is continuous and $g(x,y)=0$ if and only if x and y are well ordered. By (2.4), every couple of points in the closure of $\operatorname{spt}\gamma\cap\{c<\infty\}$ is well ordered, so we are left to prove that the closure of $\operatorname{spt}\gamma\cap\{c<\infty\}$ is $\operatorname{spt}\gamma$. This is in turn true, because $\operatorname{spt}\gamma$ is closed and $\operatorname{spt}\gamma\cap\{c<\infty\}$ is a concentration set for γ, since $(K)<+\infty$. Then $\gamma(\{c=\infty\})=0$. ∎

Corollary 2.6 *Let f, c, and ρ be defined as in Theorem 1.1. Let γ be an optimal transport plan for the problem (K) with $(K)<\infty$. Then*

$$\operatorname{spt}\gamma\cap\{x_i=x_j\}=\varnothing\quad\textit{for all } i,j\in\{1,\dots,N\} \textit{ with } i\neq j.$$

Proof Let us assume by contradiction that there exists $x\in\operatorname{spt}(\gamma)$ such that $x_i=x_j=t$ for some $i\neq j$; this means that there exists $k\in\{1,\dots,N\}$ such that $x_k^*=x_{k+1}^*=t$. Now take another point $y\in\operatorname{spt}(\gamma)$; by Corollary 2.5 we know that x and y are well ordered, in particular, either $x_k^*\le y_k^*\le x_{k+1}^*$ or $x_k^*\le y_{k+1}^*\le x_{k+1}^*$. In both cases we have that a coordinate of y is equal to t, so we have that

$$\operatorname{spt}\gamma\subseteq\{x_1=t\}\cup\{x_2=t\}\cup\cdots\cup\{x_N=t\}.$$

Since, by assumption, ρ has no atoms and by the condition on the marginals, we have that $\gamma(\{x_i=t\})=\rho(\{t\})=0$. Since the support of a probability measure γ cannot be contained in a finite union of γ-null sets, we obtain a contradiction. ∎

3 A Geometric Characterization of c-monotonicity

We now focus on the proof of Proposition 2.4. In the following inequalities, the analysis of equality cases will play a crucial role in understanding c-monotonicity. We point out that this analysis can be written in a simplified way if one assumes that all the components of x and y in Proposition 2.4 (and in the following lemmas) are different real numbers. In the general case, to fully describe equality cases we need a notion of equality of discrete sets that counts the elements with multiplicity. To this aim, we give the following definition.

Definition 3.1 Given a vector $x=(x_1,\dots,x_N)\in(\mathbb{R}^m)^N$ we define the measure

$$\mu_x=\sum_{i=1}^N\delta_{x_i},$$

which represents the set $\{x_1,\dots,x_N\}$ with elements counted with multiplicity.

https://doi.org/10.4153/CJM-2014-011-x Published online by Cambridge University Press

Given $x = (x_1, \dots, x_N), y = (y_1, \dots, y_N) \in (\mathbb{R}^m)^N$, we say that two sets $\{x_1, \dots, x_N\}$ and $\{y_1, \dots, y_N\}$ counted with multiplicities are equal if the corresponding measures are equal, namely $\mu_x = \mu_y$. With this notation it is clear that

$$\mu_{(x,y)} = \mu_x + \mu_y, \qquad x, y \in (\mathbb{R}^m)^N$$

and that

$$\text{for every } x, x' \in \mathbb{R}^N \text{ we have } \mu_x = \mu_{x'} \text{ if and only if } x^* = x'^*. \tag{3.1}$$

The next lemma can be read as a description of optimal transport plans between purely atomic measures on $\mathbb{R}$, following the spirit of [2, Proposition 4.5]. Since the assumptions on our function ϕ are different from the ones in [2], we prove the lemma for the sake of completeness.

Lemma 3.2 *Let $\phi\colon \mathbb{R} \to [0, \infty]$ be a convex function, which is strictly convex in its finiteness domain. Let $N \in \mathbb{N}$, and let $x_1 \leq \cdots \leq x_N$ and $y_1 \leq \cdots \leq y_N$ be two N-uples of real numbers.*

Then

$$\sum_{i=1}^{N} \phi(x_i - y_i) \leq \sum_{i=1}^{N} \phi(x_i - y_{\sigma(i)}) \qquad \forall \sigma \in \mathcal{S}_N. \tag{3.2}$$

Moreover, if the left-hand side in (3.2) is finite, we have equality in (3.2) for some $\sigma \in \mathcal{S}_n$ if and only if we have the equality of measures

$$\sum_{i=1}^{N} \delta_{(x_i, y_i)} = \sum_{i=1}^{N} \delta_{(x_i, y_{\sigma(i)})}. \tag{3.3}$$

Proof *Step one.* If $i, j \in \{1, \dots, N\}$ are such that

$$x_i \leq x_j \quad \text{and} \quad y_{\sigma(j)} \leq y_{\sigma(i)}, \tag{3.4}$$

then we have that

$$\phi(x_i - y_{\sigma(j)}) + \phi(x_j - y_{\sigma(i)}) \leq \phi(x_i - y_{\sigma(i)}) + \phi(x_j - y_{\sigma(j)}). \tag{3.5}$$

Moreover, if both the inequalities in (3.4) are strict and the left-hand side in (3.5) is finite, then the inequality in (3.5) is also strict.

In fact we have that $x_i - y_{\sigma(i)} \leq x_i - y_{\sigma(j)} \leq x_j - y_{\sigma(j)}$ (with strict inequalities if $y_{\sigma(i)} < y_{\sigma(j)}$ and $x_i < x_j$), and so there exists $t \in [0, 1]$ ($t \in (0, 1)$ if the inequalities are strict) such that

$$x_i - y_{\sigma(j)} = t(x_i - y_{\sigma(i)}) + (1 - t)(x_j - y_{\sigma(j)}).$$

A brief computation shows that

$$x_j - y_{\sigma(i)} = (1 - t)(x_i - y_{\sigma(i)}) + t(x_j - y_{\sigma(j)}).$$

Applying Jensen's inequality to these points and summing, we get (3.5) (with strict inequalities if $t \in (0, 1)$ and the values in the left-hand side are finite, thanks to the strict convexity of ϕ).

Step two. We prove (3.2) and we characterize the equality cases.

We notice that if (3.3) holds true, we have equality in (3.2). Without loss of generality we can assume that the right-hand side in (3.2) is finite. Let us set $A_\sigma =$

$\sum_{i=1}^N \phi(x_i - y_{\sigma(i)})$. We prove by induction on N that $A_{\mathrm{Id}} \le A_\sigma$ for every $\sigma \in \mathcal{S}_N$ with $A_\sigma < \infty$, where Id is the identical permutation, and that if equality holds in (3.2), then (3.3) holds. For $N = 1$ there is nothing to prove. Let $1 < N$. If $\sigma(1) = 1$, we can eliminate the first term in both sides of the inequality and then use the inductive hypothesis. Hence,

$$A_{\mathrm{Id}} \le A_\tau \qquad \text{for all } \tau \in \mathcal{S}_N \text{ such that } \tau(1) = 1 \text{ and } A_\tau < \infty, \tag{3.6}$$

with equality if and only if

$$\sum_{i=1}^N \delta_{(x_i,y_i)} = \sum_{i=1}^N \delta_{(x_i,y_{\tau(i)})}. \tag{3.7}$$

Suppose that $\sigma(1) \neq 1$, and consider k such that $\sigma(k) = 1$. From Step 1 applied to $i = 1$ and $j = k$, noticing that (3.4) is satisfied because $x_1, \dots, x_N$ and $y_1, \dots, y_N$ are ordered, we have that

$$\phi(x_1 - y_1) + \phi(x_k - y_{\sigma(1)}) \le \phi(x_1 - y_{\sigma(1)}) + \phi(x_k - y_1). \tag{3.8}$$

There is equality in (3.8) if and only if $x_1 = x_k$ or $y_{\sigma(1)} = y_1$, or, in other words, if and only if

$$\big\{ (x_1, y_{\sigma(1)}),\ (x_k, y_1) \big\} = \big\{ (x_k, y_{\sigma(1)}),\ (x_1, y_1) \big\}. \tag{3.9}$$

If we consider $\tau \in \mathcal{S}_N$ defined by $\tau(1) = 1$, $\tau(k) = \sigma(1)$, and $\tau(j) = \sigma(j)$ for every $j \neq 1, k$ then (3.8) yields that $A_\sigma \ge A_\tau$, but since $\tau(1) = 1$ by (3.6), we get that $A_{\mathrm{Id}} \le A_\sigma$.

Regarding the equality cases, if equality holds true in (3.2) for some $\sigma \in \mathcal{S}_N$, then we have that both (3.7) and (3.9) hold true, which together prove (3.3). ∎

To prove Proposition 2.4 we rewrite the cost function as a sum on l-neighbors and prove an inequality for fixed l, which is the main core of the proof (Lemma 3.4). In this way, when exchanging the coordinates to obtain a decreasing variation of the cost, one has to track the effect on fewer terms. Before stating the results, we introduce this notation (see Figure 3.1 for two examples with $N = 3$).

Definition 3.3 Let $x \in \mathbb{R}^N$ and let $x_1^* \le x_2^* \cdots \le x_N^*$ be the coordinates of x^*. We define the $N - l$-tuple $V^l(x) \in (\mathbb{R}^2)^{N-l}$ of couples $(x_i^*, x_{i+l}^*) \in \mathbb{R}^2$ with $i \in \{1, \dots, N - l\}$, and we associate with $V^l(x)$ the measure

$$\mu_{V^l(x)} = \sum_{i=1}^{N-l} \delta_{(x_i^*, x_{i+l}^*)},$$

which represents the set of l-neighbors counted with multiplicity. We also define $V(x)$ as the $\binom{N}{2}$-tuple of $V^l(x)$ neighbors as l varies, and we associate with $V(x)$ the measure

$$\mu_{V(x)} := \sum_{l=1}^{N-1} \mu_{V^l(x)} = \sum_{l=1}^{N-1} \sum_{i=1}^{N-l} \delta_{(x_i^*, x_{i+l}^*)} = \sum_{1 \le i < j \le N} \delta_{(x_i^*, x_j^*)}.$$

We remark that, given f satisfying (1.6) and c defined as (1.5), we have that

$$c(x) = c(x^*) = \int_{\mathbb{R}^2} f(z - z')\, d\mu_{V(x)}(z, z').$$

https://doi.org/10.4153/CJM-2014-011-x Published online by Cambridge University Press

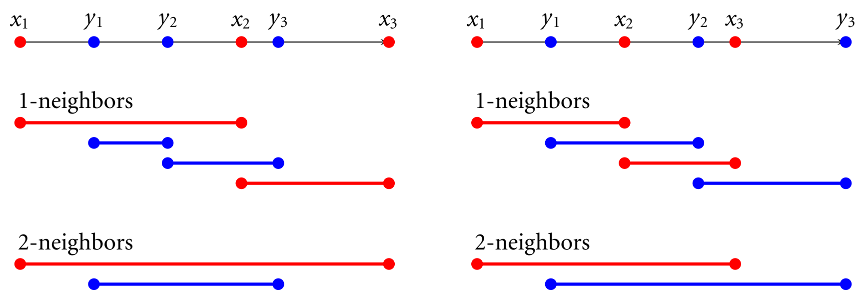

Figure 3.1: The figure shows the coordinates of $x, y \in \mathbb{R}^3$ in two different situations; in particular, in the second case the coordinates of x and y are chosen to make x and y well ordered. In both cases, the sets of 1-neighbors $V^1(x) \cup V^1(y)$ and 2-neighbors $V^2(x) \cup V^2(y)$ are represented, drawing a segment between the endpoints of each couple of neighboring coordinates.

Given a couple of points $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$, let $x_1^* \le x_2^* \cdots \le x_N^*$ be the coordinates of x^* , and $y_1^* \le y_2^* \le \cdots \le y_N^*$ be the coordinates of y^*; we consider the set of l-neighbors of x and of y, namely

$$\mu_{(V^l(x),V^l(y))} = \mu_{V^l(x)} + \mu_{V^l(y)} = \sum_{i=1}^{N-l} \delta_{(x_i^*,x_{i+l}^*)} + \sum_{i=1}^{N-l} \delta_{(y_i^*,y_{i+l}^*)}.$$

In the same way we have that

$$\mu_{(V(x),V(y))} = \sum_{l=1}^{N-1} \mu_{(V^l(x),V^l(y))}$$

and

$$c(x) + c(y) = c(x^*) + c(y^*) = \int_{\mathbb{R}^2} f(z - z')\, d\mu_{(V(x),V(y))}(z, z') \tag{3.10}$$
$$= \sum_{l=1}^{N-1} \int_{\mathbb{R}^2} f(z - z')\, d\mu_{(V^l(x),V^l(y))}(z, z').$$

The following lemma represents our fundamental estimate on l-neighbors for fixed l.

Lemma 3.4 *Let f be defined as in Theorem 1.1. Let m, M be positive integers such that $m < M$. Let I be a subset of $\{1, \dots, M\}$ with m elements and $s\colon I \to \{1, \dots, M\}$ be an injective map such that $i \le s(i)$ for all i. Let $z_1 \le z_2 \le \cdots \le z_M$.*

Then we have that

$$\sum_{i=1}^{m} f(z_{i+M-m} - z_i) \le \sum_{i \in I} f(z_{s(i)} - z_i). \tag{3.11}$$

Moreover, if the left-hand side in (3.11) is finite, equality holds in (3.11) for some s if and only if the sets $\{(z_i, z_{i+M-m})\}_{i\in\{1,\dots,m\}}$ and $\{(z_i, z_{s(i)})\}_{i\in\{1,\dots,m\}}$ coincide with

https://doi.org/10.4153/CJM-2014-011-x Published online by Cambridge University Press

multiplicity, namely

$$\sum_{i=1}^{m} \delta_{(z_i, z_{i+M-m})} = \sum_{i \in I} \delta_{(z_i, z_{s(i)})}. \tag{3.12}$$

Proof It is clear that (3.12) implies equality in (3.11). We assume in the following that the right-hand side in (3.11) is finite; otherwise, there is nothing to prove.

Step one: from s to its monotone rearrangement. Let us call $s(I)$ the image of I through s and let $s^*\colon I \to s(I)$ be the strictly monotone function that maps I onto $s(I)$, which exists and is unique thanks to the fact that I and $s(I)$ have the same number of elements. Then we have that

$$\sum_{i \in I} f(z_{s^*(i)} - z_i) \le \sum_{i \in I} f(z_{s(i)} - z_i). \tag{3.13}$$

Let $\phi\colon \mathbb{R} \to [0, \infty]$ be defined as

$$\phi(x) = \begin{cases} +\infty & \text{if } x \in (-\infty, 0), \\ f(x) & \text{if } x \in [0, \infty). \end{cases}$$

Thanks to Lemma 3.2 applied with $N = m$, to the ordered m-uples of elements of $(z_i)_{i \in I}$ and $(z_{s^*(i)})_{i \in I}$, we obtain

$$\sum_{i \in I} \phi(z_{s^*(i)} - z_i) \le \sum_{i \in I} \phi(z_{s(i)} - z_i) = \sum_{i \in I} f(z_{s(i)} - z_i), \tag{3.14}$$

where the last equality follows from the assumption $i \le s(i)$ for every $i \in I$ and the monotonicity of $z_1, \dots, z_M$. As a consequence, since the right-hand side in (3.14) is finite we also obtain that $z_i \le z_{s^*(i)}$ for every $i \in I$. We can then rewrite (3.14) as (3.13). Moreover, we have equality in (3.14) if and only if

$$\sum_{i \in I} \delta_{(z_i, z_{s^*(i)})} = \sum_{i \in I} \delta_{(z_i, z_{s(i)})}. \tag{3.15}$$

Step two: conclusion. Denote by $i_1 < i_2 < \cdots < i_m$ the elements of I. Then

$$l \le i_l, \qquad s^*(i_l) \le M - m + l, \text{for } l \in \{1, \dots, m\}. \tag{3.16}$$

Indeed thanks to the fact that $i_l < i_{l+1}$ are positive integers for every $l = 1, \dots, m$, it follows easily by induction that $l \le i_l$ for every $l = 1, \dots, m$. In the same way from $s^*(i_l) < s^*(i_{l+1})$ for every $l = 1, \dots, m$ it follows, by induction, that $i_l \le l + M - m$ for every $l = 1, \dots, m$.

From (3.16) and (3.13), since $z_i \le z_{s^*(i)}$ for every $i \in I$ and by the monotonicity of f in $[0, +\infty)$, it follows that

$$\sum_{l=1}^{m} f(z_{l+M-m} - z_l) \le \sum_{l=1}^{m} f(z_{s^*(i_l)} - z_{i_l}) = \sum_{i \in I} f(z_{s^*(i)} - z_i) \le \sum_{i \in I} f(z_{s(i)} - z_i). \tag{3.17}$$

In particular, since f is strictly decreasing in $[0, \infty)$ because it is strictly convex, if equality holds in (3.17), then we have that $z_{l+M-m} = z_{s^*(i_l)}$ and $z_l = z_{i_l}$ for every $l = 1, \dots, m$. If equality holds, then we have that

$$\sum_{l=1}^{m} \delta_{(z_l, z_{l+M-m})} = \sum_{l=1}^{m} \delta_{(z_{i_l}, z_{s^*(i_l)})} = \sum_{i \in I} \delta_{(z_i, z_{s^*(i)})}. \tag{3.18}$$

From (3.15) and (3.18) we obtain that the equality cases in (3.11) are described by (3.12). ∎

In the following lemma we prove that the set of neighbors $V(x) \cup V(y)$ of some $x, y \in \mathbb{R}^N$ determines uniquely (x, y), in the sense that if the set of neighbors of two couples of points (x, y) and (x', y') coincide, then the two couples coincide up to the order. Since the statement is of purely combinatorial nature, the reader can skip the proof of the lemma at a first reading.

Lemma 3.5 *Let $(x, y), (x', y') \in \mathbb{R}^{2N}$, whose sets of coordinates counted with multiplicities are equal,* i.e.,

$$\mu_{(x,y)} = \mu_{(x',y')} = \mu. \tag{3.19}$$

Let us assume that the set of neighbors (counted with multiplicities) is the same for both the couples, that is,

$$\mu_{(V(x),V(y))} = \mu_{(V(x'),V(y'))}. \tag{3.20}$$

Then we have that

$$\{x^*, y^*\} = \{x'^*, y'^*\}. \tag{3.21}$$

Proof We first observe that the relation $\mu_{(x,y)} = \mu_{(x',y')}$ can be written as

$$\sum_{i=1}^{N} \delta_{x_i^*} + \delta_{y_i^*} = \sum_{i=1}^{N} \delta_{x_i'^*} + \delta_{y_i'^*},$$

and it also gives us that

$$\sum_{i=1}^{N} \delta_{(x_i^*,x_i^*)} + \delta_{(y_i^*,y_i^*)} = \sum_{i=1}^{N} \delta_{(x_i'^*,x_i'^*)} + \delta_{(y_i'^*,y_i'^*)}. \tag{3.22}$$

Writing (3.20) and then exchanging the coordinates of $\mathbb{R}^2$ we also obtain

$$\sum_{i<j} \delta_{(x_i^*,x_j^*)} + \delta_{(y_i^*,y_j^*)} = \sum_{i<j} \delta_{(x_i'^*,x_j'^*)} + \delta_{(y_i'^*,y_j'^*)}, \tag{3.23}$$

$$\sum_{j<i} \delta_{(x_i^*,x_j^*)} + \delta_{(y_i^*,y_j^*)} = \sum_{j<i} \delta_{(x_i'^*,x_j'^*)} + \delta_{(y_i'^*,y_j'^*)}. \tag{3.24}$$

Now, summing up (3.22), (3.23), and (3.24) we obtain

$$\sum_{i,j=1}^{N} \delta_{(x_i^*,x_j^*)} + \delta_{(y_i^*,y_j^*)} = \sum_{i,j=1}^{N} \delta_{(x_i'^*,x_j'^*)} + \delta_{(y_i'^*,y_j'^*)},$$

that can be seen as

$$\mu_x \otimes \mu_x + \mu_y \otimes \mu_y = \mu_{x'} \otimes \mu_{x'} + \mu_{y'} \otimes \mu_{y'}, \tag{3.25}$$

where $\otimes$ represents the tensor product between measures on $\mathbb{R}$. Let $\phi \in \mathcal{C}_c(\mathbb{R})$ and let us consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = \phi(x)\phi(y)$. We test it with the measure equality (3.25) to infer

$$\Big(\int_{\mathbb{R}} \phi \, d\mu_x\Big)^2 + \Big(\int_{\mathbb{R}} \phi \, d\mu_y\Big)^2 = \Big(\int_{\mathbb{R}} \phi \, d\mu_{x'}\Big)^2 + \Big(\int_{\mathbb{R}} \phi \, d\mu_{y'}\Big)^2. \tag{3.26}$$

From (3.26) and

$$\int_{\mathbb{R}} \phi \, d\mu_x + \int_{\mathbb{R}} \phi \, d\mu_y = \int_{\mathbb{R}} \phi \, d\mu_{x'} + \int_{\mathbb{R}} \phi \, d\mu_{y'},$$

which in turn follows from (3.19), we obtain with some computations[1] that

$$\Big(\int_{\mathbb{R}} \phi \, d\mu_x - \int_{\mathbb{R}} \phi \, d\mu_{x'}\Big) \cdot \Big(\int_{\mathbb{R}} \phi \, d\mu_x - \int_{\mathbb{R}} \phi \, d\mu_{y'}\Big) = 0.$$

As a consequence we get that for every $\phi \in \mathcal{C}_c(\mathbb{R})$ either

$$\int_{\mathbb{R}} \phi \, d\mu_x = \int_{\mathbb{R}} \phi \, d\mu_{x'} \quad \text{or} \quad \int_{\mathbb{R}} \phi \, d\mu_x = \int_{\mathbb{R}} \phi \, d\mu_{y'}. \tag{3.27}$$

We claim that either the first equality or the second equality in (3.27) is true for every $\phi \in \mathcal{C}_c(\mathbb{R})$, which allows us to conclude by duality that either $\mu_x = \mu_{x'}$ or $\mu_x = \mu_{y'}$. Namely by (3.1) either $x^* = x'^*$ or $x^* = y'^*$. Thus (3.21) follows thanks to (3.19). To prove the claim, we notice that if the first equality in (3.27) fails for some $\phi \in \mathcal{C}_c(\mathbb{R})$, and the second equality in (3.27) fails for $\psi \in \mathcal{C}_c(\mathbb{R})$, we obtain a contradiction because the function $\phi + \psi$ does not satisfy any of the two conditions in (3.27). ∎

Proof of Proposition 2.4 Let x, y be a couple of well ordered points in $\mathbb{R}^N$; we denote by $z_1 \leq z_2 \leq \cdots \leq z_{2N}$ the cumulative coordinates, so for $i = 1, \dots, N$, we have either

$$\begin{cases} x_i^* = z_{2i+1} \\ y_i^* = z_{2i} \end{cases} \qquad \text{or} \qquad \begin{cases} y_i^* = z_{2i+1} \\ x_i^* = z_{2i} \end{cases}$$

because they are well ordered. By (3.10), inequality (2.2) is equivalent to

$$\sum_{l=1}^{N-1} \int_{\mathbb{R}^2} f(z-z') \, d\mu_{(V^l(x),V^l(y))}(z,z') \leq \sum_{l=1}^{N-1} \int_{\mathbb{R}^2} f(z-z') \, d\mu_{(V^l(X),V^l(Y))}(z,z').$$

This last inequality is true, because we claim that Lemma 3.4 implies that, for every $l = 1, \dots, N-1$,

$$\int_{\mathbb{R}^2} f(z-z') \, d\mu_{(V^l(x),V^l(y))}(z,z') \leq \int_{\mathbb{R}^2} f(z-z') \, d\mu_{(V^l(X),V^l(Y))}(z,z'), \tag{3.28}$$

and so, adding up over l, we get the conclusion. To prove the claim, we rewrite (3.28) in terms of the z_i and we see in fact that the left-hand side is simply

$$\int_{\mathbb{R}^2} f(z-z') \, d\mu_{(V^l(x),V^l(y))}(z,z') = \sum_{i=1}^{2l} f(z_i - z_{2N-2l+i}),$$

while the right-hand side can be described as

$$\int_{\mathbb{R}^2} f(z-z') \, d\mu_{(V^l(X),V^l(Y))}(z,z') = \sum_{i \in I} f(z_i - z_{s(i)}), \tag{3.29}$$

[1] from $a^2 + b^2 = \alpha^2 + \beta^2$ and $a + b = \alpha + \beta$ we obtain $ab = \alpha\beta$. Then $a(a + b - \alpha - \beta) = 0$, developing this equality and replacing ab by $\alpha\beta$ one gets $(a - \alpha)(a - \beta) = 0$.

https://doi.org/10.4153/CJM-2014-011-x Published online by Cambridge University Press

where $I \subseteq \{1,\dots,2N\}$ is a set with $2l$ elements and $s\colon I \to \{1,\dots,2N\}$ is an injective function with $s(i) > i$ (s may not be unique).[2]

Now we can apply Lemma 3.4 to $z_1,\dots,z_{2N}$ with $M = 2N$ and $m = 2l$, I and s as in (3.29), and we get (3.28). Tracing back the equality cases under the assumption that the left-hand side in (2.2) is finite, we have equality in (2.2) if and only if we have equality in (3.28) for every l. In particular, if equality holds in (2.2) by Lemma 3.4, we have $\mu_{(V^l(x),V^l(y))} = \mu_{(V^l(X),V^l(Y))}$ for every l and, summing over l,

$$\mu_{(V(x),V(y))} = \mu_{(V(X),V(Y))}.$$

Therefore we apply Lemma 3.5 because clearly $\mu_{(x,y)} = \mu_{(X,Y)}$, and we get $\{x^*,y^*\} = \{X^*,Y^*\}$ in case of equality in (2.2).

With this we proved the first part, that is if (x,y) are well ordered then $c(x)+c(y)$ is a minimizer among the values of $c(X)+c(Y)$ varying the partition. However the description of equality cases also lets us conclude the second part. Let us take a couple x,y that is a minimizer among all partitions, namely satisfies (2.3); we know that there exists a partition such that X,Y are well ordered, and by the first part, this couple is also a minimizer. In particular we get $c(x)+c(y) = c(X)+c(Y)$ and, by the characterization of equality cases, we conclude that $\{x^*,y^*\} = \{X^*,Y^*\}$, and so x,y are also well ordered. ∎

4 Proof of the Main Theorem: Existence, Uniqueness, and Description

We now enter into the proof of Theorem 1.1; the existence part of the proof also holds under slightly weaker assumptions on f, as stated in the next remark.

Remark 4.1 If we drop the strict convexity assumption for f on $[0,\infty)$, assuming only convexity, we still have that the minimum in (M) and (K) is realized by the map described in Theorem 1.1, with a completely analogous proof. In this case, however, there is no longer uniqueness in the class $\Pi_{\mathrm{sym}}(\rho)$; the lack of uniqueness happens also, for instance, in the classical 2-marginal case.

We recall here a standard result in optimal transport theory that will be useful in the last part of the proof of the main theorem.

Lemma 4.2 *Let ρ_1 and ρ_2 be two non-atomic probability measures on $\mathbb{R}$. Let $c(x,y)\colon \mathbb{R}^2 \to [0,\infty]$ be a lower semicontinuous convex function of $x-y$, strictly convex on its finiteness domain, and let us assume that*

$$\min\Big\{\int_{\mathbb{R}\times\mathbb{R}} c(x,y)\,d\gamma(x,y) : \pi^1_\sharp\gamma = \rho_1,\ \pi^2_\sharp\gamma = \rho_2\Big\} < +\infty. \tag{4.1}$$

Then there exists a unique minimizer γ for problem (4.1); moreover, $\gamma = (\mathrm{Id},T)_\sharp\rho_1$, where T is the unique (up to ρ_1-null sets) increasing function such that $\rho_2 = T_\sharp\rho_1$.

[2] For example, if $X^*_i = z_{r(i)}$ and $Y^*_i = z_{t(i)}$ for some strictly increasing functions $r,t\colon \{1,\dots,N\} \to \{1,\dots,2N\}$ whose ranges are disjoint, we can take $I = \{r(1),t(1),\dots,r(N-l),t(N-l)\}$ and define $s\colon I \to \{1,\dots,2N\}$ by $s(i) := r(r^{-1}(i)+l)$ for $i \in \{r(1),\dots,r(N-l)\}$ and $s(i) := t(t^{-1}(i)+l)$ for $i \in \{t(1),\dots,t(N-l)\}$.

Proof Despite quite different assumptions on the cost, the proof goes exactly as in [1, Theorem 3.1]. For a complete proof, see [8, Theorem 1.4] (in italian). ■

Proof of Theorem 1.1 Let us consider an optimal plan $\gamma \in \Pi(\rho)$; we can assume without loss of generality that γ is symmetric thanks to (1.4). Let $O \subset \mathbb{R}^N$ be the subset of ordered N-uples

$$O = \{(x_1, \dots, x_N) : x_1 < x_2 < \cdots < x_N\},$$

and let $\gamma^* = \gamma|_O$. From Corollary 2.6 the optimal plan γ is concentrated outside the union of all diagonals $D = \bigcup_{i \neq j}\{x_i = x_j\}$. Since γ is symmetric, we obtain that

$$\gamma = \gamma|_{\mathbb{R}^N \setminus D} = \sum_{\sigma \in \mathcal{S}_N} \gamma|_{\tau_\sigma(O)} = \sum_{\sigma \in \mathcal{S}_N} (\tau_\sigma)_\sharp \gamma|_{\tau_\sigma(O)} = \sum_{\sigma \in \mathcal{S}_N} (\tau_\sigma)_\sharp \gamma^*. \tag{4.2}$$

Let Γ be the support of γ and $\Gamma^* = \Gamma \cap O$, so that Γ^* is also the support of γ^*. Let us define

$$d_i^+ = \max_{x \in \Gamma^*} x_i \qquad d_i^- = \min_{x \in \Gamma^*} x_i, \qquad \rho_i = \rho|_{[d_i^-, d_i^+]}.$$

We show some properties of $d_i^\pm$, namely that

(i) $d_i^+ \leq d_{i+1}^-$ for all $i = 1, \dots, N-1$;
(ii) $\Gamma^* \subseteq \prod_i [d_i^-, d_i^+]$;
(iii) $\rho([d_i^-, d_i^+]) = 1/N$ for all $i = 1, \dots, N$;
(iv) $(N-1)!\gamma^*$ is a transport plan whose marginals are exactly ρ_i, namely $(N-1)!(\pi^i)_\sharp \gamma^* = \rho_i$ for all $i = 1, \dots, N-1$.

We start with assertion (i). Let us fix an index $i \in \{1, \dots, N-1\}$. We know from Corollary 2.5 that every $x, x' \in \Gamma^*$ are well ordered, but, since they are already ordered, we have that either $x_i \leq x_i' \leq x_{i+1} \leq x_{i+1}'$ or $x_i' \leq x_i \leq x_{i+1}' \leq x_{i+1}$. In both cases we have that $x_i \leq x_{i+1}'$. We let x, x' vary in Γ^*, take the infimum on the right-hand side and the supremum on the left hand side and we obtain (i).

Assertion (ii) follows directly from the definition of $d^\pm$.

To prove (iii) we assume to fix the ideas that $i = 1$, the general case is completely analogous. First we remark that $1 = \gamma(\mathbb{R}^N) = N!\gamma^*(\mathbb{R}^N)$ and that it follows from (ii) that

$$\operatorname{spt}((\tau_\sigma)_\sharp \gamma^*) = \tau_\sigma(\Gamma^*) \subseteq \prod_{i=1}^N [d_{\sigma(i)}^-, d_{\sigma(i)}^+], \qquad \text{for} \sigma \in \mathcal{S}_N. \tag{4.3}$$

Since ρ is non-atomic and $(\pi^1)_\sharp \gamma = \rho$, we have that for every $\sigma \in \mathcal{S}_N$,

$$(\tau_\sigma)_\sharp \gamma^* \big(\{d_1^\pm\} \times \mathbb{R}^{N-1} \big) \leq \gamma \big(\{d_1^\pm\} \times \mathbb{R}^{N-1} \big) = \rho\big(\{d_1^\pm\}\big) = 0.$$

Therefore by (4.2) we know that

$$\begin{aligned} \rho\big([d_1^-, d_1^+]\big) & \\ &= \gamma\big([d_1^-, d_1^+] \times \mathbb{R}^{N-1}\big) = \sum_{\sigma \in \mathcal{S}_N} (\tau_\sigma)_\sharp \gamma^* \big([d_1^-, d_1^+] \times \mathbb{R}^{N-1}\big) \\ &= \sum_{\sigma(1)=1} (\tau_\sigma)_\sharp \gamma^* \big([d_1^-, d_1^+] \times \mathbb{R}^{N-1}\big) + \sum_{\sigma(1) \neq 1} (\tau_\sigma)_\sharp \gamma^* \big((d_1^-, d_1^+) \times \mathbb{R}^{N-1}\big). \end{aligned} \tag{4.4}$$

Recalling (4.3) we have two cases: if $\sigma(1) \neq 1$ the set $(d_1^-, d_1^+) \times \mathbb{R}^{N-1}$ does not intersect $\mathrm{spt}((\tau_\sigma)_\sharp\gamma^*)$, if $\sigma(1) = 1$, we have $\mathrm{spt}((\tau_\sigma)_\sharp\gamma^*) \subseteq [d_1^-, d_1^+] \times \mathbb{R}^{N-1}$ and so $[d_1^-, d_1^+] \times \mathbb{R}^{N-1}$ has full $(\tau_\sigma)_\sharp\gamma^*$ measure. In particular, we get

$$\rho([d_1^-, d_1^+]) = \sum_{\sigma(1)=1} (\tau_\sigma)_\sharp\gamma^*(\mathbb{R}^N) = (N-1)! \cdot \gamma^*(\mathbb{R}^N) = \frac{1}{N}$$

from (4.4), which proves (iii).

Since ρ is a non-atomic probability measure, by (i) and (iii) it follows that the sets $[d_1^-, d_1^+], \dots, [d_n^-, d_n^+]$ are a partition of $\mathbb{R}$ up to sets of ρ-measure 0. In order to fulfill (1.7), we can take $d_i = d_i^+$ for $i = 1, \dots, N-1$, so that $\rho_i = 1_{[d_{i-1}, d_i]}\rho$ and $\rho = \rho_1 + \cdots + \rho_N$.

To prove (iv) we note that

$$\sum_{i=1}^N \rho_i = \rho = (\pi^1)_\sharp\gamma = \sum_{\sigma\in\mathcal{S}_N} (\pi^1)_\sharp((\tau_\sigma)_\sharp\gamma^*) = \sum_{\sigma\in\mathcal{S}_N} \pi_\sharp^{\sigma(1)}\gamma^* = (N-1)! \sum_{i=1}^N \pi_\sharp^i\gamma^*.$$

Since the measures $\pi_\sharp^i\gamma^*$ have essentially disjoint supports (in fact we have $\mathrm{spt}\pi_\sharp^i\gamma^* \subseteq [d_i^-, d_i^+]$, we obtain exactly that $(N-1)!\pi_\sharp^i\gamma^* = \rho_i$, which proves (iv).

We turn now to the proof of (1.9). We introduce the plan associated with T, which is given by $\gamma_T = (\mathrm{Id}, T, \dots, T^{(N-1)})_\sharp\rho$, and we rewrite its cost using that the cost c is symmetric, $[T^{(i-1)}]_\sharp\rho_1 = \rho_i$, and $T^{(N)} = \mathrm{Id}$:

$$\begin{aligned}
\int_{\mathbb{R}} c\, d\gamma_T &= \int_{\mathbb{R}} c(x, T(x), \dots, T^{(N-1)}(x))\, d\rho \\
&= \sum_{i=1}^N \int_{\mathbb{R}} c(x, T(x), \dots, T^{(N-1)}(x))\, d\rho_i \\
&= \sum_{i=1}^N \int_{\mathbb{R}} c(T^{(i-1)}(x), T^{(i)}(x), \dots, T^{(i+N-2)(x)})\, d\rho_1 \\
&= N \int_{\mathbb{R}} c(x, T(x), \dots, T^{(N-1)}(x))\, d\rho_1.
\end{aligned} \tag{4.5}$$

Let us define

$$\gamma_{i,j} = (N-1)!(\pi_i, \pi_j)_\sharp\gamma^*, \qquad 1 \le i < j \le N,$$

which is a transport plan between ρ_i and ρ_j by (iv). If we consider

$$\bar{f}(x, y) = \begin{cases} f(x-y) & \text{if } x \ge y, \\ +\infty & \text{otherwise,} \end{cases}$$

we have that $c(x_1, \dots, x_N) = \sum_{i<j} \bar{f}(x_j, x_i)$ if $(x_1, \dots, x_n) \in O$. Now we can write that

$$\begin{aligned}
\frac{1}{N}\int_{\mathbb{R}^N} c\, d\gamma &= (N-1)! \int_O c\, d\gamma^* = \sum_{i<j} \int_O (N-1)! \bar{f}(x_j, x_i)\, d\gamma^*(x_1, \dots, x_N) \\
&= \sum_{i<j} \int_O \bar{f}\, d\gamma_{i,j}.
\end{aligned} \tag{4.6}$$

Since $\bar{f}(x, y)$ is a strictly convex function of $x-y$ on its finiteness domain, Lemma 4.2 gives us that for every $i, j \in \{1, \dots, N\}$ there is a unique optimal plan relative to the

https://doi.org/10.4153/CJM-2014-011-x Published online by Cambridge University Press

cost $\bar{f}(x, y)$, given by the unique monotone increasing map that maps ρ_i to ρ_j. We claim that this monotone map is $T^{(j-i)}$, with T defined as in (1.8). In fact T, by definition, is an increasing map restricted to the support of every ρ_i and maps ρ_i into ρ_{i+1} cyclically. From this we can deduce that $T^{(j-i)}$ is also monotone restricted to the support of every ρ_i and maps ρ_i into ρ_j. So we get that for every $i, j \in \{1, \dots, N\}$,

$$\int \bar{f}(T^{(i)}, T^{(j)})\, d\rho_1 = \int \bar{f}(x, T^{(j-i)}(x))\, d\rho_i \le \int \bar{f}(x, y)\, d\gamma_{i,j}. \tag{4.7}$$

Summing up (4.7) for $1 \le j < i \le N$ and thanks to (4.5) and (4.6), we obtain

$$\frac{1}{N}\int c\, d\gamma_T = \int c(x, T(x), T^{(2)}(x), \dots, T^{(N-1)}(x))\, d\rho_1 \le \frac{1}{N}\int c\, d\gamma, \tag{4.8}$$

Now, (4.8) shows (1.9); moreover, since γ is optimal, we have equality in (4.7) for every $j > 1$, and so, thanks to the uniqueness in the classical case, we get that $\gamma_{1,j}$ is concentrated on the graph of $(\mathrm{Id}, T^{(j-1)})$ for every $j > 1$, namely on $\{y = T^{(j-1)}(x)\}$. Hence γ^* is concentrated on $\{x_j = T^{(j-1)}(x_1)\}$ for every $j = 2, \dots, N$, and so it is concentrated on the intersection of these sets, that is, the graph of $(T, T^2, \dots, T^{(N-1)})$ over x_1. Following [1, Proposition 2.1], we get that γ^* is determined by T:

$$(N-1)!\gamma^* = (\mathrm{Id}, T, T^2, \dots, T^{(N-1)})_\sharp \rho_1,$$

and thus γ is also determined (using (4.2)) and is therefore unique. ■

References

[1] L. Ambrosio, *Lecture notes on optimal transport problems.* In: Mathematical aspects of evolving interfaces (Funchal, 2000), Lecture Notes in Math., 1812, Springer, Berlin, 2003, pp. 1–52.
[2] L. Ambrosio, B. Kirchheim, and A. Pratelli, *Existence of optimal transport maps for crystalline norms.* Duke Math. J. **125**(2004), no. 2, 207–241. http://dx.doi.org/10.1215/S0012-7094-04-12521-7
[3] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi, *Optimal transport formulation of electronic density-functional theory.* Phys. Rev. A **85** (2012), 062502.
[4] T. Champion and L. De Pascale, *On the twist condition and c-monotone transport plans.* Discrete Contin. Dyn. Sys. **34**(2014), no. 4, 1339–1353. http://dx.doi.org/10.3934/dcds.2014.34.1339
[5] M. Colombo, S. Di Marino, *Equality between Monge and Kantorovich multimarginal problems with Coulomb cost.* Ann. Mat. Pura Appl., to appear. http://dx.doi.org/10.1007/s10231-013-0376-0
[6] C. Cotar, G. Friesecke, C. Klüppelberg, *Density functional theory and optimal transportation with Coulomb cost.* Comm. Pure Appl. Math. **66**(2013), no. 4, 548–599. http://dx.doi.org/10.1002/cpa.21437
[7] C. Cotar, G. Friesecke, B. Pass, *Infinite-body optimal transport with Coulomb Cost.* arxiv:1307.6540
[8] S. Di Marino, *Trasporto ottimo e problemi di evoluzione per sistemi di particelle.* http://cvgmt.sns.it/paper/1862/
[9] A. Fathi and A. Figalli, *Optimal transportation on non-compact manifolds.* Israel J. Math. **175**(2010), 1–59. http://dx.doi.org/10.1007/s11856-010-0001-5
[10] W. Gangbo and R. J. McCann, *The geometry of optimal transportation.* Acta Math. **177**(1996), no. 2, 113–161. http://dx.doi.org/10.1007/BF02392620
[11] N. Ghoussoub and B. Maurey, *Remarks on multidimensional symmetric Monge-Kantorovich problems.* Discrete Contin. Dyn. Syst. **34**(2014), no. 4, 1465–1480.
[12] N. Ghoussoub and A. Moameni, *Symmetric Monge-Kantorovich problems and polar decompositions of vector fields.* arxiv:1302.2886
[13] F. Malet, A. Mirtschink, J. C. Cremon, S. M. Reimann, and P. Gori-Giorgi, *Kohn-Sham density functional theory for quantum wires in arbitrary correlation regimes.* Phys. Rev. B **87**(2013), 115146.
[14] C. B. Mendl and L. Lin, *Towards the Kantorovich dual solution for strictly correlated electrons in atoms and molecules.* arxiv:1210.7117

[15] B. Pass, *On the local structure of optimal measures in the multi-marginal optimal transportation problem*. Calc. Var. Partial Differential Equations **43**(2012), no. 3–4, 529–536. http://dx.doi.org/10.1007/s00526-011-0421-z
[16] ———, *Uniqueness and Monge solutions in the multimarginal optimal transportation problem*. SIAM J. Math. Anal. **43**(2011), no. 6, 2758–2775. http://dx.doi.org/10.1137/100804917
[17] ———, *Remarks on the semi-classical Hohenberg-Kohn functional.* Nonlinearity **26**(2013), no. 9. 2731–2744. http://dx.doi.org/10.1088/0951-7715/26/9/2731
[18] A. Pratelli, *On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation.* Ann. Inst. H. Poincaré Probab. Statist. **43**(2007), no. 1, 1–13. http://dx.doi.org/10.1016/j.anihpb.2005.12.001
[19] M. Seidl, *Strong-interaction limit of density-functional theory*. Phys. Rev. A **60**(1999), 4387.
[20] C. Smith and M. Knott, *On Hoeffding-Fréchet bounds and cyclic monotone relations.* J. Multivariate Anal. **40**(1992), no. 2, 328–334. http://dx.doi.org/10.1016/0047-259X(92)90029-F
[21] C. Villani, *Optimal transport. Old and new*. Grundlehren der Mathematischen Wissenschaften, 338, Springer-Verlag, Berlin, 2009.

Scuola Normale Superiore, 56126 Pisa, Italy
e-mail: maria.colombo@sns.it simone.dimarino@sns.it

Dipartimento di Matematica, Universitá di Pisa, Pisa, Italy
e-mail: depascal@dm.unipi.it