Continuity of multi-marginal optimal transport with repulsive cost

Maria Colombo* Simone Di Marino† Federico Stra‡

January 17, 2019

Abstract

We provide sharp conditions for the finiteness and the continuity of multi-marginal optimal transport with repulsive cost, expressed in terms of a suitable concentration property of the measure. To achieve this result, we analyze the Kantorovich potentials of the optimal plans and we estimate the distance of any optimal plan from the regions where the cost is infinite.

1 Introduction

In recent years, a new mathematical model for the strong interaction limit of the density functional theory (DFT) has been considered. For instance, in [6], Buttazzo, De Pascale and Gori-Giorgi show 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 [10], Cotar, Friesecke and Klüppelberg show that an analogous optimal transportation problem describes the semiclassical limit of DFT in the case of 2 electrons and provides estimates from below in the general case.

In this article we prove the finiteness and continuity of multi-marginal optimal transport with repulsive cost under the assumption that the measure does not concentrate too much. The article is a refinement of the results presented in [5], especially from the point of view of the assumptions, which in our work are shown to be sharp. We aknowledge also the recent preprint [2] in which the finiteness of the cost is proved in a similar fashion by dimension reduction.

To describe the problem, we fix a complete and separable (Polish) metric space (X, d). We consider a repulsive interaction cost given by a symmetric lower semi-continuous function $c : X \times X \to [0, \infty]$ for which there exist two non-increasing right-continuous (or, equivalently, lower semi-continuous) functions $m, M : (0, \infty) \to [0, \infty)$ satisfying

$$m\big(\mathsf{d}(x_1, x_2)\big) \le c(x_1, x_2) \le M\big(\mathsf{d}(x_1, x_2)\big), \qquad \text{for all } x_1, x_2 \in X. \tag{1.1}$$

Moreover sometimes we will need a *strong repulsion* assumption, namely

$$c(x, x) = \infty \ \ \forall x \in X \qquad \text{and} \qquad \lim_{r\to 0^+} m(r) = \lim_{r\to 0^+} M(r) = \infty. \tag{1.2}$$

---

*ETH, Zürich, maria.colombo@math.uzh.ch

†INdAM, Scuola Normale Superiore, Pisa, simone.dimarino@altamatematica.it

‡Scuola Normale Superiore, Pisa, federico.stra@sns.it

Extending $m(0) = M(0) = \infty$ for this last case, the inequality (1.1) still holds for all $x_1, x_2 \in X$. The Coulomb cost fits into this framework as $c(x_1, x_2) = \frac{1}{|x_1 - x_2|}$ on $X = \mathbb{R}^d$, $d \in \mathbb{N}$ and one can take $m(r) = M(r) = \frac{1}{r}$

For every integer $N \geq 2$, define the symmetric interaction cost $c : X^N \to [0, \infty]$ by

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

the cost of a plan $C : \mathscr{P}\big(X^N\big) \to [0, \infty]$ by

$$C(\pi) \coloneqq \int_{X^N} c(x_1, \dots, x_N) \, \mathrm{d}\pi(x_1, \dots, x_N) \tag{1.3}$$

and lastly the optimal transport cost $\mathscr{C} : \mathscr{P}(X) \to [0, \infty]$ associated to a marginal by

$$\mathscr{C}(\rho) \coloneqq \inf\{C(\pi) : \pi \in \Pi_N(\rho)\}, \tag{1.4}$$

where

$$\Pi_N(\rho) \coloneqq \big\{\pi \in \mathscr{P}\big(X^N\big) : P^i_{\#}\pi = \rho \text{ for } i = 1, \dots, N\big\}$$

denotes the set of admissible transport plans and $P^i : X^N \to X$ are the projections on the i-th component for $i = 1, \dots, N$. The existence of a minimizer for the infimum problem in (1.4) follows from standard methods in the calculus of variations as long as c is lower semicontinuous (see for example [1, 30, 28]).

Besides exploring the connection with density functional theory [6, 10, 3, 19], several authors investigated the mathematical properties of the minimizer. A natural question is whether the optimizer is induced by a map, namely if there exists a Borel map $T : X \to X$ such that $T_{\#}\rho = \rho$ (where $T_{\#}\rho$ represents the pushforward measure of the measure ρ through the Borel map T) and an optimizer π in the minimization problem in (1.4) can be represented as $(Id, T, T^{(2)}, \dots, T^{(n-1)})_{\#}\rho$. This question is still widely open, though some results have been obtained in [8, 7, 9, 29] regarding the possibility to approximate the cost of a minimizer with costs of these particular plans, the 1-dimensional case, and the radial case (see also the survey [14]). It is important to mention here also the negative result in [17], where they show (in the case of the repulsive harmonic cost $c(x, y) = -|x - y|^2$) an explicit density ρ, absolutely continuous with respect to the Lebesgue measure, such that there is a unique optimal symmetric plan, which is not induced by a map.

The asymptotic behavior as $N \to \infty$ for the Coulomb were presented in [11, 27] and then more precisely (up to the first order) in [24, 12].

In [13, 5], instead, the authors prove a duality result, which shows that the value in (1.4) can be represented via a duality argument as

$$\sup\Big\{N \int_X \varphi \, \mathrm{d}\rho : \varphi(x_1) + \dots + \varphi(x_N) \leq c(x)\Big\}; \tag{1.5}$$

the proof is carried out for the Coulomb cost, but adapts to the assumption that c is lower continuous; moreover, the existence of an optimal potential φ in the dual formulation (1.5) is also proved (see also [16] for a generalization to costs not necessarily bounded from below). We remark that a general duality result has been already proven by Kellerer in [21], but the hypotesis on the cost function couldn't be adapted to a Coulomb-type cost.

Finally, we remark that in the context of multimarginal optimal transport problem several of the questions mentioned above are open even with more classical cost functions, such as the quadratic cost; recent developments can be found in [15, 26, 23, 18, 14].

In this paper we prove the finiteness and continuity of multi-marginal optimal transport with repulsive cost under the assumption that the measure does not concentrate too much. This issue was already partly addressed in [5], where the authors present sufficient conditions for the continuity of the cost and where they analyze the Kantorovich potentials. We improve the previous results especially from the point of view of the assumptions, which in our work are shown to be sharp. The continuity of the optimal cost under the sharp conditions of the present paper is useful for instance in [4] to study the bond dissociation problem of certain molecules in density functional theory.

Our results depend on assumptions regarding the concentration of mass of the marginal ρ, therefore we introduce two quantities measuring it. Given $\mu \in \mathscr{P}(X)$, we consider the pointwise concentration of μ, namely its biggest atom

$$a(\mu) := \max_{x \in X} \mu(\{x\}), \tag{1.6}$$

and the concentration on balls defined as

$$\kappa(\mu, r) := \sup_{x \in X} \mu\big(\bar{B}(x, r)\big),$$

which gives a more quantitative information, since $\lim_{r \to 0} \kappa(\mu, r) = a(\mu)$. In terms of the first concentration property, we characterize the finiteness of the cost.

Theorem 1.1 (Finiteness of the cost)**.** *Let $\rho \in \mathscr{P}(X)$, let $\mathscr{C}$ be the cost introduced in (1.4) under the assumptions (1.1), (1.2) and (1.3), let $a(\rho)$ as in (1.6). Then we have that*

(i) if $a(\rho) < \frac{1}{N}$ then $\mathscr{C}(\rho) < \infty$;

(ii) if $a(\rho) > \frac{1}{N}$ then $\mathscr{C}(\rho) = \infty$;

(iii) if $a(\rho) = \frac{1}{N}$ then $\mathscr{C}(\rho) < \infty$ if and only if

$$\int_{X \setminus \{\bar{x}\}} c(\bar{x}, x) \, \mathrm{d}\rho(x) < \infty \qquad \forall \bar{x} \, : \, \rho(\{\bar{x}\}) = \frac{1}{N}. \tag{1.7}$$

We acknowledge here also the recent paper [2] in which the first two parts of the theorem above are proved by a slicing method and by induction on the number of atoms. Our proof is considerably shorter and it is based instead on a dimension reduction argument, which allows to reduce to a 1 dimensional problem via a suitably chosen projection.

Next, we address in our main theorem the problem of the continuity of the cost, in which we will also use the "enlarged diagonal" for any $\alpha > 0$

$$D_\alpha = \big\{x = (x_1, \dots, x_N) \in X^N : \mathsf{d}(x_i, x_j) < \alpha \text{ for some } i \neq j\big\}.$$

Theorem 1.2. *Let $\mathscr{C}$ be the cost introduced in (1.4) under the assumptions (1.1), (1.2) and (1.3). Let $r > 0$, $\delta \in (0, 1/N)$, and consider the set*

$$\mathscr{K}_{r,\delta} := \{\rho \in \mathscr{P}(X) : \kappa(\rho, r) < \delta\}. \tag{1.8}$$

Then we have the following.

(i) If the cost c is lower semicontinuous, $\mathscr{C}$ is Lipschitz in $\mathscr{K}_{r,\delta}$ with respect to the strong topology on $\mathscr{P}(X)$.

(ii) If, in addition, the cost c is continuous, $\mathscr{C}$ is continuous in $\{\rho \in \mathscr{P}(X) : a(\rho) < 1/N\}$ with respect to the weak topology, or equivalently with respect to the Wasserstein distance W_1.

(iii) If the cost c is Lipschitz outside D_α for every $\alpha > 0$, then $\mathscr{C}$ is Lipschitz in $\mathscr{K}_{r,\delta}$ with respect to the Wasserstein W_1 distance on $\mathscr{P}(X)$.

A consequence of Theorem 1.2(i) is that $\mathscr{C}$ is locally Lipschitz on $\{\rho \in \mathscr{P}(X) : a(\rho) < 1/N\}$ with respect to the strong topology on $\mathscr{P}(X)$. Similarly, under the assumptions on the cost in Theorem 1.2(iii), $\mathscr{C}$ is locally Lipschitz in $\{\rho \in \mathscr{P}(X) : a(\rho) < 1/N\}$ with respect to the Wasserstein distance. On the other hand, it is *not* true that the cost is continuous, even with the strong topology, on the set $\{\rho \in \mathscr{P}(X) : \mathscr{C}(\rho) < \infty\}$, even in the case of the Coulomb cost in $X = \mathbb{R}^d$. Indeed, it is clear that if $a(\rho) = 1/N$ for some ρ of finite cost, we can enlarge the Dirac delta of size $1/N$ in ρ (reducing the mass slightly elsewhere) to obtain a sequence of measures of infinite cost which converge strongly to ρ. On the other side, with a little bit more work we can approximate any ρ of this type with measures of finite energy creating a discontinuity in ρ, see Proposition 3.5 below.

The result in Theorem 1.2 is based on two key ideas. First, in Theorem 1.3 we provide quantitative bounds regarding the distance of the support from the diagonals $x_i = x_j$, $i \neq j$ (where the cost is infinite). Once this result is established, the minimization problem (1.4) becomes fully equivalent to the same problem *with a cost truncated from above* (see Lemma 5.1). Under the boundedness assumption for the cost, we can refer to more classical optimal transport results to obtain the existence of optimal potentials as well as their estimates. Since these results might have an interest which goes beyond the proof of Theorem 1.2, we describe them in the next two subsections.

1.1 Diagonal bounds

For $\alpha > 0$ define the "enlarged diagonals"

$$\begin{aligned} D_\alpha &= \big\{x = (x_1, \dots, x_N) \in X^N : \mathsf{d}(x_i, x_j) < \alpha \text{ for some } i \neq j\big\}, \\ \bar{D}_\alpha &= \big\{x = (x_1, \dots, x_N) \in X^N : \mathsf{d}(x_i, x_j) \leq \alpha \text{ for some } i \neq j\big\}. \end{aligned}$$

Notice that in general $\bar{D}_\alpha$ is not the closure of D_α (which would be denoted by $\overline{D_\alpha}$ if needed), but rather contains it. We also introduce an enlarged diagonal which is more intrinsic in terms of the cost, rather than of the distance

$$D^h = \big\{x = (x_1, \dots, x_N) \in X^N : c(x_i, x_j) > h \text{ for some } i \neq j\big\}.$$

The set D^h is a more tailored version of D_α but of course we have $D^{M(\alpha)} \subseteq D_\alpha \subseteq D^{m(\alpha)}$ and for example in the case $c(x,y) = f(d(x,y))$ they coincide up to a composition with f or its inverse.

We can provide some diagonal bounds for the optimal plan, improving the corresponding result in [5, Theorem 2.4]. Notice that in this Theorem we don't require the *strong repulsion* assumption (1.2); this is important since in the crucial Lemma 5.1 we in fact apply it to a bounded cost.

Theorem 1.3 (Diagonal Bounds)**.** *Let $\rho \in \mathscr{P}(X)$, $r > 0$; let us consider an optimal plan $\pi \in \Pi(\rho)$ in (1.4) under the assumptions (1.1) and (1.3), let $\kappa(\rho, r)$ as in (1.6). Then we have*

(i) if $\kappa(\rho,r) < \frac{1}{2(N-1)}$ and $h > 2(N-1)M(r)$ then $\pi(D^h) = 0$.

(ii) if $\kappa(\rho,r) < \frac{1}{N}$ whenever we have $h, \beta > 0$ satisfying

$$h > 2(N-1)M(\beta/2), \qquad m(\beta) > \frac{\mathscr{C}(\rho)}{1 - N\kappa(\rho,r)}, \qquad \beta/2 \le r \tag{1.9}$$

we have $\pi(D^h) = 0$. Moreover under assumption 1.2 there always exist such $h, \beta > 0$, and they can be also chosen depending only on r, N and $\delta := 1 - N\kappa(\rho, r)$.

In particular in both cases we have $\pi(D_\alpha) = 0$ whenever $m(\alpha) \ge h$.

Notice that while in (ii) the assumption on the measure is sharp, the behavior of the estimate of h with respect to N and r is not very nice. Indeed it depends on $\mathscr{C}(\rho)$, which can be avoided by using Theorem 4.1, at the cost of losing a factor $\binom{N}{2}$. This is why we kept also estimate (i) which is sharper in the behavior on N and r despite being not optimal on the assumption.

1.2 Kantorovich potentials

We recall here the existence of optimal potentials, which is the main result of [13]. In the paper the proof is written for the Coulomb cost and for probabilities ρ with no atoms. Using some ideas present in [5], we provide a sharper version, that works for every ρ such that $a(\rho) < 1/N$.

Theorem 1.4. *Let $\mathscr{C}$ be the cost introduced in (1.4) under the assumptions (1.1), (1.2) and (1.3), $\rho \in \mathscr{P}(X)$ with $a(\rho) < 1/N$. Then the duality formula holds*

$$\mathscr{C}(\rho) = \sup \Big\{ N \int_X \varphi \, \mathrm{d}\rho : \varphi(x_1) + \cdots + \varphi(x_N) \le c(x) \Big\}$$

and the supremum in the right-hand side is realized by a potential $\varphi \in L^1 \cap L^\infty(\rho)$.

Proof. We deduce this theorem since by Lemma 2.1 we have $\rho \in \mathscr{K}_{r,\delta}$ for some $r > 0$ and $\delta < 1/N$; then we use Lemma 5.1, which in turn uses the existence of optimal potentials for bounded costs, proven in [21]. ☐

To prove the continuity of the cost in Theorem 1.2, we need to obtain uniform estimates on these potentials when ρ varies in a set $\mathscr{K}_{r,\delta}$. Notice that in the following theorem we do not assume the strong repulsion assumption (1.2).

Proposition 1.5 (L^∞ and Lip bounds on the Kantorovich potential)**.** *Let $\mathscr{C}$ be the cost introduced in (1.4) under the assumptions (1.1) and (1.3). Let $r > 0$, $\delta \in (0, 1/N)$, and consider the set $\mathscr{K}_{r,\delta}$ introduced in (1.8).*

Then we have that there exists a function $h := h(r, \delta)$ (given for example by (5.3)) such that

(i) if the cost c is lower semicontinuous, then there exist Kantorovich potentials φ_ρ, which are uniformly bounded in $\mathscr{K}_{r,\delta}$

$$\sup_{\rho \in \mathscr{K}_{r,\delta}} \|\varphi_\rho\|_{L^\infty(X)} < \binom{N}{2} \cdot h(r, \delta)$$

(ii) if the cost c is Lipschitz outside D_α for every $\alpha > 0$, then there exist Kantorovich potentials φ_ρ which are uniformly Lipschitz in $\mathscr{K}_{r,\delta}$

$$\sup_{\rho \in \mathscr{K}_{r,\delta}} \|\varphi_\rho\|_{\mathrm{Lip}(X)} < (N-1) \cdot \|c^{h(r,\delta)}\|_{\mathrm{Lip}(X)}$$

where c^h is defined as in (5.1).

1.3 Examples

We summarize here three particular examples that fall inside this setting:

- Coulomb in $\mathbb{R}^d$,
- $c = \phi \circ \mathsf{d}$,
- $c = G$ Green function of Δ on a manifold.

Coulomb in $\mathbb{R}^d$. The model case is the Coulomb interaction in $\mathbb{R}^3$. This is how the problem originated in the context of Density Functional Theory. The ambient space is $\mathbb{R}^d$ and the cost $c(x, y) = 1/|x - y|$. In this case we have $m(t) = M(t) = \frac{1}{t}$. Since this is maybe the most interesting example, we provide specific and quantitative estimate for every theorem in section 6.

Case $c = \phi \circ \mathsf{d}$. A specific instance of this kind would be a cost of the form $c(x_1, x_2) = \phi\big(\mathsf{d}(x_1, x_2)\big)$, where $\phi : [0, \infty) \to [0, \infty]$ is a lower semi-continuous function such that

- $\phi(0) = \infty$, hence $\lim_{r \to 0^+} \phi(r) = \infty$,
- and $\phi|_{[r,\infty)}$ is bounded for every $r > 0$.

In this case, m and M could be given by

$$m(r) \coloneqq \min_{r' \in [0,r]} \phi(r'), \qquad M(r) \coloneqq \sup_{r' \in [r,\infty)} \phi(r').$$

From the definition follows that m and M are non-increasing and right-continuous, $m(r) \leq \phi(r) \leq M(r)$ and $\lim_{r \to 0^+} m(r) = \infty$. We define also the pseudo-inverse $m^{-1} : [0, \infty) \to (0, \infty]$ by

$$m^{-1}(t) \coloneqq \max\{r \in (0, \infty] : m(r) \geq t\}.$$

Then m^{-1} is non-increasing, left-continuous and satisfies the important relation $m\big(m^{-1}(t)\big) \geq t$.

Green function of Δ. Noticing that the potential $1/|x-y|$ is the fundamental solution of the Laplacian in $\mathbb{R}^3$, the first case can be generalized to a Riemannian manifold M where the cost is given by $c(x,y)=G(x,y)$, the fundamental solution of $\Delta_x G(x,y)=\delta_y$. If the manifold is compact then it is clear that c satisfies the previous hypotheses, but they could be verified also on some non-compact manifolds, like they are in $\mathbb{R}^d$ because of the translation invariance.

2 Preliminary results

For all the functions introduced so far, sometimes we will drop the N dependence, whenever it will be clear by the context.

We will use the notation $P^i : X^N \to X$ to denote the projection on the i-th coordinate and also $P^{i_1,\dots,i_k} : X^N \to X^k$ to denote the projection on the coordinates $i_1,\dots,i_k$. Moreover, given $\pi \in \mathscr{P}(X^N)$ we denote by $P_{sym}(\pi) = \frac{1}{N!}\sum_{\sigma\in S_N}\sigma_\#\pi$ and notice that $C(\pi)=C(P_{sym}(\pi))$ thanks to the symmetry of the cost.

2.1 Properties of the concentration

Clearly the uniform concentration condition measured by κ is stronger than the pointwise one encoded by a. However, thanks to a compactness argument, the next lemma shows that the two are in fact almost equivalent.

Lemma 2.1. *Let $\rho \in \mathscr{P}(X)$ and assume that $a(\rho) < \delta$. Then there exists $r > 0$ such that $\kappa(\rho, r) < \delta$.*

Proof. Fix δ' such that $a(\rho) < \delta' < \delta$. Since ρ is tight, we can find a compact subset $K \subset X$ such that $\rho(K^c) < \delta'$. Given $x \in X$, one has $\lim_{r\to 0^+}\rho\big(\bar{B}(x,r)\big) = \rho(\{x\}) \le a(\rho) < \delta'$, therefore for every x there exists a positive radius r_x such that

$$\rho\big(\bar{B}(x,3r_x)\big) < \delta'.$$

Since K is compact, we can find a finite number of points $x_1,\dots,x_k$ such that $K \subset \bigcup_{i=1}^k \bar{B}(x_i, r_{x_i})$. Let $r=\min\{r_{x_1},\dots,r_{x_k}\}$. If $\mathsf{d}(x,K) > r$, then $\bar{B}(x,r)\subset K^c$, hence $\rho\big(\bar{B}(x,r)\big) < \delta'$. If $\mathsf{d}(x,K)\le r$, then $\mathsf{d}(x,x_i) \le r + r_{x_i} \le 2r_{x_i}$ for some $i=1,\dots,k$, therefore $\bar{B}(x,r)\subset \bar{B}(x_i,3r_{x_i})$, hence $\rho\big(\bar{B}(x,r)\big) < \delta'$. This implies $\kappa(\rho,r)\le\delta'<\delta$. ☐

Lemma 2.2. *Assume that $\rho,\eta \in \mathscr{P}(X)$. Then for every $r, r' > 0$ we have*

$$(r-r')\cdot(\kappa(\eta,r')-\kappa(\rho,r)) \le W_1(\rho,\eta).$$

Proof. By symmetry we can assume $r > r'$. We can also assume $\kappa(\eta,r') > \kappa(\rho,r)$ otherwise the inequality would be trivial. Let us now take $x \in X$ such that $\eta(B(x,r')) \ge \kappa(\eta,r')-\varepsilon$. Then we can consider an optimal plan γ between η and ρ, and let $A = B(x,r')\times(X\setminus B(x,r))$; we have that

$$\begin{aligned}\gamma(A) &= \gamma(B(x,r')\times X) - \gamma(B(x,\tilde{r})\times B(x,r))\\ &\ge \gamma(B(x,r')\times X) - \gamma(X\times B(x,r))\\ &= \eta(B(x,r')) - \rho(B(x,r)) \ge \kappa(\eta,r')-\varepsilon-\kappa(\rho,r).\end{aligned}$$

In particular for ε sufficiently small we have $\gamma(A) > 0$ and so we can compute

$$W_1(\rho, \eta) \geq \int_A d(x, y) \geq (r - r')\gamma(A) \geq (r - r') \cdot (\kappa(\eta, r') - \varepsilon - \kappa(\rho, r)).$$

and we can conclude by the arbitrariness of ε. □

Lemma 2.3. *Assume that $\rho \in \mathscr{P}(X)$ satisfies $\kappa(\rho, r) < \delta$ for some $r > 0$ and let $\rho_n \rightharpoonup \rho$. Then for every $r' \in (0, r)$ one has $\kappa(\rho_n, r') < \delta$ for n large enough.*

In particular, if $a(\rho) < \delta$, then $a(\rho_n) < \delta$ definitely in n.

Proof. We can assume that the distance is bounded, considering the modified distance $d_M(x, y) = \min\{M, d(x, y)\}$, for M big enough. If the distance is bounded we have that $\rho_n \to \rho$ if and only if $W_1(\rho_n, \rho) \to 0$. But then for $r' < r$ we can apply Lemma 2.2 in order to get

$$\limsup_{n\to\infty} \kappa(\rho_n, r') \leq \limsup_{n\to\infty} \left\{ \kappa(\rho, r) + \frac{W_1(\rho_n, \rho)}{r - r'} \right\} = \kappa(\rho, r) < \delta. \qquad \square$$

Proposition 2.4 (Good projection)**.** *Let $\rho \in \mathscr{P}(X)$ with $a(\rho) < \delta$. Then there exists $P \in \mathrm{Lip}_1(X)$ such that $a(P_{\#}\rho) < \delta$. Such a P will be called a good projection.*

Proof. We start from the case where X is a finite-dimensional normed vector space, i.e. $X \simeq \mathbb{R}^d$. It is sufficient to show that there exists $P_d \in \mathrm{Lip}(\mathbb{R}^d; \mathbb{R}^{d-1})$ such that $a(P_{d\#}\rho) < \delta$. Then we conclude by taking $P = P_2 \circ \cdots \circ P_d$. The statement is true if we are able to find a direction $v \in \mathbb{R}^d$ such that $\rho(l) < \delta$ for every line l parallel to v. In fact, then we can write $\mathbb{R}^d \simeq \mathbb{R}^{d-1} \oplus \langle v \rangle$ and take P_d to be the projection onto the first factor. Fix a positive $\varepsilon < [\delta - a(\rho)]/2$. Let $\{x_i\}_i$ be the at most countable set of atoms of ρ. Take out a finite number of them, $x_1, \ldots, x_n$, such that the mass of the remaining ones is small, namely

$$\sum_{i>n} \rho(\{x_i\}) < \varepsilon.$$

The directions $v_{ij} = x_i - x_j$ are forbidden. Consider the non-atomic measure

$$\tilde{\rho} = \rho - \sum_{i\geq 1} \rho(\{x_i\})\delta_{x_i}.$$

This measure is additive on finite unions of distinct lines, because the intersections are finite sets of points, which have zero measure w.r.t. $\tilde{\rho}$. Therefore there is only a finite number of lines $l_1, \ldots, l_k$ with $\tilde{\rho}(l_i) \geq \varepsilon$. Let v_i denote a direction parallel to l_i. This procedure rules out another finite number of directions, $v_1, \ldots, v_k$. Now take a direction v which is not parallel to any of the v_{ij} or v_i. If l is a line parallel to v, l can contain at most one of the points $x_1, \ldots, x_n$ (otherwise v would be parallel to some v_{ij}) and $\tilde{\rho}(l) < \varepsilon$ (otherwise v would be parallel to some v_i). Therefore

$$\rho(l) \leq \tilde{\rho}(l) + \max_{i=1,\ldots,n} \rho(\{x_i\}) + \sum_{i>n} \rho(\{x_i\}) < \varepsilon + a(\rho) + \varepsilon < \delta.$$

Assume now that $X = \ell^\infty$ and $\rho \in \mathscr{P}(\ell^\infty)$ is tight. It is well know (see for instance [25, Lemma 5.7]) that ℓ^∞ has the metric approximation property, that is, for every compact set $K \subset \ell^\infty$ and every $\varepsilon > 0$ there is a linear operator $T : \ell^\infty \to \ell^\infty$ of finite rank with operator norm $\|T\| \leq 1$ and $\sup_{x\in K}\|Tx - x\|_\infty \leq \varepsilon$. Since ρ is tight, there are increasing compact sets K_n such that $\rho(K_n^c) < 1/n$. ρ is clearly concentrated on the set $H = \bigcup_n K_n$. Let $T_n : \ell^\infty \to \ell^\infty$ be a finite-rank linear operator with $\|T_n\| \leq 1$ and $\sup_{x\in K_n}\|T_n x - x\|_\infty \leq 1/n$. For every $x \in H$ we have $T_n x \to x$ as $n \to \infty$, therefore $T_{n\#}\rho \rightharpoonup \rho$.[1] But then, by Lemma 2.3, $a(T_{n\#}\rho) < \delta$ for n sufficiently large. The measure $T_{n\#}\rho$ is supported on a finite-dimensional vector subspace of ℓ^∞ (the image of T_n), therefore we already know that there is a good projection Q for it. A good projection for ρ itself is then given by $P = Q \circ T_n$.

In the general case of a Polish space (X, d), we simply need to embed it isometrically $\iota : X \to \ell^\infty$ by means of $\iota(x) = (\varphi_n(x))_n$, where $\varphi_n(x) = \mathsf{d}(x, x_n) - \mathsf{d}(x, x_0)$ and $\{x_n\}_n \subset X$ is a countable dense set. By Ulam lemma ρ is tight, and so is $\iota_\#\rho \in \mathscr{P}(\ell^\infty)$. Clearly $a(\iota_\#\rho) = a(\rho) < \delta$, therefore we can find a good projection Q for $\iota_\#\rho$ and a good projection for ρ is given by $P = Q \circ \iota$. □

Remark 2.5. The previous proposition remains true when ρ is a tight finite non-negative measure on a generic metric space X. The only modification is to observe that we just need to embed only $\operatorname{supp}(\rho) \hookrightarrow \ell^\infty$, which is σ-compact and closed, thus Polish.

Proposition 2.4 will be used to prove the finiteness of the cost under the assumption that $a(\rho) < 1/N$. To deal with the other concentration condition $\kappa(\rho, r) < 1/N$, one could hope to extend the good projection in the following way. However we have not been able to establish the truth of the next conjecture, therefore we had to find another way to get the bound of the cost (see Theorem 4.1). The conjecture, however, seems interesting enough from the measure teoretic perspective, so we state it anyway.

Conjecture 2.6 (Good projection, quantitative version). *Let $\rho \in \mathscr{P}(\mathbb{R}^d)$ with $\kappa(\rho, r) < \delta$. Then for every $\varepsilon > 0$ there exists $P \in \mathrm{Lip}_1(\mathbb{R}^d)$ such that $\kappa(P_\#\rho, r') < \delta + \varepsilon$ for some $r'(r, d, \delta, \varepsilon) > 0$.*

3 Characterization of finiteness of the cost

Lemma 3.1 (Monotone plan). *Let $\rho \in \mathscr{P}(\mathbb{R})$. Then there exists $\pi_\rho \in \Pi_N(\rho)$ such that for π_ρ-a.e. $x \in \mathbb{R}^N$ and for every $i \neq j$ we have $\rho([x_i, x_j]) \geq \frac{1}{N}$. In particular if $\kappa(\rho, r) < \frac{1}{N}$ then $\pi_\rho(D_{2r}) = 0$.*

Remark 3.2. In the following proof we will consider π_ρ as the unique symmetric *monotone* plan (which is a plan with the property that every x, y in its support are *well ordered*, as defined in [7]). Since we are only interested in its final properties we will not prove that it is monotone (even if it is obvious) nor we will discuss its uniqueness: a proof of these properties in the case ρ atomless can be found in [7].

Proof. Let us consider $F : [0, 1] \to \mathbb{R}$ defined as $F(t) = \sup\{x \ : \ \rho((-\infty, x]) < t\}$. Then let us define

$$\pi_\rho = N \cdot P_{sym}\Big(\big(F(t), F(t + \tfrac{1}{N}), \ldots, F(t + \tfrac{N-1}{N})\big)_\# \mathcal{L}|_{[0, \frac{1}{N}]}\Big).$$

---

[1] Indeed, if $f \in C_b(\ell^\infty)$, one has $\int f \, \mathrm{d}T_{n\#}\rho = \int f \circ T_n \, \mathrm{d}\rho \to \int f \, \mathrm{d}\rho$ by dominated convergence.

We claim that π_ρ is a plan with the properties we want.

It is clear to see that F is a pseudo-inverse of the cumulative distribution function of ρ. In particular we have $F_\# \mathcal{L}_{[0,1]} = \rho$, and so we obtain that $\pi_\rho \in \Pi_N(\rho)$. Moreover, from the definition of F we deduce

$$\rho\big((-\infty, F(t)\big) \leq t \leq \rho\big((-\infty, F(t)]\big).$$

In particular we get immediately $\rho\big([F(t+\frac{i}{N}), F(t+\frac{j}{N})]\big) \geq \frac{|i-j|}{N}$, which implies the wanted property on π_ρ.

Now, if we add the hypothesis that $\kappa(\rho, r) < \frac{1}{N}$, we have that that $\rho([x_i, x_j]) \geq \frac{1}{N}$ imples $|x_i - x_j| \geq 2r$. In fact if it was not the case then

$$\frac{1}{N} \leq \rho([x_i, x_j]) \leq \rho\Big(B\big(\tfrac{x_i+x_j}{2}, r\big)\Big) \leq \kappa(\rho, r) < \frac{1}{N}.$$

But then $D_{2r} \subseteq \bigcup_{i \neq j} \{x \ : \ \rho([x_i, x_j]) < \frac{1}{N}\}$, and since we know that the second set is π_ρ-null we have also that $\pi_\rho(D_{2r}) = 0$. ☐

We present here a simple proof of the finiteness of the cost depending on the existence of good projections, before moving on to the more powerful, but maybe less intuitive, Theorem 4.1. This result appears also in the recent preprint [2, Theorem 1.1], where it is proved in a longer way using at the core a dimension reduction argument similar to our good projection, but working always in the original ambient space and therefore not fully exploiting the simpler structure of the one-dimensional problem.

Proof of Theorem 1.1(i). We claim that, given $\rho \in \mathscr{P}(X)$ be such that $a(\rho) < 1/N$, there exists a plan $\pi \in \Pi(\rho)$ such that $\pi(D_\alpha) = 0$ for some $\alpha > 0$. In particular, the statement follows since

$$\mathscr{C}(\rho) \leq C(\pi) \leq \binom{N}{2} M(\alpha) < \infty.$$

To show the claim, take a good projection $P \in \mathrm{Lip}_1(X)$ given by Proposition 2.4 and consider the measure $\nu = P_\# \rho$; in particular we have $a(\nu) < \frac{1}{N}$ and, thanks to Lemma 2.1, also that $\kappa(\nu, r) < \frac{1}{N}$ for some $r > 0$. By the disintegration theorem there are probabilities $\rho_t \in \mathscr{P}(X)$ such that $\rho = \rho_t \otimes \nu(t)$.

Let $\tilde{\pi} \in \Pi(\nu)$ be a plan given by Lemma 3.1 and let $\pi \in \Pi(\mu)$ be any plan such that $(P, \dots, P)_\# \pi = \tilde{\pi}$. Such a plan can be build by mapping arbitrarily the measures ρ_t on one another. In particular we will have that for every $(x_1, \dots, x_N) \in \mathrm{supp}(\pi)$ we have $(P(x_1), \dots, P(x_n)) \in \mathrm{supp}(\tilde{\pi})$ and so we get that $\pi(D_\alpha) = 0$ as long as $\tilde{\pi}(D_\alpha) = 0$, thanks to the fact that P is 1-Lipschitz.

Since we have $\kappa(\nu, r) < \frac{1}{N}$, Lemma 3.1 gives that $\tilde{\pi}(D_{2r}) = 0$ and so we can conclude $\pi(D_{2r}) = 0$. ☐

Proof of Theorem 1.1(ii). We prove that every $\rho \in \mathscr{P}(X)$ such that $\mathscr{C}(\rho) < \infty$ satisfies $a(\rho) \leq 1/N$. Let $\pi \in \Pi(\rho)$ be an optimal plan. Since $\mathscr{C}(\rho) = C(\pi) < \infty$, we infer that $\pi(D) = 0$. Let $\bar{x} \in \arg\max\{\rho(\{x\}) : x \in X\}$, so that $\rho(\{\bar{x}\}) = a(\rho)$, and define $X_* = \{\bar{x}\}^c$. For every $i = 1, \dots, N$ one has

$$\rho(\{\bar{x}\}) = P^i_\# \pi(\{\bar{x}\}) = \pi\big(X^{N-1} \times_i \{\bar{x}\}\big) = \pi\big(X_*^{N-1} \times_i \{\bar{x}\}\big).$$

Notice that the N sets $X_*^{N-1} \times_i \{\bar{x}\}$ are disjoint, therefore, adding over $i = 1, \ldots, N$, we get

$$N\rho(\{\bar{x}\}) = \sum_{i=1}^{N} \pi\big(X_*^{N-1} \times_i \{\bar{x}\}\big) = \pi\left(\bigcup_{i=1}^{N} X_*^{N-1} \times_i \{\bar{x}\}\right) \le \pi\big(X^N\big) = 1,$$

from which $a(\rho) \le 1/N$. □

Proposition 3.3. *Let $\mathscr{C}$ be the cost introduced in (1.4) under the assumptions (1.1), (1.2) and (1.3), let $a(\rho)$ as in (1.6). Let $\rho \in \mathscr{P}(X)$ and $\bar{x} \in X$ such that $\rho(\{\bar{x}\}) = \frac{1}{N}$. Then, letting $X^* = X \setminus \{\bar{x}\}$ and $\tilde{\rho} = \frac{N}{N-1}\rho|_{X_*}$, we have*

$$\mathscr{C}_N(\rho) = N \int_{X_*} c(\bar{x}, y)\,\mathrm{d}\rho(y) + \mathscr{C}_{N-1}(\tilde{\rho}). \tag{3.1}$$

Proof. Let $\pi \in \Pi_{N-1}(\tilde{\rho})$: then we have that $P_{sym}(\delta_{\bar{x}} \otimes \pi) \in \Pi_N(\rho)$ and so

$$\mathscr{C}_N(\rho) \le C_N(P_{sym}(\delta_{\bar{x}} \otimes \pi)) = C_N(\delta_{\bar{x}} \otimes \pi) = N \int_{X_*} c(\bar{x}, y)\,\mathrm{d}\rho(y) + C_{N-1}(\pi);$$

talking the infimum in π we obtain the first inequality.

In order to prove the other inequality, we can assume $\mathscr{C}_N(\rho) < \infty$. Let us consider π is a symmetric optimal coupling for $\mathscr{C}_N(\rho)$, which is therefore concentrated outside the diagonals, and let us define $X_i = X_*^{N-1} \times_i \{\bar{x}\}$; we know that X_i are disjoint and $\pi(X_i) = P^i_\# \pi(\{\bar{x}\}) = \frac{1}{N}$. This means that π is concentrated on $\bigcup X_i$.

We can define π_1 through the implicit equality

$$\pi|_{\{\bar{x}\} \times X^{N-1}} = \pi|_{X_1} = \frac{1}{N}\delta_{\bar{x}} \times \pi_1$$

with $\pi_1(X_*^{N-1}) = \pi_1(X^{N-1}) = 1$ and, thanks to the symmetry of π, by considering only permutations which fix the first coordinate, we deduce that also π_1 is symmetric in its $N-1$ variables. A simple computation then shows that

$$\pi = P_{sym}(\delta_{\bar{x}} \otimes \pi_1) \tag{3.2}$$

and that

$$P^i_\#(\pi_1) = \tilde{\rho} = \frac{N}{N-1}\rho|_{X_*} \qquad \text{for every } i = 1, \ldots, N-1. \tag{3.3}$$

Indeed, for every permutation which fixes x_1, we know that the measure is unchanged. On the other side, every permutation of coordinates can be written as the composition of one of these permutations with a $p^i : X^N \to X^N$ which exchanges x_1 and x_i for some $i = 2, \ldots, N$ and leaves all other coordinates fixed. For every $i = 2, \ldots, N$, we know that $p^i_\#(\delta_{\bar{x}} \otimes \pi_1)$ goes to a nonnegative measure of total mass $1/N$, concentrated on X_i and hence orthogonal to $\delta_{\bar{x}} \otimes \pi_1$ (and to any other permutation p^j with $j \neq i$), which is also a submeasure of π. Hence, we conclude that (3.2) holds. Define $\tilde{\rho}$ to be any marginal of π_1, so that $\delta_{\bar{x}} \otimes \pi_1$ has all marginals equal to $\tilde{\rho}$ apart from one, which equals $\delta_{\bar{x}}$. Symmetrizing, we find that ρ, which is any marginal of π, equals $1/N\delta_{\bar{x}} + (N-1)/N\tilde{\rho}$, which proves (3.3).

Then, thanks to (3.2) and (3.3), we can rewrite the energy of π as

$$\begin{aligned}\mathscr{C}_N(\rho) = C_N(\delta_{\bar{x}} \otimes \pi_1) &= N \int_{X_*} c(\bar{x}, y)\, \mathrm{d}\rho(y) + C_{N-1}(\pi_1) \\ &\geq N \int_{X_*} c(\bar{x}, y)\, \mathrm{d}\rho(y) + \mathscr{C}_{N-1}(\tilde{\rho}). \qquad \square\end{aligned}$$

At the threshold level $1/N$ anything can happen: the cost can be finite or infinite, depending on the specific distribution of the mass.

Remark 3.4. If X is a space with at least one accumulation point, $\mathscr{C}$ and a are as in Proposition 3.3, then there exists $\rho \in \mathscr{P}(X)$ such that $a(\rho) = 1/N$ and $\operatorname{supp}(\pi) \cap D \neq \emptyset$ for every $\pi \in \Pi(\rho)$ (thus $\pi(D_\alpha) > 0$ for every $\alpha > 0$). This shows that the assumption on the concentration in Theorem 1.3 is necessary. Moreover, there is one such ρ with $\mathscr{C}(\rho) < \infty$ and one with $\mathscr{C}(\rho) = \infty$.

Indeed, let $x \in X$ be a limit point, $(x_n)_{n\in\mathbb{N}} \subset X \setminus \{x\}$ be a sequence of distinct points converging to x, and

$$\rho := \frac{1}{N}\delta_x + \frac{N-1}{N}\sum_n p_n \delta_{x_n},$$

where $(p_n)_{n\in\mathbb{N}} \in \ell^1$ with $p_n \in (0, 1/N)$ and $\sum_{n=1}^{\infty} p_n = 1$.

With the notation of the proof of Proposition 3.3, we have

$$\begin{aligned}\pi(D_\alpha) \geq \pi\Big(\{x\} \times X_*^{N-2} \times B(x,\alpha)\Big) &= \pi_1\Big(X_*^{N-2} \times B(x,\alpha)\Big) \\ &= P_\#^{n-1}\pi_1(B(x,\alpha)) = \tilde{\rho}(B(x,\alpha)) > 0.\end{aligned}$$

Finally, since $a(\tilde{\rho}) < \frac{1}{N} < \frac{1}{N-1}$, we have $\mathscr{C}_{N-1}(\sum_n p_n \delta_{x_n}) < \infty$, and so, again by Proposition 3.3, $\mathscr{C}_N(\rho)$ is finite if and only if

$$\int_{X_*} c(\bar{x}, y)\, \mathrm{d}\rho(y) = \sum_n p_n c(\bar{x}, x_n)$$

is finite; one can choose the weights $(p_n)_n$ appropriately, taking into account (1.2), in order to make the cost finite or infinite.

Proof of Theorem 1.1(iii). We argue inductively on the number i of atoms in μ of mass $1/N$. If μ has exactly one atom of mass $1/N$ at $\bar{x}$, namely, if $i = 1$, from Proposition 3.3 we know that the cost of $\mathscr{C}_N$ is finite if and only if condition (1.7) is in force at $\bar{x}$ and $\frac{N}{N-1}\rho|_{X\setminus\{\bar{x}\}}$ has finite $\mathscr{C}_{N-1}$ cost. On the other side, this second condition is always verified because $\frac{N}{N-1}\rho|_{X\setminus\{\bar{x}\}}$ doesn't have atoms of the critical mass $1/(N-1)$, hence the first part of Theorem 1.1 applies.

If we assume the statement to be true when there are i of atoms in μ of mass $1/N$ and we want to prove it for $i + 1$, we consider μ with $i + 1$ atoms of mass $1/N$, one of which at $\bar{x}$; next we apply Proposition 3.3 and we reduce to study the finiteness of the cost of $\frac{N}{N-1}\rho|_{X\setminus\{\bar{x}\}}$, which in turn has exactly i atoms of mass $1/(N-1)$. Hence, applying the inductive assumption to this measure, we conclude the proof. $\square$

For simplicity we give the following example of discontinuity of the cost for the Coulomb cost in $\mathbb{R}^d$; of course, it can be generalized to metric spaces with costs as in assumptions (1.1) and (1.2).

Proposition 3.5. *Let* $c(x,y) = |x-y|^{-1}$ *in* $\mathbb{R}^d \times \mathbb{R}^d$ *and consider the minimization problem* (1.4)*. Let* $\rho \in \mathscr{P}(\mathbb{R}^d)$ *be such that* $a(\rho) = \frac{1}{N}$*. Then for every* $\varepsilon > 0$ *small there exists* $\tilde{\rho}$ *such that* $|\rho - \tilde{\rho}|(\mathbb{R}^d) \leq \varepsilon$ *and*

$$1 + \mathscr{C}(\rho) < \mathscr{C}(\tilde{\rho}) < \infty. \tag{3.4}$$

Proof. Let $\bar{x} \in X$ and $\mu \in \mathscr{M}_+(X)$ be such that $\rho = \frac{1}{N}\delta_{\bar{x}} + \mu$. Then by Proposition 3.3 we know that

$$N \int_{\mathbb{R}^d} c(\bar{x}, y)\,\mathrm{d}\mu(y) + \mathscr{C}_{N-1}\Big(\frac{N}{N-1}\mu\Big) = \mathscr{C}_N(\rho) < \infty. \tag{3.5}$$

Let $\varepsilon < 1/N$ and $y_\varepsilon \in \mathbb{R}^d$ such that $\varepsilon < |\bar{x} - y_\varepsilon| \leq 2\varepsilon$ and $\mu(y_\varepsilon) = 0$. We consider

$$\rho_\varepsilon := \frac{1}{N}\delta_{\bar{x}} + \varepsilon\delta_{y_\varepsilon} + (1-\varepsilon)\mu \in \mathscr{P}(\mathbb{R}^d),$$

we notice that $|\rho - \rho_\varepsilon|(\mathbb{R}^d) = \varepsilon| - \delta_{y_\varepsilon} + \varepsilon\mu|(\mathbb{R}^d) \leq 2\varepsilon$, and we estimate its cost thanks to Proposition 3.3

$$\mathscr{C}_N(\rho_\varepsilon) = N \int_{\mathbb{R}^d} c(\bar{x}, y)\,\mathrm{d}(\varepsilon\delta_{y_\varepsilon} + (1-\varepsilon)\mu)(y) + \mathscr{C}_{N-1}\Big(\frac{N}{N-1}(\varepsilon\delta_{y_\varepsilon} + (1-\varepsilon)\mu)\Big)$$

The last term is finite because $a(\varepsilon\delta_{y_\varepsilon} + (1-\varepsilon)\mu) \leq \frac{1-\varepsilon}{N}$, and the first one thanks to (3.5). On the other side, we estimate the cost from below by

$$\begin{aligned}\mathscr{C}_N(\rho_\varepsilon) &\geq N\varepsilon c(\bar{x}, y_\varepsilon) + N(1-\varepsilon)\int_{\mathbb{R}^d} c(\bar{x}, y)\,\mathrm{d}\mu(y) + \mathscr{C}_{N-1}\Big(\frac{N}{N-1}(\varepsilon\delta_{y_\varepsilon} + (1-\varepsilon)\mu)\Big)\\ &\geq \frac{N}{2} + N(1-\varepsilon)\int_{\mathbb{R}^d} c(\bar{x}, y)\,\mathrm{d}\mu(y) + \mathscr{C}_{N-1}\Big(\frac{N}{N-1}(\varepsilon\delta_{y_\varepsilon} + (1-\varepsilon)\mu)\Big)\end{aligned} \tag{3.6}$$

By the lower semicontinuity of the cost $\mathscr{C}_{N-1}$, we know that the last two terms in the right-hand side are, in the limit, greater than or equal to to the quantity in (3.5), namely $\mathscr{C}_N(\rho)$. Hence, for ε small enough we obtain (3.4). ☐

4 Uniform bounds on the cost and diagonal bounds

Theorem 4.1 (Uniform bound on the cost in terms of the concentration)**.** *Let* $\rho \in \mathscr{P}(X)$ *be such that* $\kappa(\rho, r) \leq \frac{1}{N}$ *for some* $r > 0$*. Let* $r(x)$ *be such that* $\rho(B(x, r(x))) \leq 1/N$ *for every* x*. Then we have*

$$\mathscr{C}(\rho) \leq \binom{N}{2} \int M(\max\{\tfrac{r(x)}{2}, r\})\,\mathrm{d}\rho.$$

Remark 4.2. Under the assumptions of Theorem 4.1, the slightly weaker bound

$$\mathscr{C}(\rho) \leq \binom{N}{2} M(r)$$

can be achieved in a simpler way. Indeed, this is a straightforward application of Theorem 4.3 with the set $D = \{d(x,y) < r\}$, which guarantees us the existence of a plan $\pi \in \Pi_N(\rho)$, concentrated outside D_r; in particular we have $c(x_i, x_j) \leq M(r)$ for every $i \neq j$, for π-a.e. and so we have $C(\pi) \leq \binom{N}{2} M(r)$.

Proof of Theorem 4.1. Without loss of generality we can assume that $r(x)$ is the maximum radius such that $\rho(B(x, r(x)) \leq 1/N$. We first notice that $r(x)$ is 1-Lipschitz, in fact since $B(x, r(x)+\varepsilon) \subseteq B(y, r(x)+\varepsilon+d(x,y))$ by the maximality of $r(x)$ we deduce that $\rho(B(y, r(x)+\varepsilon+d(x,y))) > \frac{1}{N}$. This is true for every $\varepsilon > 0$, so we get

$$r(y) \leq r(x) + d(x,y).$$

Since we can reverse the roles of x and y, we obtain that r is 1-Lipschitz.

Let $D = \{2d(x,y) < \max\{r(x), r(y), 2r\}\}$; since both d and r are continuous we have that D is an open symmetric set. Moreover defining $B(x) = \{x' : (x,x') \in D\}$ we have

$$\begin{aligned} B(x) &= B(x,r) \cup B(x, \tfrac{r(x)}{2}) \cup \{x' : d(x,x') < \frac{r(x')}{2}\} \\ &\subseteq B(x,r) \cup B(x, \tfrac{r(x)}{2}) \cup \{x' : d(x,x') < \frac{d(x,x') + r(x)}{2}\} = B(x,r) \cup B(x, r(x)), \end{aligned}$$

where we used the fact that r is 1-Lipschitz. Clearly we thus have $\rho(B(x)) \leq \frac{1}{N}$; so we can use Theorem 4.3 in order to get a plan $\pi \in \Pi_N(\rho)$ such that $(x_i, x_j) \notin D$ for $i \neq j$ for π-a.e. $x \in X^k$. But this means that $d(x_i, x_j) \geq \max\{r(x_i), r(x_j)\}/2$ for π-a.e. x: we then get

$$\begin{aligned} \mathscr{C}(\rho) \leq C(\pi) &\leq \int_{X^N} \sum_{i<j} M(\tfrac{\max\{r(x_i), r(x_j)\}, 2r}{2}) \, \mathrm{d}\pi \\ &\leq \int_X \sum_{i<j} \frac{M(\max\{\frac{r(x_i)}{2}, r\}) + M(\max\{\frac{r(x_j)}{2}, r\})}{2} \, \mathrm{d}\pi \\ &= \binom{N}{2} \int_X M(\max\{\tfrac{r(x)}{2}, r\}) \, \mathrm{d}\rho. \end{aligned}$$

□

In graph theory, a consequence of the Hajnal-Szemerédi theorem [20, 22] is a simplified multimarginal version of the marriage theorem (a *multimarriage theorem*): let us suppose we have kN people, and everyone has a list of hated people, which has always less than k people in it (hatred is a reciprocal sentiment, at least in this example). Then we can form k disjoint N-tuples such that in every N-tuple we do not have people who hate each other.

The following theorem can be seen as the continuous analogue of this *multimarriage theorem* ($B(x)$ is the list of people disliked by x).

Theorem 4.3 (Existence of a plan outside the diagonal)**.** *Let X be a Polish space and let $D \subset X^2$ be a symmetric open set; let us denote $B(x) = \{x' : (x, x') \in D\}$. Let us suppose that $\rho(B(x)) \leq 1/N$ for every $x \in X$: then there exists a plan $\pi \in \Pi_N(\rho)$ that is concentrated on A^c, where*

$$A = \bigcup_{X^N} \bigcup_{i \neq j} \{(x_i, x_j) \in D\},$$

that is we have $(x_i, x_j) \notin D$ for $i \neq j$ for π-a.e. $x \in X^N$.

Proof. The proof exploits the duality formula for bounded costs. In order to show that there exists an admissible plan $\pi \in \Pi_N(\rho)$ such that $\pi(A) = 0$, we will analyze the minimizer of the multi-marginal optimal transport with respect to the following bounded cost:

$$\tilde{c}(x_1, \dots, x_N) = \inf\bigl\{d((x_1, \dots, x_N), A^c), 1\bigr\}$$

For this cost it is known that the duality formula holds ([13]):

$$\inf_{\pi \in \Pi(\rho)} \int_{X^N} \tilde{c} \,\mathrm{d}\pi = \sup_{\varphi(x_1) + \cdots + \varphi(x_N) \leq \tilde{c}(x)} N \int_X \varphi \,\mathrm{d}\rho.$$

The optimal $\pi \in \Pi_N(\rho)$ will satisfy $\pi(A) = 0$ if we show that

$$\int_X \varphi \,\mathrm{d}\rho \leq 0 \qquad \text{for all admissible } \varphi.$$

In fact, in such case the optimal value of the previous problems must be 0, therefore π has to be supported on A^c, thanks to the fact that A is open.

Actually, the crucial constraint on φ that will be needed for the proof is

$$\varphi(x_1) + \cdots + \varphi(x_N) \leq 0 \qquad \text{if } x \in A^c.$$

The only role that the cost $\tilde{c}$ on A plays is telling us that φ is bounded from above, in fact $N\varphi(x) \leq \tilde{c}(x, \dots, x) \leq 1$; indeed one would like to consider the cost which takes the value ∞ in this region, if there were not the problem of the validity of the duality formula for such a cost and the boundedness of the potential.

After having fixed a small $\varepsilon > 0$, we do the following iterative construction of η_i, z_i and B_i:

$$\begin{array}{lll}
\eta_1 = \sup_X \varphi, & z_1 \in X,\ \varphi(z_1) \geq \eta_1 - \varepsilon, & B_1 = B(z_1), \\
\eta_2 = \sup_{B_1^c} \varphi, & z_2 \in B_1^c,\ \varphi(z_2) \geq \eta_2 - \varepsilon, & B_2 = B(z_2), \\
\vdots & \vdots & \vdots \\
\eta_k = \sup_{(B_1 \cup \cdots \cup B_{k-1})^c} \varphi, & z_k \in (B_1 \cup \cdots \cup B_{k-1})^c,\ \varphi(z_k) \geq \eta_k - \varepsilon, & B_k = B(z_k).
\end{array}$$

Notice that we have the monotone sequence $r/N \geq \eta_1 \geq \eta_2 \geq \dots$ and so on.

At each step we check the sign of the quantity

$$\eta_1 + \cdots + \eta_{k-1} + (N - k + 1)\eta_k - (k - 1)\varepsilon.$$

As soon as it is non-positive we stop the process and estimate the quantity $\int_X \varphi \,\mathrm{d}\rho$. Notice that this will surely happen by the time we reach $k = N$, because if $z \in (B_1 \cup \dots \cup B_{N-1})^c$, then $(z_1, \dots, z_{N-1}, z) \in A^c$ (in fact, letting $z_N := z$ we have $(z_i, z_j) \notin D$ for every $i \neq j$), so

$$(\eta_1 - \varepsilon) + \dots + (\eta_{N-1} - \varepsilon) + \varphi(z) \leq \varphi(z_1) + \dots + \varphi(z_{N-1}) + \varphi(z) \leq 0$$

and $\eta_1 + \dots + \eta_N - (N-1)\varepsilon \leq 0$ follows by taking the supremum over z.

Calling k the smallest integer for which this happens, by construction we have

$$\eta_k \leq -\frac{1}{N-k+1} \sum_{j=1}^{k-1} \eta_j + \frac{k-1}{N-k+1}\varepsilon, \tag{4.1}$$

while the preceding inequalities are reversed.

Letting $\tilde{B}_j = B_j \setminus (B_1 \cup \dots \cup B_{k-1})$ so that they are disjoint, we can estimate

$$\begin{aligned}
\int_X \varphi \,\mathrm{d}\rho &= \sum_{i=1}^{k-1} \int_{\tilde{B}_i} \varphi \,\mathrm{d}\rho + \int_{(B_1 \cup \dots \cup B_{k-1})^c} \varphi \,\mathrm{d}\rho \\
&\leq \sum_{i=1}^{k-1} \eta_i \rho(\tilde{B}_i) + \eta_k \left(1 - \sum_{i=1}^{k-1} \rho(\tilde{B}_i)\right) \\
&\leq \sum_{i=1}^{k-1} \eta_i \rho(\tilde{B}_i) + \left(-\frac{1}{N-k+1} \sum_{j=1}^{k-1} \eta_j + \frac{k-1}{N-k+1}\varepsilon\right) \left(1 - \sum_{i=1}^{k-1} \rho(\tilde{B}_i)\right) \\
&= \sum_{i=1}^{k-1} \rho(\tilde{B}_i) \underbrace{\left(\eta_i + \frac{1}{N-k+1} \sum_{j=1}^{k-1} \eta_j - \frac{k-2}{N-k+1}\varepsilon\right)}_{\geq 0} \\
&\qquad\qquad - \frac{1}{N-k+1} \sum_{j=1}^{k-1} \eta_j + \frac{k-1}{N-k+1}\varepsilon - \frac{1}{N-k+1}\varepsilon \sum_{i=1}^{k-1} \rho(\tilde{B}_i), \\
&\leq \sum_{i=1}^{k-1} \frac{1}{N} \left(\eta_i + \frac{1}{N-k+1} \sum_{j=1}^{k-1} \eta_j - \frac{k-2}{N-k+1}\varepsilon\right) \\
&\qquad\qquad - \frac{1}{N-k+1} \sum_{j=1}^{k-1} \eta_j + \frac{k-1}{N-k+1}\varepsilon \\
&\leq \left(\frac{1}{N} + \frac{k-1}{N(N-k+1)} - \frac{1}{N-k+1}\right) \sum_{j=1}^{k-1} \eta_j + \frac{k-1}{N-k+1}\varepsilon \\
&= \frac{k-1}{N-k+1}\varepsilon \leq N\varepsilon,
\end{aligned}$$

where we used (4.1) in the second inequality and then

$$\begin{aligned}(N-k+1)\eta_i+\sum_{j=1}^{k-1}\eta_j-(k-2)\varepsilon \\ &\geq (N-k+1)\eta_{k-1}+\sum_{j=1}^{k-1}\eta_j-(k-2)\varepsilon \\ &= (N-(k-1)+1)\eta_{k-1}+\sum_{j=1}^{k-2}\eta_j-(k-2)\varepsilon \geq 0\end{aligned}$$

in the next step in order to substitute $\rho(B_i)\leq 1/N$. Letting $\varepsilon\to 0$ shows that $\int_X \varphi\,\mathrm{d}\rho\leq 0$ as desired. ☐

Proof of Theorem 1.3(ii). First of all, we can assume without loss of generality that the plan π is symmetric, since π^{sym} has the same cost of π and $\pi^{\mathrm{sym}}(\bar{D}_\alpha)=\pi(\bar{D}_\alpha)$ for every $\alpha\geq 0$.

Assume by contradiction that $\pi(\bar{D}_\alpha)>0$. Then there exists $x\in\operatorname{supp}(\pi)\cap D_\alpha$. We may assume without loss of generality that $|x_1-x_2|\leq\alpha$. For notational simplicity, let $\gamma=\beta/2\leq r$. We claim that there is a point

$$y\in\operatorname{supp}(\pi)\setminus\bar{D}_\beta\cap\left(\bar{B}(x_1,\gamma)^c\right)^N.$$

To prove that such a point exists, it is sufficient to show that

$$\pi\Big(\bar{D}_\beta^c\cap\left(\bar{B}(x_1,\gamma)^c\right)^N\Big)>0.$$

But this is true since we can estimate the mass of the complement as

$$\begin{aligned}\pi\Big(\Big[\bar{D}_\beta^c\cap\left(\bar{B}(x_1,\gamma)^c\right)^N\Big]^c\Big) &= \pi\Big(\bar{D}_\beta\cup\Big[\left(\bar{B}(x_1,\gamma)^c\right)^N\Big]^c\Big)\\ &\leq \pi(\bar{D}_\beta)+\pi\left(\bigcup_{i=1}^N X^{N-1}\times_i\bar{B}(x_1,\gamma)\right)\\ &\leq \frac{C(\pi)}{m(\beta)}+N\rho\big(\bar{B}(x_1,\gamma)\big)\\ &< 1-N\kappa(\rho,r)+N\kappa(\rho,\gamma)\leq 1.\end{aligned}$$

Next we prove that there exists $i\in\{1,\dots,N\}$ such that $d(y_i,x_j)>\gamma$ for every $j=1,\dots,N$. Indeed, by definition of y, the set $\bar{B}(x_1,\gamma)$ does not contain any of the points y_i; furthermore the $N-1$ sets $\bar{B}(x_2,\gamma),\bar{B}(x_3,\gamma),\dots,\bar{B}(x_N,\gamma)$ have diameter at most $2\gamma=\beta$, therefore at least one of the N points y_i does not belong to any of them; otherwise by the pigeonhole principle one of the aforementioned sets would contain two of the points y_i, which is impossible because they are pairwise spaced apart by more than β. Since we are dealing with a symmetric plan, we may assume that $d(y_1,x_j)>\gamma$ for every $j=1,\dots,N$.

Now we introduce the two points $\tilde{x}$ and $\tilde{y}$ obtained by swapping the coordinates x_1 and y_1, namely

$$\tilde{x}=(y_1,x_2,\dots,x_N),\qquad \tilde{y}=(x_1,y_2,\dots,y_N).$$

Thanks to the c-monotonicity we then have $c(x) + c(y) \leq c(\tilde{x}) + c(\tilde{y})$. In this last inequality many terms cancel out, in fact the interaction between x_i and x_j for $i, j \geq 2$ and between y_i and y_j for $i, j \geq 2$ are present on both sides. Thus the inequality is equivalent to

$$\sum_{i=2}^{N} c(x_1, x_i) + c(y_1, y_i) \leq \sum_{i=2}^{N} c(y_1, x_i) + c(x_1, y_i).$$

Now we can use $d(x_1, x_2) \leq \alpha$, $d(x_1, y_i) > \gamma$ and $d(y_1, x_i) > \gamma$ to get

$$m(\alpha) \leq \sum_{i=2}^{N} c(x_1, x_i) + c(y_1, y_i) \leq \sum_{i=2}^{N} c(y_1, x_i) + c(x_1, y_i) \leq 2(N-1)M(\gamma),$$

and so we reached a contradiction. ☐

Proof of Theorem 1.3(i). As in the proof of Theorem 1.3(ii) we can assume that π is symmetric. Assume by contradiction that $\pi(\bar{D}_\alpha) > 0$. Then there exists $x \in \operatorname{supp}(\pi) \cap D_\alpha$. We may assume without loss of generality that $d(x_1, x_2) \leq \alpha$. We claim that there is a point $y \in \operatorname{supp}(\pi)$ such that

$$y \in \Big(\bigcup_{i=2}^{N} \bar{B}(x_i, r)\Big)^c \times (\bar{B}(x_1, r)^c)^{N-1}.$$

For notational convenience, let us denote $A_1 = \bigcup_{i=2}^{N} \bar{B}(x_i, r)$ and $A = \bar{B}(x_1, r)$. To prove that such a point exists, it is sufficient to show that

$$\pi\big(A_1^c \times (A^c)^{N-1}\big) > 0.$$

But this is true since we can estimate the mass of the complement as

$$\begin{aligned}
\pi\Big(\big[A_1^c \times (A^c)^{N-1}\big]^c\Big) &\leq \pi\left(A_1 \times X^{N-1} \cup \bigcup_{i=2}^{N} A \times_i X^{N-1}\right) \\
&\leq \rho(A_1) + (N-1)\rho(A) \\
&\leq \sum_{i=2}^{N} \rho(B(x_i, r)) + (N-1)\rho(B(x_1, r)) < 1.
\end{aligned}$$

Now we introduce the two points $\tilde{x}$ and $\tilde{y}$ obtained by swapping the coordinates x_1 and y_1, namely

$$\tilde{x} = (y_1, x_2, \dots, x_N), \qquad \tilde{y} = (x_1, y_2, \dots, y_N).$$

Thanks to the c-monotonicity we then have $c(x) + c(y) \leq c(\tilde{x}) + c(\tilde{y})$. In this last inequality many terms cancel out, in fact the interaction between x_i and x_j for $i, j \geq 2$ and between y_i and y_j for $i, j \geq 2$ are present on both sides. Thus the inequality is equivalent to

$$\sum_{i=2}^{N} c(x_1, x_i) + c(y_1, y_i) \leq \sum_{i=2}^{N} c(y_1, x_i) + c(x_1, y_i).$$

Now we can use $d(x_1, x_2) \le \alpha$, $d(x_1, y_i) > r$ and $d(y_1, x_i) > r$ to get

$$m(\alpha) \le \sum_{i=2}^{N} c(x_1, x_i) + c(y_1, y_i) \le \sum_{i=2}^{N} c(y_1, x_i) + c(x_1, y_i) \le 2(N-1)M(r),$$

and so we reached a contradiction. □

5 Estimates on the potentials and continuity of the cost

Putting together the previous results, it is possible to show the continuity of the cost function $\mathscr{C}$ under a more general hypothesis than the one assumed in [5], following the same strategy; we sketch the short argument for completeness. Moreover, as Remark 3.4 tells us, the threshold $1/N$ is sharp. An important role will be played by truncated cost, which we therefore introduce: we define

$$c^h(x, y) := \min\{c(x, y), h\}. \tag{5.1}$$

Then, similarly to (1.3)-(1.4) we define $c^h(x_1, \dots, x_N) = \sum_{i<j} c^h(x_i, x_j)$ and

$$C^h(\pi) = \int_{X^N} c^h \, d\pi, \qquad \mathscr{C}^h(\rho) = \min\Big\{ C^h(\pi) : \pi \in \Pi(\rho) \Big\} \tag{5.2}$$

The following lemma uses the diagonal bounds on the truncated cost to prove that for h sufficiently big we have that the minimizing problems with $\mathscr{C}^h$ and $\mathscr{C}$ have in fact the same minimizers.

Lemma 5.1 (Equivalence with a truncated cost)**.** *Let us consider a cost c satisfying assumptions (1.1) and (1.2), and let $r > 0$ and $\delta < \frac{1}{N}$. Then there exists $h = h(r, \delta)$ such that for every ρ belonging to the set $\mathscr{K}_{r,\delta}$ introduced in (1.8) we have*

(i) π is an minimizer for $\mathscr{C}^h(\rho)$ if and only if it is a minimizer for $\mathscr{C}(\rho)$.

(ii) if φ^h is an optimal potential for $\mathscr{C}^h(\rho)$, it is also an optimal potential for $\mathscr{C}(\rho)$.

In particular we have that $\mathscr{C} = \mathscr{C}^h$ on $\mathscr{K}_{r,\delta}$.

Proof. Let us consider $\rho \in \mathscr{K}_{r,\delta}$.

(i) Let us consider a plan π which is optimal for the problem $\mathscr{C}(\rho)$ and a plan π^h which is optimal for $\mathscr{C}^h(\rho)$. First of all we prove that we can choose β in Theorem 1.3(ii) depending only on r, δ and not on ρ specifically. In fact we could consider β such that

$$m(\beta) > \frac{\binom{N}{2} M(r)}{1 - N\delta},$$

and this is sufficient in order to statisfy (1.9); in fact thanks to Remark 4.2 we have $\mathscr{C}(\rho) \le \binom{N}{2} M(r)$ and so in particular

$$m(\beta) > \frac{\binom{N}{2} M(r)}{1 - N\delta} \ge \frac{\mathscr{C}(\rho)}{1 - N\kappa(\rho, r)}.$$

Now, fix $h > h' > \max\{2(N-1)M(\beta/2), m(\beta)\}$ and define $m^h(r) = \min\{m(r), h\}$ and $M^h(r) = \min\{M(r), h\}$. We apply again Theorem 1.3(ii) with the cost c^h, $m^h(r)$ and $M(r)$, which satisfy assumption (1.1) too. Nnotice that $h > m(\beta)$ by construction and in particular we have $m(\beta) = m^h(\beta)$, so we get

$$m^h(\beta) = m(\beta) > \frac{\binom{N}{2} M(r)}{1 - N\delta} \geq \frac{\mathscr{C}^h(\rho)}{1 - N\kappa(\rho, r)}.$$

Since $h' > 2(N-1)M(\beta/2)$, we must have $\pi_h(D^{h'}_{c^h}) = 0$ for every π_h optimal plan for the problem $\mathscr{C}^h(\rho)$. However, since $h' < h$, it is clear that whenever $c^h(x,y) \leq h'$ we have $c^h(x,y) = c(x,y)$. But then we have $c = c^h$ on the support of π_h and so

$$\int c^h \, \mathrm{d}\pi_h = \int c \, \mathrm{d}\pi_h \geq \int c \, \mathrm{d}\pi \geq \int c^h \, \pi \geq \int c^h \, \mathrm{d}\pi_h,$$

where all the other inequalities are true for the optimality, or from $c \geq c^h$. Since the first term and the last term are equal we deduce that they are all equal and in particular π is a minimizer also for $\mathscr{C}^h$ and π_h is a minimizer also for $\mathscr{C}$, concluding the proof.

(ii) First of all if φ^h is an optimal potential for $\mathscr{C}^h(\rho)$, then of course is admissible also for $c \geq c^h$. Moreover, by (i), we have $\mathscr{C}(\rho) = \mathscr{C}^h(\rho) = \int \varphi^h \, \mathrm{d}\rho$, proving also the maximality of φ^h. □

We state explicitely what we can choose for $h(r, \delta)$:

$$h(r,\delta) > 2(N-1) \cdot M\left(\frac{1}{2} m^{-1}\Big(\frac{\binom{N}{2} M(r)}{1 - N\delta}\Big)\right). \tag{5.3}$$

This expression is very complicated since it is also in term of m, M; however if, for example, we have $M(r) \leq \alpha m(r)$ for some $\alpha > 0$, we can choose a more explicit form for $h(r, \delta)$:

$$h(r,\delta) > \frac{\alpha N (N-1)^2}{1 - N\delta} \cdot M(r).$$

In the following proposition we want to recall typical regularity results that the potentials can inherit from the cost.

Theorem 5.2 (Regularity of the potential)**.** *Assume that $c : X \times X \to [0, \infty)$ is a lower semicontinuous bounded cost. Then for every $\rho \in \mathscr{P}(X)$ there exists an optimal potential φ such that*

- *φ is bounded and*

$$-\frac{(N-1)^2}{2} \|c\|_\infty \leq \varphi(x) \leq \frac{N-1}{2} \|c\|_\infty$$

- *if moreover c is Lipschitz, φ is Lipschitz and*

$$\|\varphi\|_{\mathrm{Lip}} \leq (N-1) \|c\|_{\mathrm{Lip}}$$

- *if X is geodesic and c is K-concave then φ can be assumed to be $(N-1)K$-concave*

Proof. Let $\varphi_1, \dots, \varphi_N$ be admissible potentials. We will now construct new potentials $\tilde{\varphi}_1, \dots, \tilde{\varphi}_N$ which will be admissible, satisfy the regularity assumption and moreover $\varphi_i \leq \tilde{\varphi}_i$.

First of all we have that $\sup \varphi_i = t_i < \infty$, otherwise they are not admissible potentials; moreover we have, for the admissibility, that $t_1 + \cdots + t_N \leq \binom{N}{2} \|c\|_\infty$.

We can then modify the potential φ_1 taking

$$\tilde{\varphi}_1(x) = \inf \left\{ c(x, \dots, x_n) - \sum_{i=2}^{N} \varphi_i(x) \right\}.$$

Of course by construction $(\tilde{\varphi}_1, \varphi_2, \dots, \varphi_N)$ are admissible potentials, and moreover we have $\varphi_1 \leq \tilde{\varphi}_1$. Denoting $\tilde{t}_1 = \sup \tilde{\varphi}_1$, in particular we can now say that

$$\tilde{\varphi}_1(x) \geq -t_2 - \cdots - t_N = \tilde{t}_1 - (\tilde{t}_1 + \cdots + t_N) \geq \tilde{t}_1 - \binom{N}{2} \|c\|_\infty$$

$$\inf \tilde{\varphi}_1 - \sup \tilde{\varphi}_1 \geq - \binom{N}{2} \|c\|_\infty.$$

Notice also that if c is L-Lipschitz (respectively K-concave), then $\tilde{\varphi}_1$ is an infimum of $(N-1)L$-Lipschitz (respectilvely $(N-1)K$-concave) functions and so we have that $\tilde{\varphi}_1$ is $(N-1)L$-Lipschitz (respectilvely $(N-1)K$-concave).

We can iterate this construction in order to get $\tilde{\varphi}_1, \dots, \tilde{\varphi}_N$ that are still an admissible N-tuple of potentials such that $\varphi_i \leq \tilde{\varphi}_i$ and

$$\inf \tilde{\varphi}_i - \sup \tilde{\varphi}_i \leq - \binom{N}{2} \|c\|_\infty.$$

Now if we had $\varphi_1, \dots, \varphi_N$ were maximizing potentials (their existence is proven for example in [21], see Proposition 2.3 and Theorem 2.21), we will have that $\tilde{\varphi}_i$ are also maximizing potentials. We can then assume that $\sup \tilde{\varphi}_i = t > 0$ is independent of i, implying also $t \leq \frac{N-1}{2} \|c\|_\infty$.

Then we can consider $\varphi(x) = \frac{1}{N} \sum_{i=1}^{N} \tilde{\varphi}_i(x)$ which will be a maximizing potential with the required property (also in the Lipschiz and concave hypothesis). □

Proof of Proposition 1.5. We can just apply Lemma 5.1 and then Theorem 5.2. □

We are now ready to prove one of the main results.

Proof of Theorem 1.2. We deal separately with each individual point.

(i) Let $\rho, \mu \in \mathscr{K}_{r,\delta}$ and $h := h(r, \delta)$ given by (5.3). We can assume $\mathscr{C}(\rho) \geq \mathscr{C}(\mu)$; let us consider then a potential φ_μ relative to μ given by Proposition 1.5. In particular we have

$$\mathscr{C}(\mu) - \mathscr{C}(\rho) \leq \int \varphi_\mu \, \mathrm{d}(\mu - \rho) \leq \|\varphi_\mu\|_\infty \cdot \|\mu - \rho\|_{\mathrm{TV}} \leq \binom{N}{2} h \cdot \|\mu - \rho\|_{\mathrm{TV}}.$$

(ii) If $\rho_n \rightharpoonup \rho$, they all satisfy $\kappa(\rho_n, r') < 1/N - \delta$ for some $r' > 0$, thanks to Lemma 2.3. But then by Lemma 5.1 there exists $h > 0$ such that $\mathscr{C}^h$ coincides with $\mathscr{C}$ on the whole sequence. Thanks to the fact that c^h is continuous and bounded, the corresponding functional $\mathscr{C}^h$ is weakly continuous and so we reach the conclusion.

(iii) Let $\rho, \mu \in \mathscr{K}_{r,\delta}$ and $h := h(r, \delta)$ given by (5.3). We can assume $\mathscr{C}(\rho) \geq \mathscr{C}(\mu)$; let us consider then a potential φ_μ relative to μ given by Proposition 1.5. In particular we have that $\frac{\varphi_\mu}{(N-1)\|c^h\|_{\text{Lip}}}$ is 1-Lipschitz; by the duality formula for W_1 we then have

$$\begin{aligned}\mathscr{C}(\mu) - \mathscr{C}(\rho) &\leq \int \varphi_\mu \, \mathrm{d}(\mu - \rho) = (N-1)\|c^h\|_{\text{Lip}} \int \frac{\varphi_\mu}{(N-1)\|c^h\|_{\text{Lip}}} \, \mathrm{d}(\mu - \rho) \\ &\leq (N-1)\|c^h\|_{\text{Lip}} \cdot W_1(\mu, \rho)\end{aligned}$$

□

Remark 5.3. It is not necessary for the cost to be Lipschitz outside each D_α, it would be enough to have it Lipschitz where the plans are supported, see for example subsection 6.1.

6 The case of Coulomb cost

In this section we will resume the main results of the paper in the case of $X = \mathbb{R}^d$ and $c(x, y) = \frac{1}{|x-y|}$. First of all we can take $m(r) = M(r) = r^{-1}$. In particular the assumptions (1.1) and (1.2) are satisfied. In the sequel also the gauge function g will be useful for summarizing some estimates:

$$g(\delta) = \begin{cases} 2(N-1) & \text{if } \delta < \frac{1}{2(N-1)} \\ \frac{N^2(N-1)}{2(1-N\delta)} & \text{if } \frac{1}{2(N-1)} \leq \delta < \frac{1}{N} \\ +\infty & \text{otherwise} \end{cases} \tag{6.1}$$

In this section $\mathscr{C}$ will be the cost introduced in (1.4) with the choice $c(x, y) = \frac{1}{|x-y|}$. Whenever used we will have that $\rho \in \mathscr{P}(\mathbb{R}^d)$, $r > 0$, $\delta \in (0, 1/N)$, and $\mathscr{K}_{r,\delta}$ will be defined as in (1.8). Moreover π will denote any optimal plan for the problem (1.4) relative to ρ. All the results of the paper concerning the Coulomb case are collected in Table 6.

A couple of remarks are in order: the finiteness conditions of Theorem 1.1 and the bound in Remark 4.2 are already clear. As for the diagonal bounds, when we consider the Coulomb cost, of course we have $D_r = D^{1/r}$ and so we will work directly with the more geometric D_r. The results Theorem 1.3(i) and (ii), which involve an estimate of the cost, are unified thanks to the gauge function g defined in (6.1).

It is worthwhile to add here also a statement that goes in the opposite direction.

Lemma 6.1. *Let $\rho \in \mathscr{P}(X)$ and $\pi \in \Pi_N(\rho)$. Then if $\kappa(\rho, r) > \frac{1}{N}$ we have $\pi(D_{2r}) \neq 0$.*

Proof. The same reasoning behind the proof of Theorem 1.1(ii) allows to prove the following: whenever $\rho(A) > \frac{1}{N}$ we have $\pi(\bigcup_{i \neq j}\{x_i, x_j \in A\}) \neq 0$. But then, by hypothesis there exists $x \in X$ such that $\rho(B(x, r)) > \frac{1}{N}$ and in particular there exists $(x_1, \ldots, x_N)$ in the support of π such that $x_i, x_j \in B(x, r)$ and $i \neq j$; this means that $|x_i - x_j| < 2r$ and so we can conclude that $\pi(D_{2r}) \neq 0$. □

The estimates for the potential are clear once we observe that for the Coulomb cost $c(x,y) = \frac{1}{|x-y|}$ we have:

- c^h is h^2-Lipschitz.
- c^h is $-\frac{3}{2}h^3$-concave.

Then it is sufficient to combine Theorem 5.2 with Lemma 5.1 to obtain the estimate for the potential. In the summary table we present also two sharper results for the regularity of the potential in the case $\operatorname{supp}\rho = \mathbb{R}^d$; these will be proven in subsection 6.1.

Finally, the explicit Lipschitz constants in Theorem 1.2 are found using the explicit estimates for the potentials.

6.1 Sharper estimates for the potentials

We discuss here also a different approach for the estimates of the potentials; this approach is tailored for $\mathbb{R}^d$ and the Coulomb cost, in the case where we also have $\operatorname{supp}\rho = \mathbb{R}^d$. First of all in the proof of Theorem 1.3 we can prove the sharper estimate

$$\sum_{i=2}^{N-1} \frac{1}{|x_1 - x_i|} \leq \frac{g(\delta)}{r} \qquad \text{for } \pi\text{-almost every } (x_1, x_2, \dots, x_N).$$

In particular then when we consider the optimal potential φ (which is unique thanks to the assumption $\operatorname{supp}\rho = \mathbb{R}^d$), since we have

$$\varphi(x_1) + \ldots + \varphi(x_N) \leq c(x_1, \dots, x_N) \quad \forall x_1, \dots, x_N \in \mathbb{R}^d$$

$$\varphi(x_1) + \ldots + \varphi(x_N) = c(x_1, \dots, x_N) \quad \text{ on } \operatorname{supp}\pi,$$

letting $A(x) = \Big\{ (x_2, x_3, \dots, x_N) \ : \ \sum_{i=2}^{N-1} \frac{1}{|x-x_i|} \leq \frac{g(\delta)}{r} \Big\}$ we can say

$$\begin{aligned}\varphi(x) &= \min\big\{ c(x, \dots, x_N) - \big(\varphi(x_2) + \ldots + \varphi(x_N)\big) \ : \ (x, x_2, \dots, x_N) \in \operatorname{supp}(\pi) \big\} \\ &= \min\big\{ c(x, \dots, x_N) - \big(\varphi(x_2) + \ldots + \varphi(x_N)\big) \ : \ (x_2, \dots, x_N) \in A(x) \big\}\end{aligned}$$

Now it is sufficient to study the regularity knowing this representation. In general we could think of something like

$$\varphi(x) = \min\{ f_i(x) \ : \ i \in I \} = \min\{ f_i(x) \ : \ i \in I(x) \}.$$

For representations like this the idea for controlling the regularity is to consider $I^\varepsilon(x)$ such that $I(y) \subset I^\varepsilon(x)$ for every y in a neighbourhood of x, then we estimate the Lipschitz constant and the concavity of φ pointwise and then let $\varepsilon \to 0$. If the function f_i are uniformly C^3 around x for $i \in I^\varepsilon(x)$ we could say

$$|\nabla\varphi|(x) \leq \lim_{\varepsilon\to 0} \sup_{i \in I^\varepsilon(x)} |\nabla f_i|(x) \qquad D^2\varphi(x) \leq \lim_{\varepsilon\to 0} \sup_{i \in I^\varepsilon(x)} D^2 f_i(x)$$

In our case, for example, we can say

$$|\nabla\varphi|(x) \leq \lim_{\varepsilon\to 0} \sup \left\{ |\nabla_x c(x, x_2, \dots, x_N)| \ : \ \sum_{i=2}^{N-1} \frac{1}{|x - x_i|} \leq \frac{g(\delta) + \varepsilon}{r} \right\} \leq \frac{g(\delta)^2}{r^2}$$

Cost finiteness	If $a(\rho) < \frac{1}{N}$ then $\mathscr{C}(\rho) < \infty$ If $a(\rho) > \frac{1}{N}$ then $\mathscr{C}(\rho) = \infty$ If $a(\rho) = \frac{1}{N}$ then $\mathscr{C}(\rho) < \infty$ if and only if $$\frac{1}{\lvert x - x_0 \rvert} \in L^1_{loc}(\rho) \qquad \forall x_0 \in \mathbb{R}^d \text{ s.t. } \rho(\{x_0\}) = \frac{1}{N}.$$
Cost estimate	If $\kappa(\rho, r) \leq \frac{1}{N}$ then $\mathscr{C}(\rho) \leq \binom{N}{2}\frac{1}{r}$.
Diagonal estimate	If $\rho \in \mathscr{K}_{r,\delta}$ and $r' < \dfrac{r}{g(\delta)}$ then $\pi(D_{r'}) = 0$.
Potential estimates	If $\rho \in \mathscr{K}_{r,\delta}$ there exists a potential φ such that $$-\binom{N}{2}\frac{g(\delta)}{r} \leq \varphi(x) \leq \frac{g(\delta)\cdot(N-1)}{2r},$$ $$\lVert\varphi\rVert_{\mathrm{Lip}} \leq \frac{(N-1)\cdot g(\delta)^2}{r^2},$$ $$D^2\varphi \leq \frac{3(N-1)\cdot g(\delta)^3}{2r^3} Id.$$ If moreover we have $\operatorname{supp}\rho = \mathbb{R}^d$ we can also assume: $$\lVert\varphi\rVert_{\mathrm{Lip}} \leq \frac{g(\delta)^2}{r^2}, \qquad D^2\varphi \leq \frac{3g(\delta)^3}{2r^3} Id.$$
Continuity for $\mathscr{C}$	For every $\rho_1, \rho_2 \in \mathscr{K}_{r,\delta}$ we have $$\lvert\mathscr{C}(\rho_1) - \mathscr{C}(\rho_2)\rvert \leq \frac{(N-1)\cdot g(\delta)^2}{r^2} \cdot W_1(\rho_1, \rho_2),$$ $$\lvert\mathscr{C}(\rho_1) - \mathscr{C}(\rho_2)\rvert \leq \frac{N(N-1)\cdot g(\delta)}{2r} \cdot \lVert\rho_1 - \rho_2\rVert_{TV}.$$

Table 1: This is a summary of results for the Coulomb cost. All the results are derived directly from the theorems in the paper, written down directly in the case c is the Coulomb cost. The only sharper result is for the potential estimates in the case $\operatorname{supp}\rho$ is the whole $\mathbb{R}^d$; those estimates are proved in subsection 6.1.

$$D^2\varphi(x) \leq \lim_{\varepsilon\to 0}\sup\left\{D^2_x c(x,x_2,\dots,x_N)\ :\ \sum_{i=2}^{N-1}\frac{1}{|x-x_i|}\leq\frac{g(\delta)+\varepsilon}{r}\right\}\leq\frac{3g(\delta)^3}{2r^3}Id,$$

where we used $|\nabla_x c(x,x_2,\dots,x_N)| \leq \sum_i \frac{1}{|x-x_i|^2} \leq \left(\sum_i \frac{1}{|x-x_i|}\right)^2$ and a similar reasoning for the estimate of $D^2\varphi$.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and the Wasserstein spaces of probability measures.* Lectures in Mathematics, ETH Zurich. Birkhäuser, 2005.

[2] U. Bindini. "Marginals with finite repulsive cost". In: (2017). arXiv: `1702.06301 [math.OC]`. URL: `https://arxiv.org/abs/1702.06301`.

[3] U. Bindini and L. De Pascale. "Optimal transport with Coulomb cost and the semiclassical limit of Density Functional Theory". In: *arxiv:1702.04957* ().

[4] G. Bouchitté et al. "Dissociating limit in Density Functional Theory with Coulomb optimal transport cost". In: (Nov. 2018). eprint: `1811.12085`. URL: `https://arxiv.org/abs/1811.12085`.

[5] G. Buttazzo, T. Champion, and L. De Pascale. "Continuity and estimates for multimarginal optimal transportation problems with singular costs". In: *Applied Mathematics & Optimization* 78.1 (2018), pp. 185–200.

[6] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi. "Optimal-transport formulation of electronic density-functional theory". In: *Phys. Rev. A* 85 (6 June 2012), p. 062502. DOI: `10.1103/PhysRevA.85.062502`. URL: `https://link.aps.org/doi/10.1103/PhysRevA.85.062502`.

[7] M. Colombo, L. De Pascale, and S. Di Marino. "Multimarginal optimal transport maps for 1-dimensional repulsive costs". In: *Canad. J. Math.* 67 (2015), pp. 350–368.

[8] M. Colombo and S. Di Marino. "Equality between Monge and Kantorovich multimarginal problems with Coulomb cost". In: *Ann. Mat. Pura Appl. (4)* 194.2 (2015), pp. 307–320. ISSN: 0373-3114. URL: `https://doi.org/10.1007/s10231-013-0376-0`.

[9] M. Colombo and F. Stra. "Counterexamples in multimarginal optimal transport with Coulomb cost and spherically symmetric data". In: *Math. Models Methods Appl. Sci.* 26.6 (2016), pp. 1025–1049. ISSN: 0218-2025. URL: `https://doi.org/10.1142/S021820251650024X`.

[10] C. Cotar, G. Friesecke, and C. Klüppelberg. "Density functional theory and optimal transportation with Coulomb cost". In: *Comm. Pure Appl. Math.* 66.4 (2013), pp. 548–599. ISSN: 0010-3640. URL: `https://doi.org/10.1002/cpa.21437`.

[11] C. Cotar, G. Friesecke, and B. Pass. "Infinite-body optimal transport with Coulomb cost". In: *Calculus of Variations and Partial Differential Equations* 54.1 (2015), pp. 717–742.

[12] C. Cotar and M. Petrache. “Next-order asymptotic expansion for N-marginal optimal transport with Coulomb and Riesz costs”. In: *arXiv preprint arXiv:1706.06008* (2017).

[13] L. De Pascale. “Optimal transport with Coulomb cost: approximation and duality”. In: *ESAIM: M2AN* 49 (2015), pp. 1643–1657.

[14] S. Di Marino, A. Gerolin, and L. Nenna. “Optimal Transportation Theory with Repulsive Costs”. In: *Topological Optimization and Optimal Transport.* Ed. by M. Bergounioux et al. Vol. 9. Radon Series on Computational and Applied Mathematics 17. RICAM. De Gruyter, 2017, pp. 204–256. arXiv: `1506.04565 [math.OC]`.

[15] W. Gangbo and A. Święch. “Optimal maps for the multidimensional Monge-Kantorovich problem”. In: *Comm. Pure Appl. Math.* 51.1 (1998), pp. 23–45. ISSN: 0010-3640. URL: `https://doi.org/10.1002/(SICI)1097-0312(199801)51:1%3C23::AID-CPA2%3E3.0.CO;2-H`.

[16] A. Gerolin, A. Kausamo, and T. Rajala. “Duality theory for multi-marginal optimal transport with repulsive costs in metric spaces”. In: *ESAIM: COCV* to appear.arXiv:1805.00880 (2018).

[17] A. Gerolin, A. Kausamo, and T. Rajala. “Non-existence of optimal transport maps for the multi-marginal repulsive harmonic cost”. In: *arXiv preprint arXiv:1805.00417* (2018).

[18] N. Ghoussoub and A. Moameni. “A self-dual polar factorization for vector fields”. In: *Comm. Pure Appl. Math.* 66.6 (2013), pp. 905–933. ISSN: 0010-3640. URL: `https://doi.org/10.1002/cpa.21430`.

[19] P. Gori-Giorgi, M. Seidl, and G. Vignale. “Density-Functional Theory for Strongly Interacting Electrons”. In: *Phys. Rev. Lett.* 103 (16 Oct. 2009), p. 166402. DOI: `10.1103/PhysRevLett.103.166402`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.103.166402`.

[20] A. Hajnal and E. Szemerédi. “Proof of a conjecture of P. Erdos”. In: *Combinatorial theory and its applications* 2 (1970), pp. 601–623.

[21] H. G. Kellerer. “Duality theorems for marginal problems”. In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 67.4 (Nov. 1984), pp. 399–432. ISSN: 1432-2064. DOI: `10.1007/BF00532047`. URL: `https://doi.org/10.1007/BF00532047`.

[22] H. A. Kierstead and A. V. Kostochka. “A Short Proof of the Hajnal–Szemerédi Theorem on Equitable Colouring”. In: *Combinatorics, Probability and Computing* 17.2 (2008), pp. 265–270. DOI: `10.1017/S0963548307008619`.

[23] Y.-H. Kim and B. Pass. “A general condition for Monge solutions in the multi-marginal optimal transport problem”. In: *SIAM J. Math. Anal.* 46.2 (2014), pp. 1538–1550. ISSN: 0036-1410. URL: `https://doi.org/10.1137/130930443`.

[24] M. Lewin, E. H. Lieb, and R. Seiringer. “Statistical Mechanics of the Uniform Electron Gas”. In: *arXiv preprint arXiv:1705.10676* (2017).

[25] E. Paolini and E. Stepanov. “Decomposition of acyclic normal currents in a metric space”. In: *Journal of Functional Analysis* 263.11 (2012), pp. 3358–3390. ISSN: 0022-1236. DOI: `10.1016/j.jfa.2012.08.009`. URL: `http://www.sciencedirect.com/science/article/pii/S0022123612003059`.

[26] B. Pass. “On the local structure of optimal measures in the multi-marginal optimal transportation problem”. In: *Calc. Var. Partial Differential Equations* 43.3-4 (2012), pp. 529–536. ISSN: 0944-2669. URL: `https://doi.org/10.1007/s00526-011-0421-z`.

[27] M. Petrache. “Decorrelation as an avatar of convexity”. 2015.

[28] F. Santambrogio. *Optimal transport for applied mathematicians*. Springer, Birkäuser, NY, 2015.

[29] M. Seidl et al. “The strictly-correlated electron functional for spherically symmetric systems revisited”. In: *arxiv:1702.05022* ().

[30] C. Villani. *Topics in optimal transportation*. 58. American Mathematical Soc., 2003.