

A reverse inequality in optimal mass transportation

Luigi De Pascale*

Dipartimento di Matematica e Informatica “Ulisse Dini”
Università degli Studi di Firenze
luigi.depascale@unifi.it

Igor Pinheiro†

Dipartimento di Matematica e Informatica “Ulisse Dini”
Università degli Studi di Firenze
igorvinicius.pereirapinheiro@unifi.it

October 13, 2026

Abstract

Controlling the ∞ -Wasserstein distance by the p -Wasserstein distance is interesting both for theoretical and numerical applications. Previous works established this type of inequality for costs that increase with the distance. In this paper we prove an analogous inequality for optimal transport problems with pointwise costs given by a decreasing function of the distance. In particular, we present a general framework that encompasses the known cases and covers repulsive costs.

1 Description of the problem and of the results

Let (X, d) be a metric space and let $\mu, \nu \in \mathcal{P}(X)$ be two probability measures. Let $c : X \times X \rightarrow [0, +\infty]$ be a “pointwise transportation cost” function. Denote, as usual in this domain, by

$$\mathcal{T}(\mu, \nu) := \{T : X \rightarrow X : T_{\#}\mu = \nu\},$$

where T is a Borel map and $\#$ denotes the push-forward of measures. The Monge or optimal transport problem related to the cost c consists of

$$\min_{\mathcal{T}(\mu, \nu)} \int_X c(x, T(x)) d\mu.$$

It is, by now, well known that the Monge’s problem is difficult and may have no solutions so one resorts to its Kantorovich relaxation. We refer, for these questions, to

*OrcId: 0000-0002-1054-5544

†OrcId: 0000-0001-8739-9146

any of the many books on optimal transport, for example [19, 20, 17]. The Kantorovich relaxation is

$$\mathcal{C}(\mu, \nu) = \min_{\lambda \in \Pi(\mu, \nu)} \int_{X \times X} c(x, y) d\lambda, \quad (1)$$

where $\Pi(\mu, \nu)$ is the space of probability measures λ on $X \times X$ such that $\pi_{\#}^1 \lambda = \mu$ and $\pi_{\#}^2 \lambda = \nu$. There is a natural identification between $T \in \mathcal{T}(\mu, \nu)$ and $(id \times T)_{\#} \mu \in \Pi(\mu, \nu)$. Existence of at least one optimizer for (1) holds under lower semi-continuity assumption on c and some mild integrability conditions. We will mostly focus on the Kantorovich version of the problem and we will point, when of interest, to the consequence for the minimizers of the Monge version, if any. We also introduce the *sup*, or supremal, version of the problem.

$$\mathcal{C}_{\infty}(\mu, \nu) := \min_{\lambda \in \Pi(\mu, \nu)} \lambda - \text{ess sup } c(x, y), \quad (2)$$

which is, by now, well known in the literature. The monotonicity of the integral implies that

$$\mathcal{C}(\mu, \nu) \leq \mathcal{C}_{\infty}(\mu, \nu). \quad (3)$$

This paper will be focused on inequalities that “reverse” (3) in the sense that allows to control $\mathcal{C}_{\infty}$ by $\mathcal{C}$. We will focus on the case of repulsive costs and, in a subsection, we will describe the literature for the attractive case and we will add some original remarks on the topic. Since the costs we consider are repulsive, they make sense also in the case $\mu = \nu = \rho$.

Definition 1.1. Let $h : [0, \infty) \rightarrow [0, \infty]$ be a non-increasing function and $\rho \in \mathcal{P}(X)$. We define the following transport costs,

$$\mathcal{C}(\rho) := \inf_{\lambda \in \Pi(\rho)} \int_{X \times X} h(d(x, y)) d\lambda(x, y). \quad (4)$$

and

$$\mathcal{C}_{\infty}(\rho) := \inf_{\lambda \in \Pi(\rho)} \lambda - \text{ess sup } h(d(x, y)). \quad (5)$$

Remark 1.2. The model case for the definition above is the so-called Coulomb interaction: The ambient space is $\mathbb{R}^3$ with the Euclidean metric and $h(t) = \frac{1}{t}$. We note that, in this case, h is not defined at $t = 0$, but instead we have $h(0^+) := \lim_{t \rightarrow 0^+} h(t) = +\infty$. The key instance in which $h(0^+) = +\infty$ plays a role is the finiteness of $\mathcal{C}(\rho)$ as noted in [4] and then studied in [7, 2] (see Prop. 2.15 below). Regardless, our results hold either if $h(0^+) < \infty$ or $h(0^+) = \infty$.

The type of reverse inequality that we will study is the following: there exists a function $\omega : [0, \infty) \rightarrow \mathbb{R}$ such that, $\omega(t) > 0$ for $t > 0$ and

$$\omega(\mathcal{C}_{\infty}(\rho)) \leq \mathcal{C}(\rho), \quad (6)$$

where, of course, ω should be explicit. This formulation of the problem was first given in [14]. The function ω , should allow situations as the one of the next example and this will need some study, later on, in this paper. The next example illustrates the need for some caution in the study of this problem.

Example 1.3. Let $h(t) = t^{-1}$ and $\rho_{\epsilon} \in \mathcal{P}(\mathbb{R})$ given by $\rho_{\epsilon}(x) = \frac{1}{2}\delta_0 + \epsilon\delta_1 + (\frac{1}{2} - \epsilon)\delta_{\frac{1}{\epsilon}}$. Then, $\mathcal{C}(\rho_{\epsilon}) = 2\epsilon(\frac{3}{2} - \epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$, but $\mathcal{C}_{\infty}(\rho_{\epsilon}) = 1$.

1.1 The case of Wasserstein distances

When $c(x, y) = d^p(x, y)$ the cost $(\min_{\lambda \in \Pi(\mu, \nu)} \int_X d^p(x, y) d\lambda)^{1/p} := \mathcal{W}_p(\mu, \nu)$ defines the so called p -Wasserstein distance between μ and ν . This is, certainly, one of the most significative cases. In [3] the authors prove that, if Ω is a bounded, connected, open subset of $\mathbb{R}^d$ with Lipschitz boundary then for every $\mu \in \mathcal{P}_{ac}(\Omega)$, every $\nu \in \mathcal{P}(\bar{\Omega})$ and every $p > 1$,

$$(\mathcal{W}_\infty(\mu, \nu))^{p+d} \leq C_{p,d}(\Omega) \|1/f\|_\infty \mathcal{W}_p^p(\mu, \nu),$$

where the constant $C_{p,d}(\Omega)$ only depends on Ω and $f = \frac{d\mu}{dx}$ is the Radon-Nikodym derivative of μ with respect to the Lebesgue measure.

So, if we consider the class of μ such that $\frac{d\mu}{dx} \geq a$ for a certain positive constant a the inequality become

$$\frac{a(\mathcal{W}_\infty(\mu, \nu))^{p+d}}{C_{p,d}} \leq \mathcal{W}_p^p(\mu, \nu). \quad (7)$$

This last inequality is a form of (6) for a specific attractive cost. We will discuss later on the different parameters in this inequality. The paper [3] contains some open problems and several other results, among which an L^∞ estimate for the p -optimal transport map.

Inequality (7) played a key role in several contexts ranging from regularity theory to PDEs. For instance, in [10] the authors used (7) to improve the $\mathcal{W}_2$ -convergence rate between stationary state solutions of the Porous Media Equation; in [18] the authors obtained uniform estimates on the L^∞ norm of the gradient of the Kantorovich potential corresponding to a JKO step; in [15] the authors developed local versions of (7), extended in [13] to more general costs, to obtain a purely variational approach to the partial regularity of the optimal transportation map which was later generalized in [16] to more generic settings than $\mathbb{R}^d$ and in [12] to the Coulomb cost.

The exponent $p+d$ in estimate (7) is related to the Lipschitz boundary of Ω , in fact, if Ω is not a Lipschitz domain, one could still have for a similar estimate, but with exponent $p+s$ and $s > d$.

Example 1.4. Let $\Omega = \{(x, y) \in \mathbb{R}^2 \mid \sqrt{|x|} \leq y \leq 1\}$ and $\mu \in \mathcal{P}(\Omega)$ be the uniform distribution over Ω . Fix $\alpha > 0$ and define $\nu = \mu|_{B^c((0,0), \alpha)} + \mu(B((0,0), \alpha))\delta_{(0,0)}$. One can see that, $\mathcal{W}_\infty(\mu, \nu) = \alpha$. Computing $\mathcal{W}_p(\mu, \nu)^p$ (up to first order) gives $\mathcal{W}_p(\mu, \nu)^p \approx C\alpha^{p+3}$, thus,

$$\frac{\mathcal{W}_p(\mu, \nu)^p}{\mathcal{W}_\infty(\mu, \nu)^{p+2}} \rightarrow 0 \text{ as } \alpha \rightarrow 0,$$

on the other hand, there exists $C_{p,\Omega} > 0$ such that

$$\frac{\mathcal{W}_\infty(\mu, \nu)^{p+3}}{C_{p,\Omega}} \leq \mathcal{W}_p^p(\mu, \nu). \quad (8)$$

An in depth analysis was done a few years later in [14] in which the authors proved that:

Theorem 1.5 (Th. 1.1, [14]). *Let (X, d) be a complete metric space, $h : [0, +\infty) \rightarrow [0, +\infty)$ a non-decreasing function with $h(t) > 0$ for all $t > 0$ and let $\mu \in \mathcal{P}(X)$. Then,*

there exists a non-decreasing function $\omega : [0, +\infty) \rightarrow [0, +\infty)$ with $\omega(t) > 0$ for all $t > 0$ such that

$$\omega(\mathcal{W}_\infty(\mu, \nu)) \leq \inf_{\lambda \in \Pi(\mu, \nu)} \int h(d(x, y)) d\lambda \quad \text{for all } \nu \in \mathcal{P}(\text{spt}(\mu))$$

if and only if $\text{spt}(\mu)$ is compact and connected. Moreover, in such case one can take $\omega(t) = \frac{1}{2}m(t/17)h(t/17)$ where

$$m(t) := \inf_{x \in \text{spt}(\mu)} \mu(B(x, t)). \quad (9)$$

Remark 1.6. The function m in (9) explains how the assumption or regularity on $\partial\Omega$ and the lower bound on the density of μ effect the constant and the exponents in inequality (7) and also explains the power 3 (which is equal to $d + 1$) in inequality (8). Then a strategy of study could be first to prove an inequality as the one in the theorem above and then to find a significant class of measures in which such inequality is more explicit.

2 Main theorem

In what follows, we consider (X, d) a complete and separable metric space, i.e. a Polish space, $\rho \in \mathcal{P}(X)$ a probability measure on X and a continuous and decreasing function $h : [0, \infty) \rightarrow (0, \infty]$. For $N \geq 1$ fixed we also consider the cost function $c : X^N \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ given by

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

And also the functionals $\mathcal{C}, \mathcal{C}_\infty : \mathcal{P}(X^N) \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ given by

$$\mathcal{C}(\lambda) = \int c(x_1, \dots, x_N) d\lambda(x_1, \dots, x_N) \quad \mathcal{C}_\infty(\lambda) = \lambda - \text{ess sup } c(x_1, \dots, x_N).$$

$\mathcal{C}(\rho)$ and $\mathcal{C}_\infty(\rho)$ will denote, respectively, the infimum of each functional over all couplings λ in the set $\Pi(\rho) = \{\lambda \in \mathcal{P}(X^N) : \pi_\#^i \lambda = \rho, 1 \leq i \leq N\}$.

Definition 2.1. Let $\rho \in \mathcal{P}(X)$ and $\alpha > 0$. The concentration of ρ at level α is given by the quantity

$$\kappa_\rho(\alpha) = \sup_{x \in X} \rho(\overline{B}(x, \alpha)).$$

The pointwise concentration of ρ is given by

$$\kappa(\rho) = \sup_{x \in X} \rho(\{x\}).$$

Remark 2.2. The pointwise concentration, $\kappa(\rho)$, first appeared in [11] where, in the study of optimal transport with Coulomb cost and N marginal constraints, it was observed that if $\kappa(\rho) < C(N)$ then the integral transport cost is finite and the optimal transport plan has support at positive distance from the diagonal. For the same reason it was used in [4]. In [2, 7] it was shown that $C(N) = \frac{1}{N}$ is the sharp constant. In particular, Th. 1.1 in [7] and Th.1.1 in [2] prove that the condition $\kappa(\rho) < \frac{1}{N}$ implies the existence of

a coupling $\lambda \in \Pi(\rho)$ which has support a positive distance from the set $\{x \in X^N \mid x_i = x_j \text{ for some } i \neq j\}$.

In the borderline case $\kappa(\rho) = \frac{1}{N}$ nothing can be said in general and one can give examples in which both $\mathcal{C}(\rho) = +\infty$ and $\mathcal{C}_\infty(\rho) = +\infty$ or both are finite of only $\mathcal{C}_\infty(\rho) = +\infty$.

In our case, if $h(0^+) := \lim_{t \rightarrow 0^+} h(t) < +\infty$ then the condition $\kappa(\rho) < \frac{1}{N}$ is not relevant for the finiteness of $\mathcal{C}_\infty(\rho)$ nor $\mathcal{C}(\rho)$. In the interest of generality, we prefer to allow $h(0^+) = +\infty$ and introduce the following assumption.

Assumption (A). *If $h(0^+) = +\infty$, then $\kappa(\rho) < \frac{1}{N}$.*

Our main goal is to prove the following Theorem.

Theorem 2.3. *Let $\rho \in \mathcal{P}(X)$, $h : [0, +\infty) \rightarrow (0, +\infty]$ be a decreasing continuous function and assume (A) holds. Then,*

$$\mathcal{C}(\rho) \geq h \left(2h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)} \right) \right) \left[\frac{N\kappa_\rho \left(h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)} \right) \right) - 1}{N(N-1)} \right]$$

Remark 2.4. The function

$$t \mapsto \mathbf{m}(t) := \frac{N\kappa_\rho(t) - 1}{N(N-1)}$$

plays the same role as the function m in (9). In [14] the authors guarantee that the function m is strictly positive by considering only the measures $\mu \in \mathcal{P}(X)$ with compact and connected support (see Lemma 2.1 in [14]). One may note that also $\mathbf{m}$ is not strictly positive. In fact, if $\kappa_\rho(t) \leq 1/N$, it will not be positive. A key step is to determine the level, as a function of $\mathcal{C}_\infty$, that one should compute $\mathbf{m}$ to guarantee positiveness. This is the content of Corollary 2.16.

Remark 2.5. The function $\mathbf{m}$, as in [14], also depends on the marginal measure ρ . We will show in Section 4 that this dependence is only on how ρ is distributed. Namely, for a class of measures that share certain common features, it is possible to prove that there exists a positive constant C , independent of each specific measure ρ , such that $\mathbf{m}(h^{-1}(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)})) \geq C > 0$.

Remark 2.6. The function

$$t \mapsto h(2h^{-1}(t))$$

is increasing. If $h(t) = t^{-1}$ then the formula above looks

$$\mathcal{C}(\rho) = \int_{X^N} \sum_{1 \leq i < j \leq N} \frac{1}{d(x_i, x_j)} d\lambda(x_1, \dots, x_N) \geq \mathcal{C}_\infty(\rho) \frac{N\kappa_\rho \left(\left(\frac{N(N-1)}{\mathcal{C}_\infty(\rho)} \right) \right) - 1}{2N^2(N-1)^2}.$$

Now we develop the necessary tools to prove Theorem 2.3.

Definition 2.7. Let (X, d) be a Polish space and $B, \beta > 0$. We define the following sets

$$\mathcal{B}^B := \{(x_1, \dots, x_N) : c(x_1, \dots, x_N) > B\} \text{ and}$$

$$\mathcal{B}_\beta := \{(x_1, \dots, x_N) : \min_{i \neq j} d(x_i, x_j) < \beta\}.$$

Remark 2.8. Note that, for the cost defined as

$$\sum_{1 \leq i < j \leq N} h(d(x_i, x_j))$$

where $h : (0, \infty) \rightarrow (0, \infty)$ is decreasing, we have $\mathcal{B}_\beta \subset \mathcal{B}^{h(\beta)}$ since

$$\min_{i \neq j} d(x_i, x_j) < \beta \Rightarrow \max_{i \neq j} h(d(x_i, x_j)) > h(\beta) \Rightarrow \sum_{1 \leq i < j \leq N} h(d(x_i, x_j)) > h(\beta).$$

Moreover, $\mathcal{B}^{\binom{N}{2}h(\beta)} \subset \mathcal{B}_\beta$ since

$$\begin{aligned} \sum_{1 \leq i < j \leq N} h(d(x_i, x_j)) &> \binom{N}{2} h(\beta) \Rightarrow \max_{i < j} h(d(x_i, x_j)) > h(\beta) \\ &\Rightarrow \min_{i < j} d(x_i, x_j) < \beta. \end{aligned}$$

Given two subsets $A, B \subset X$ and a set $\{k_1, \dots, k_L\} \subseteq \{1, \dots, N\}$ we introduce the shorthand notation

$$A^{N-L} \times_{k_1 \dots k_L} B = \prod_{k=1}^N Y_k \quad \text{where} \quad Y_k = \begin{cases} B & \text{if } k \in \{k_1, \dots, k_L\} \\ A & \text{otherwise} \end{cases}$$

Lemma 2.9. *Let $\rho \in \mathcal{P}(X)$. If $\kappa_\rho(\alpha) > \frac{1}{N}$, then for every $\lambda \in \Pi_N(\rho)$ we have $\lambda(\mathcal{B}_{2\alpha}) > 0$.*

Proof. Assume that $\lambda(\mathcal{B}_{2\alpha}) = 0$ and fix a point $z \in X$. In particular, for any $i, j \in \{1, \dots, N\}$ with $i \neq j$, we have

$$\lambda(X^{N-2} \times_{ij} B(z, \alpha)) = 0.$$

Thus $\rho(B(z, \alpha)) = \lambda(X^{N-1} \times_i B(z, \alpha)) = \lambda(B^c(z, \alpha)^{N-1} \times_i B(z, \alpha))$ where the family $\{B^c(z, \alpha)^{N-1} \times_i B(z, \alpha)\}_{1 \leq i \leq N}$ is a family of disjoint sets. Therefore,

$$N\rho(B(z, \alpha)) = \sum_{i=1}^N \lambda(B^c(z, \alpha)^{N-1} \times_i B(z, \alpha)) = \lambda\left(\bigcup_{i=1}^N B^c(z, \alpha)^{N-1} \times_i B(z, \alpha)\right) \leq 1$$

and taking the supremum over all $z \in X$ proves the contrapositive of the Lemma. $\square$

Theorem 2.10. *Given $\lambda \in \Pi_N(\rho)$ and $\alpha > 0$, then*

$$\lambda(\mathcal{B}_{2\alpha}) \geq \frac{N\kappa_\rho(\alpha) - 1}{N(N-1)} \quad (10)$$

Proof. For fixed $\alpha > 0$ and $x \in X$ denote $B := B(x, \alpha)$ and define $D_B := \{(x_1, \dots, x_N) : x_i, x_j \in B, \text{ for some } i \neq j\}$. Clearly, $D_B \subset \mathcal{B}_{2\alpha}$ and we have

$$N\rho(B) = \sum_{i=1}^N \lambda(X^{N-1} \times_i B) = \sum_{i=1}^N \lambda((X^{N-1} \times_i B) \cap D_B) + \sum_{i=1}^N \lambda((X^{N-1} \times_i B) \setminus D_B).$$

First we study the second summation. Given $(x_1, \dots, x_N) \in (X^{N-1} \times_i B) \setminus D_B$, then $x_i \in B$ and $x_j \notin B$ for all $j \neq i$. Hence, for each i ,

$$(X^{N-1} \times_i B) \setminus D_B = (B^c)^{N-1} \times_i B$$

moreover, for $i_1 \neq i_2$, $[(B^c)^{N-1} \times_{i_1} B] \cap [(B^c)^{N-1} \times_{i_2} B] = \emptyset$ therefore

$$\sum_{i=1}^N \lambda((X^{N-1} \times_i B) \setminus D_B) = \lambda\left(\bigcup_{i=1}^N (B^c)^{N-1} \times_i B\right) \leq 1.$$

Now, we handle the first summation. Note that $(x_1, \dots, x_N) \in (X^{N-1} \times_i B) \cap D_B$ means that $x_i \in B$ and $x_j \in B$ for some $j \neq i$. We have, for i fixed,

$$\begin{aligned} \lambda((X^{N-1} \times_i B) \cap D_B) &= \lambda\left([X^{N-1} \times_i B] \cap \bigcup_{j \neq i} [X^{N-1} \times_j B]\right) \\ &\leq \sum_{j \neq i} \lambda([X^{N-1} \times_i B] \cap [X^{N-1} \times_j B]) \\ &\leq (N-1)\lambda(D_B) \end{aligned}$$

for any $j \neq i$, where in the last inequality we used $[X^{N-1} \times_i B] \cap [X^{N-1} \times_j B] \subset D_B$. Finally,

$$\sum_{i=1}^N \lambda((X^{N-1} \times_i B) \cap D_B) \leq \sum_{i=1}^N (N-1)\lambda(D_B) = N(N-1)\lambda(D_B).$$

Putting everything together and recalling that $D_B \subset \mathcal{B}_{2\alpha}$ leads to

$$N\rho(B) \leq N(N-1)\lambda(\mathcal{B}_{2\alpha}) + 1 \Rightarrow \lambda(\mathcal{B}_{2\alpha}) \geq \frac{N\rho(B) - 1}{N(N-1)}$$

recalling that $B := B(x, \alpha)$, taking supremum over all $x \in X$ leads to the result. $\square$

Remark 2.11. It is worth mentioning that inequality (10) will be the key step to prove Theorem 2.3 in the same way as Lemma 2.2. in [14]. However, they are not obtained in the same way: In [14] the authors prove that $\lambda(\{d(x, y) \geq r\}) \geq \frac{m(r)}{2}$ for each $\lambda \in \Pi(\mu, \nu)$ as long as $r < \frac{W_\infty(\mu, \nu)}{17}$, while we prove that $\lambda(\mathcal{B}_{2\alpha}) \geq \mathbf{m}(\alpha)$ for every $\lambda \in \Pi(\rho)$ and $\alpha > 0$, then we determine a specific level $\tilde{\alpha}$ such that $\mathbf{m}(\tilde{\alpha}) > 0$ (see Corollary 2.16).

Remark 2.12. Observe that substituting $N = 2$ in inequality (10) we get $\lambda(\mathcal{B}_{2\alpha}) \geq \frac{2\kappa_\rho(\alpha) - 1}{2}$. In the following section (see Lemma 3.2) we will show that, for $N = 2$, it is possible to improve the estimate to $\lambda(\mathcal{B}_{2\alpha}) \geq 2\kappa_\rho(\alpha) - 1$. Therefore, it seems reasonable to expect that the bounding function can still be improved to $\frac{N\kappa_\rho(\alpha) - 1}{N-1}$. In fact, in the following example we show that such improved bound may be reached if $\kappa_\rho > 1/N$.

Example 2.13. Let $\alpha > 0$, $p \in (\frac{1}{N}, 1]$ and $(x_1, \dots, x_N) \in \mathbb{R}^{Nd}$ such that $d(x_i, x_j) > 2\alpha$ for every $i \neq j$. Define

$$\rho(x) = p\delta_{x_1} + \frac{1-p}{N-1} \sum_{i=2}^N \delta_{x_i}.$$

We denote by $[!N]$ the set of derangements (permutations without fixed points) of N elements such that if $\sigma, \varsigma \in [!N]$, then $\sigma(i) \neq \varsigma(i)$ for every $i = 1, \dots, N$.

Consider the following plan

$$\lambda := \frac{1-p}{N-1} \left[\delta_{x_1} \otimes \dots \otimes \delta_{x_N} + \sum_{\sigma \in [!N]} \delta_{x_{\sigma(1)}} \otimes \dots \otimes \delta_{x_{\sigma(N)}} \right] + \frac{Np-1}{N-1} \delta_{x_1} \otimes \dots \otimes \delta_{x_1}. \quad (11)$$

Clearly,

$$\lambda(\mathcal{B}_{2\alpha}) = \frac{Np-1}{N-1} = \frac{N\kappa_\rho(\alpha)-1}{N-1} > \frac{N\kappa_\rho(\alpha)-1}{N(N-1)}. \quad (12)$$

This is indeed a consequence of ρ having atoms with more than $1/N$ mass. In the following example we provide a situation in which $\kappa(\rho) \leq 1/N$ holds and one still gets the strict inequality in (10), even for the improved lower bound.

Example 2.14. Consider $\rho \in \mathcal{P}(\mathbb{R})$ given by $\rho(x) = \frac{1}{2}\chi_{[-1,1]}$. Let $\lambda = (\text{Id}, T)_\# \rho$ where $T(x) = x+1$ for $x \leq 0$ and $T(x) = x-1$ for $x > 0$. Then, for every $\alpha > 0$ we have

$$\lambda(\mathcal{B}_{2\alpha}) = \begin{cases} 0 & \text{if } 0 \leq \alpha < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \leq \alpha \leq 1 \end{cases}$$

while $2\kappa_\rho(\alpha) - 1 = 2\alpha - 1 < 1$ for $0 \leq \alpha < 1$.

The estimate of Theorem 2.3 is meaningful if $\mathbf{m}\left(h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)\right)$ is strictly positive. That is, we need to guarantee that $\kappa_\rho\left(h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)\right) > \frac{1}{N}$. Then, as said in the introduction, it would be useful to provide some class of probability measures such that $\mathbf{m}\left(h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)\right) > C$, for some constant $C > 0$. We start with the following Proposition, which was proven in [7] using the duality formulation for bounded costs.

Proposition 2.15 (Theorem 4.1 in [7]). *Let $\rho \in \mathcal{P}(X)$. If $\kappa_\rho(\alpha) \leq \frac{1}{N}$, then there exists $\lambda \in \Pi^N(\rho)$ such that $\lambda(\mathcal{B}_\alpha) = 0$.*

Corollary 2.16. *Let $\rho \in \mathcal{P}(X)$, $h : [0, +\infty) \rightarrow (0, +\infty]$ a decreasing and continuous function, then*

$$\kappa_\rho\left(h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)\right) > \frac{1}{N}$$

Proof. If $\kappa_\rho\left(h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)\right) \leq \frac{1}{N}$, then by Proposition 2.15, there exists $\lambda \in \Pi^N(\rho)$ such that $\lambda(\mathcal{B}_r) = 0$ for $r = h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)$. This means that, for every $x \in \text{supp}(\lambda)$

$$\min_{i \neq j} d(x_i, x_j) \geq h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)$$

observing that $N(N-1) = 2\binom{N}{2}$ and h non-increasing we have

$$\sum_{1 \leq i < j \leq N} h(d(x_i, x_j)) \leq \binom{N}{2} \max_{i \neq j} h(d(x_i, x_j)) \leq \frac{\mathcal{C}_\infty(\rho)}{2} < \mathcal{C}_\infty(\rho)$$

which leads to a contradiction as soon as one takes the essential supremum with respect to λ on the left hand side. $\square$

Remark 2.17. Observe that if $h(0^+) = +\infty$ and $\mathcal{C}_\infty(\rho) = +\infty$, then $\kappa_\rho(h^{-1}(+\infty)) = \kappa_\rho(0) = \kappa(\rho)$ and the Corollary above amounts to saying that $\kappa(\rho) > 1/N$.

Finally, we have all the ingredients for the proof of Theorem 2.3.

Proof of Theorem 2.3. Let $\lambda \in \Pi^N(\rho)$, we have

$$\begin{aligned} \int_{X^N} \sum_{1 \leq i < j \leq N} h(d(x_i, x_j)) d\lambda(x_1, \dots, x_N) \\ &\geq \int_{\mathcal{B}_{2h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)}} \sum_{1 \leq i < j \leq N} h(d(x_i, x_j)) d\lambda(x_1, \dots, x_N) \\ &\geq h\left(2h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)\right) \lambda\left(\mathcal{B}_{2h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)}\right) \\ &\geq h\left(2h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)\right) \left[\frac{N\kappa_\rho\left(h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)}{N(N-1)}\right)\right) - 1}{N(N-1)}\right] \end{aligned}$$

$\square$

3 The case of two marginals

In this section we restrict our attention to the case of two marginals and we provide simpler proofs and sharper conclusions of the results presented in the previous section. One of the key reasons for these improvements is that, for two marginals, the costly sets $\mathcal{B}^B$ and $\mathcal{B}_\beta$ in Definition 2.7 coincide. Another reason is the, so called, Fréchet Bounds stated below which correspond to the estimate in (10).

Theorem 3.1 (Fréchet Bounds). *Let (X, d) be a polish space and $\rho \in \mathcal{P}(X)$. Given $\lambda \in \Pi(\rho)$, for any measurable set $A \subset X$ it holds that*

$$\max\{0, 2\rho(A) - 1\} \leq \lambda(A \times A) \leq \rho(A)$$

Proof. The right-hand side inequality is easy: $\lambda(A \times A) \leq \lambda(A \times X) \leq \rho(A)$. For the left-hand side inequality, if $\rho(A) \leq \frac{1}{2}$ then $2\rho(A) - 1 \leq 0$, and since λ is a positive measure, we still get the inequality. If $\rho(A) > \frac{1}{2}$ then, by Lemma 2.9, $\lambda(A \times A) > 0$ and

$$\lambda(A \times A) = \lambda(A \times X) - \lambda(A \times A^c) \geq \lambda(A \times X) - \lambda(X \times A^c) = \rho(A) - \rho(A^c) = 2\rho(A) - 1$$

$\square$

Theorem 3.2. *Let (X, d) be a Polish space and $\rho \in \mathcal{P}(X)$. Given $\lambda \in \Pi(\rho)$, for every $\alpha > 0$ the following inequality holds*

$$\lambda(\mathcal{B}_{2\alpha}) \geq 2\kappa_\rho(\alpha) - 1 \quad (13)$$

Proof. Fix $\alpha > 0$. If $\kappa_\rho(\alpha) \leq \frac{1}{2}$, then the inequality is trivially true, as λ is a positive measure. Assume $\kappa_\rho(\alpha) > \frac{1}{2}$. Then, there exists $z \in X$ such that $\rho(B(z, \alpha)) > \frac{1}{2}$ and by the proof of Lemma 2.9 we have that $\lambda(B(z, \alpha) \times B(z, \alpha)) > 0$. By the Fréchet bounds it follows that

$$\lambda(\mathcal{B}_{2\alpha}) \geq \lambda(B(z, \alpha) \times B(z, \alpha)) \geq 2\rho(B(z, \alpha)) - 1$$

and taking the supremum over all $z \in X$ we obtain the result. $\square$

Example 3.3 (Strict inequality). Consider $0 < \delta < \frac{1}{2}$ and $\rho \in \mathcal{P}(\mathbb{R})$ given by $\rho = \frac{1}{3\delta}\chi_{(0,\delta)} + \frac{1}{3\delta}\chi_{(1,1+\delta)} + \frac{1}{3\delta}\chi_{(2,2+\delta)}$. For $\lambda : \frac{1}{3}\rho|_{(0,\delta)} \otimes \rho|_{(1,1+\delta)} + \frac{1}{3}\rho|_{(1,1+\delta)} \otimes \rho|_{(2,2+\delta)} + \frac{1}{3}\rho|_{(2,2+\delta)} \otimes \rho|_{(0,\delta)}$ and $2\alpha = 1 + \delta$, we have

$$\lambda(\mathcal{B}_{1+\delta}) = \frac{2}{3} > 2\kappa\left(\rho, \frac{1+\delta}{2}\right) - 1 = \frac{1}{3}.$$

The theorem below corresponds to Proposition 2.15, but we provide a different proof for the case of two marginals.

Theorem 3.4. *Let (X, d) be a polish space and $\rho \in \mathcal{P}(X)$. If $\kappa_\rho(\alpha) \leq \frac{1}{2}$, then there exists $\lambda \in \Pi(\rho)$ such that $\lambda(\mathcal{B}_\alpha) = 0$.*

Proof. Fix $\alpha > 0$ and let us introduce the cost function $c(x, y) = \chi_{\mathcal{B}_\alpha}(x, y)$ ¹. Since $\mathcal{B}_\alpha$ is open, the cost c is lower semicontinuous and we can guarantee the existence of a minimizer for

$$\inf_{\lambda \in \Pi(\rho)} \int \chi_{\mathcal{B}_\alpha}(x, y) d\lambda(x, y).$$

Assume, for the sake of contradiction, that $\lambda(\mathcal{B}_\alpha) > 0$ and let $(x, y) \in \text{supp}(\lambda) \cap \mathcal{B}_\alpha$.

From $\kappa_\rho(\alpha) \leq \frac{1}{2}$ it follows that $\lambda(B^c(y, \alpha) \times B^c(x, \alpha)) > 0$. In fact,

$$\begin{aligned} \lambda(B^c(y, \alpha) \times B^c(x, \alpha)) &= 1 - \lambda([B^c(y, \alpha) \times B^c(x, \alpha)]^c) \\ &= 1 - \lambda([B(y, \alpha) \times X] \cup [X \times B(x, \alpha)]) \\ &> 1 - \rho(B(y, \alpha)) - \rho(B(x, \alpha)) \geq 0. \end{aligned}$$

where the strictly inequality comes from $\lambda(B(y, \alpha) \times B(x, \alpha)) > 0$ since $(x, y) \in B(y, \alpha) \times B(x, \alpha)$.

Let $(\tilde{x}, \tilde{y}) \in [B^c(y, \alpha) \times B^c(x, \alpha)] \cap \text{supp}(\lambda)$. By the c -cyclicall monotonicity of the support of λ (see [17], Theorem 1.38) it follows that

$$1 \leq c(x, y) + c(\tilde{x}, \tilde{y}) \leq c(x, \tilde{y}) + c(\tilde{x}, y) = 0.$$

$\square$

¹ $\chi_{\mathcal{B}_\alpha}$ denotes the characteristic function of the set $\mathcal{B}_\alpha$ which takes the value 1 if $x \in \mathcal{B}_\alpha$ and 0 otherwise.

Corollary 3.5. *Let (X, d) be a polish space, $\rho \in \mathcal{P}(X)$ and $h : [0, +\infty) \rightarrow (0, +\infty]$ a decreasing continuous function. Then,*

$$\kappa_\rho \left(h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right) \right) > \frac{1}{2}$$

Proof. . If $\kappa \left(h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right) \right) \leq \frac{1}{2}$ then by Theorem 3.4 there exists $\lambda \in \Pi(\rho)$ such that $\lambda(\mathcal{B}_\alpha) = 0$ for $\alpha = h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right)$. Therefore, for all $(x, y) \in \text{supp}(\lambda)$,

$$d(x, y) \geq h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right)$$

and, since h is decreasing,

$$\lambda - \text{ess sup } h(d(x, y)) \leq \frac{\mathcal{C}_\infty(\rho)}{2}$$

which is a contradiction. $\square$

Theorem 3.6. *Let $\rho \in \mathcal{P}(X)$, $h : [0, +\infty) \rightarrow (0, +\infty]$ be a decreasing continuous function and assume **(A)** holds. Then,*

$$\mathcal{C}(\rho) \geq h \left(2h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right) \right) \left[2\kappa_\rho \left(h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right) \right) - 1 \right]. \quad (14)$$

Proof. From Corollary 3.5 we have that $2\kappa_\rho \left(h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right) \right) - 1 > 0$. Hence, using $\alpha = h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right)$ in Lemma 3.2 we have $\lambda \left(\mathcal{B}_{2h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right)} \right) \geq 2\kappa_\rho \left(h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right) \right) - 1$ for every $\lambda \in \Pi(\rho)$. Therefore, proceeding in the same way as in Theorem 2.3 we have

$$\int_{\mathbb{R}^{2d}} h(d(x, y)) \, d\lambda \geq h \left(2h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right) \right) \left[2\kappa_\rho \left(h^{-1} \left(\frac{\mathcal{C}_\infty(\rho)}{2} \right) \right) - 1 \right]$$

$\square$

Remark 3.7. In the case of $X = \mathbb{R}^d$, $\rho \in \mathcal{P}(\mathbb{R}^d)$ and $h(t) = \frac{1}{t}$ the inequality (14) becomes

$$\inf_{\lambda \in \Pi(\rho)} \int_{\mathbb{R}^{2d}} \frac{1}{|x - y|} \, d\lambda(x, y) \geq \frac{\mathcal{C}_\infty(\rho)}{4} \left[2\kappa_\rho \left(\frac{2}{\mathcal{C}_\infty(\rho)} \right) - 1 \right] \quad (15)$$

In the example below we return to Example 1.3 to illustrate the role of the function $\mathfrak{m}$.

Example 3.8 (Example 1.3 BIS). Let $\rho_\epsilon \in \mathcal{P}(\mathbb{R})$ given by $\rho_\epsilon(x) = \frac{1}{2}\delta_0 + \epsilon\delta_1 + (\frac{1}{2} - \epsilon)\delta_{\frac{1}{\epsilon}}$. Then, $\mathcal{C}(\rho_\epsilon) = 2\epsilon(\frac{3}{2} - \epsilon)$, $\mathcal{C}_\infty(\rho_\epsilon) = 1$ and $2\kappa_{\rho_\epsilon} \left(\frac{4}{\mathcal{C}_\infty(\rho_\epsilon)} \right) - 1 = 2(\frac{1}{2} + \epsilon) - 1 = 2\epsilon$. Then, for $\epsilon > 0$ small enough

$$\mathcal{C}(\rho_\epsilon) = 2\epsilon(\frac{3}{2} - \epsilon) > \frac{2\epsilon}{8} = \frac{\mathcal{C}_\infty(\rho_\epsilon)}{8} \left[2\kappa \left(\frac{4}{\mathcal{C}_\infty(\rho_\epsilon)} \right) - 1 \right]$$

4 Classes of probability measures

The function $t \mapsto \mathbf{m}(t) = \frac{N\kappa_\rho(h^{-1}(\frac{t}{4})) - 1}{N(N-1)}$ is decreasing and plays a fundamental role in Theorem 2.3. Examples 1.3 and 3.8 show that it is not possible to obtain a constant $C > 0$ (independent of ρ) such that

$$C \mathcal{C}_\infty(\rho) \leq \mathcal{C}(\rho). \quad (16)$$

Throughout this section we provide two classes of probability measures in which it is possible to control $\mathbf{m}$ from below and then to obtain a constant $C > 0$ depending only on the class, and not on the particular probability measure, which gives (16).

The key idea is as follows. Observe that, for any $\rho \in \mathcal{P}(X)$

$$\mathcal{C}_\infty(\rho) = \min_{\lambda \in \Pi(\rho)} h(\sup\{\alpha : \lambda(\mathcal{B}_\alpha) = 0\}) \quad (17)$$

thus, for any $\lambda \in \Pi(\rho)$,

$$h^{-1}(\mathcal{C}_\infty(\rho)) \geq \sup\{\alpha : \lambda(\mathcal{B}_\alpha) = 0\}. \quad (18)$$

If, for $\tilde{\alpha} > 0$ fixed, $\mathfrak{R}_{\tilde{\alpha}} \subset \mathcal{P}(X)$ is a collection of measures such that, for each $\rho \in \mathfrak{R}_{\tilde{\alpha}}$ there exists $\lambda \in \Pi(\rho)$ such that $\lambda(\mathcal{B}_{\tilde{\alpha}}) = 0$, it follows from (18) that $h^{-1}(\mathcal{C}_\infty(\rho)) \geq \tilde{\alpha} > 0$. If one can guarantee that, for such $\tilde{\alpha}$ the quantity $\mathbf{m}(\tilde{\alpha})$ is strictly positive, then the inequality in Theorem 2.3 becomes, for $\rho \in \mathfrak{R}_{\tilde{\alpha}}$,

$$\mathcal{C}(\rho) \geq C_{\tilde{\alpha}} h\left(2h^{-1}\left(\frac{\mathcal{C}_\infty(\rho)^{-1}}{N(N-1)}\right)\right). \quad (19)$$

where $C_{\tilde{\alpha}} = \mathbf{m}(\tilde{\alpha})$.

In this section we will provide two classes $\mathfrak{R}_{\tilde{\alpha}}$: The first will be a collection of unimodal isotropic probability measures on $\mathbb{R}^d$, and the second will be the collection of discrete measures. To avoid obscuring the meaning of the results by notational heaviness, we will restrict ourselves to the case of two marginals and to $h(s) = s^{-1}$, but the results can be naturally posed for an arbitrary number marginals and more general decreasing functions h .

We start with a generic condition that allows one to obtain the constant $C_{\tilde{\alpha}}$.

Theorem 4.1. *Let $\rho \in \mathcal{P}(X)$, $h : (0, \infty) \rightarrow (0, \infty)$ be given by $h(s) = s^{-1}$. Define $r_\rho = \sup\{r > 0 \mid \kappa_\rho(r) \leq 1/2\}$. Let $0 < \delta < \frac{1}{2}$. Then there exists $C = C(\delta) > 0$ (not depending on ρ) such that, if*

$$\inf_{x \in \mathbb{R}^d} \rho(B^c(x, 2r_\rho)) \leq \delta < \frac{1}{2}, \quad (20)$$

then

$$\mathcal{C}_\infty(\rho) \leq C \mathcal{C}(\rho).$$

Proof. By Theorem 3.4 with $\alpha = r_\rho$ we have $\mathcal{C}_\infty(\rho) \leq \frac{1}{r_\rho}$, from which we derive

$$\frac{2}{\mathcal{C}_\infty(\rho)} \geq 2r_\rho \quad (21)$$

From $\inf_{x \in \mathbb{R}^d} \rho(B^c(x, 2r_\rho)) \leq \delta$ it follows that $\kappa_\rho(2r_\rho) \geq 1 - \delta > \frac{1}{2}$. From (21) we obtain

$$2\kappa_\rho\left(\frac{2}{\mathcal{C}_\infty}\right) - 1 \geq 2\kappa_\rho(2r_\rho) - 1 > 1 - 2\delta,$$

which implies that $\mathfrak{m}(\frac{2}{\mathcal{C}_\infty(\rho)}) > 1 - 2\delta$ where $\mathfrak{m}$ is the function in Theorem 3.6. We conclude by taking $C := \frac{1-2\delta}{4} > 0$. $\square$

4.1 Unimodal isotropic probability measures

Several classes of measures satisfy the assumptions of Theorem 4.1. We will analyze the case of unimodal isotropic probability distributions which includes the Cauchy distribution, the Gaussian distribution, the Laplace distribution and the Logistic distribution, to cite a few. We will also show that for absolutely continuous probability measures which are unimodal and isotropic, the quantity $\mathcal{C}_\infty$ can be fully characterized as inverse of the diameter of ball (Proposition 4.5) which allows us to give a better expression of $\mathfrak{m}$.

Lemma 4.2. *Let $\rho \in \mathcal{P}(\mathbb{R}^d)$, then*

$$\mathcal{C}_\infty(\rho) \geq \sup_{\alpha} \left\{ \frac{1}{2\alpha} : \kappa_\rho(\alpha) - \frac{1}{2} > 0 \right\}$$

Proof. Fix $\alpha > 0$ such that $\kappa_\rho(\alpha) > \frac{1}{2}$. Let $z \in \text{supp}(\rho)$ be such that $\rho(\bar{B}(z, \alpha)) > \frac{1}{2}$. By the proof of Lemma 2.9 for every $\lambda \in \Pi(\rho)$ we have that $\lambda(\bar{B}(z, \alpha) \times \bar{B}(z, \alpha)) > 0$, that is, there exists a pair $(x, y) \in \text{supp}(\lambda)$ such that $|x - y| \leq 2\alpha$ which implies that

$$\lambda - \text{ess sup} \frac{1}{|x - y|} \geq \frac{1}{2\alpha}, \quad \forall \lambda \in \Pi(\rho)$$

taking supremum over all α such that $\kappa(\alpha) > 1/2$, and then infimum over all $\lambda \in \Pi(\rho)$ leads to

$$\mathcal{C}_\infty(\rho) \geq \sup \left\{ \frac{1}{2\alpha} : \kappa_\rho(\alpha) - \frac{1}{2} > 0 \right\}.$$

$\square$

The following Theorem was already proven in [5] (see Appendices A and B).

Theorem 4.3. *Given $\rho \in \mathcal{P}_{ac}(\mathbb{R}^d)$, there exists $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $T_\# \rho = \rho$ and T is an optimizer in $\mathcal{C}(\rho)$. Moreover, if ρ is radially symmetric, i.e. $\rho(x) = \rho(|x|)$ then $T(x) = -\frac{x}{|x|}\tau(|x|)$ and τ is given by ,*

$$\tau(r) = F^{-1}(1 - F(r)), \tag{22}$$

where $F(r) = \int_0^r d\omega_d \rho(s) s^{d-1} ds$.

In the next result we will leverage Theorem 4.3 to obtain a characterization of $\mathcal{C}_\infty(\rho)$ in terms of the inverse of the diameter the ball. Existence of optimal maps of the form (22) have been proven only in the case $N = 2$ for any dimension d (see [9]), and for

any number of marginals N only in dimension 1 (see [6]). Counterexamples have been provided in [8] for spherically symmetric probability measures in $\mathbb{R}^2$ and 3 marginals. Therefore one should not expect the strategy for this characterization to work in case $N \geq 3$ and $d \geq 2$. We start by recalling the definition of unimodal distribution.

Definition 4.4. An absolutely continuous probability measure $\rho \in \mathcal{P}(\mathbb{R}^d)$ is said to be unimodal if, for every $A > 0$ the set $\{x \in X : \rho(x) > A\}$ is convex.

Proposition 4.5. Let $\rho \in \mathcal{P}_{ac}(\mathbb{R}^d)$ be unimodal. If ρ is radially symmetric, i.e. $\rho(x) = \rho(|x|)$, then

$$\mathcal{C}_\infty(\rho) = \sup_r \left\{ \frac{1}{2r} : \kappa_\rho(r) - \frac{1}{2} > 0 \right\}$$

Proof. Let T be the optimal map in Theorem 4.3 and τ the respective 1-dimensional map. We want to estimate the quantity $\|\text{Id} - T\|_\infty$ from below.

Let r_ρ be the number such that $\kappa_\rho(r_\rho) = \frac{1}{2}$. This number exists thanks to the continuity and positivity of the density ρ . Moreover, since ρ is unimodal, $r \mapsto \rho(r)$ is nonincreasing, thus $\kappa_\rho(r) = \rho(B(0, r)) = F(r)$ where F is the 1 dimensional cumulative distribution function defined in Theorem 4.3. In particular, $F(r_\rho) = \frac{1}{2}$.

Substituting r_ρ in (22) we obtain $\tau(r_\rho) = r_\rho$. Now we estimate $\|\text{Id} - T\|_\infty$.

$$\|\text{Id} - T\|_\infty \geq |x - T(x)| = |x| + \tau(|x|),$$

and choosing $r = r_\rho$ we obtain $\|\text{Id} - T\|_\infty \geq 2r_\rho$.

Define $\lambda_T := (\text{Id}, T)_\# \rho$, thanks to Theorem 4.2

$$\frac{1}{2r_\rho} \geq \lambda_T - \text{ess sup}_{|x-y|} \frac{1}{|x-y|} \geq \mathcal{C}_\infty(\rho) \geq \sup_r \left\{ \frac{1}{2r} : \kappa_\rho(r) > \frac{1}{2} \right\}$$

and noting that the left-most term equals the right-most term, we conclude. $\square$

The following example highlights that, without unimodality, one should not expect the $\mathcal{C}_\infty(\rho)$ characterization, as the estimate in Lemma 4.2 might be strict.

Example 4.6. Consider $h(t) = \frac{1}{t}$ and $\rho \in \mathcal{P}(\mathbb{R})$ given by $\rho = \frac{1}{3\delta} \chi_{(0,\delta)} + \frac{1}{3\delta} \chi_{(1,1+\delta)} + \frac{1}{3\delta} \chi_{(2,2+\delta)}$. Then, the smallest α such that $\kappa(\alpha) = \frac{1}{2}$ is $\alpha = \frac{2+\delta}{4}$. On the other hand, for any $\lambda \in \Pi(\rho)$ we have $\lambda - \text{ess sup}_{|x-y|} \frac{1}{|x-y|} \geq 1$, thus

$$\mathcal{C}_\infty(\rho) \geq \lambda - \text{ess sup}_{|x-y|} \frac{1}{|x-y|} \geq 1 > \frac{2}{2+\delta} = \frac{1}{2\alpha}$$

An immediate consequence of the characterization of $\mathcal{C}_\infty(\rho)$ is an improvement on Corollary 3.5 along with an expression for $\mathfrak{m}$ that no longer involves $\mathcal{C}_\infty(\rho)$ explicitly.

Corollary 4.7. Let $\rho \in \mathcal{P}_{ac}(\mathbb{R}^d)$ be a unimodal and radially symmetric probability distribution, then

$$\kappa_\rho(\mathcal{C}_\infty(\rho)^{-1}) = \kappa_\rho(2r_\rho) > \frac{1}{2}.$$

Furthermore,

$$\mathcal{C}(\rho) \geq C_\rho \mathcal{C}_\infty(\rho)$$

and $C_\rho = \kappa_\rho(2r_\rho) - \frac{1}{2}$.

In the following example we highlight that C_ρ depends only on how ρ is distributed. We show that, for a Gaussian distribution, C_ρ depends only on the dimension and not on the variance. This can be interpreted as the 68–95–99.7 rule in Statistics which states that, in a normal distribution, 68% of the values fall within one standard deviation of the mean, 95% within two and 99.7% within three standard deviations. Our example highlights that, in a normal distribution, 96.87% of the data falls within two median radius of the mean, where the median radius is the r_ρ which satisfies $\kappa_\rho(r_\rho) = 1/2$.

Example 4.8. Consider the Gaussian probability measure with variance σ^2 and mean μ , which we assume to be zero as it does not play a role in the computations,

$$\rho(x) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} e^{-\frac{|x|^2}{2\sigma^2}}.$$

By Corollary 4.7, $C_\rho = 2\kappa_\rho(2r_\rho) - 1$ depends on σ and d . One can compute

$$\kappa_\rho(t) = \sup_{x \in \mathbb{R}^d} \rho(B(x, t)) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \int_{B(0, t)} e^{-\frac{|x|^2}{2\sigma^2}} dx = \frac{d\omega_d}{2\pi^{\frac{d}{2}}} \gamma\left(\frac{d}{2}, \frac{t^2}{2\sigma^2}\right),$$

where γ is the lower incomplete Gamma function. From $2\kappa_\rho(r_\rho) = 1$ we have

$$\frac{\pi^{\frac{d}{2}}}{d\omega_d} = \gamma\left(\frac{d}{2}, \frac{r_\rho^2}{2\sigma^2}\right)$$

which shows that the function $\sigma \mapsto \gamma\left(\frac{d}{2}, \frac{r_\rho^2}{2\sigma^2}\right)$ is constant. Therefore, the function $\sigma \mapsto 2\kappa_\rho(2r_\rho) - 1$ is constant and C_ρ depends only on the dimension d .

In the particular case of $d = 2$ we can compute κ_ρ explicitly as $\kappa_\rho(t) = 1 - e^{-\frac{t^2}{2\sigma^2}}$, thus

$$2\kappa_\rho(t) - 1 = 1 - 2e^{-\frac{t^2}{2\sigma^2}} \text{ and } r_\rho = (2\sigma^2 \ln 2)^{\frac{1}{2}}.$$

Therefore, $\kappa_\rho(2r_\rho) - \frac{1}{2} = \frac{15}{32}$.

We conclude this section with an application of Theorem 4.1.

Proposition 4.9. *Let $\rho \in \mathcal{P}(\mathbb{R}^d)$ be an absolutely continuous probability measure with density of the form $\rho(x) = Ke^{-g(|x|)}$ where $K > 0$ is a normalization constant and $g : [0, +\infty) \rightarrow [0, +\infty)$ is a positive and increasing convex function. For all measures of this form*

$$\mathcal{C}(\rho) \geq \frac{1}{4} \mathcal{C}_\infty(\rho).$$

Proof. Let $\tilde{\rho} \in \mathcal{P}([0, +\infty))$ be the probability distribution with density given by $\tilde{\rho}(t) = K \exp(-g(t))t^{d-1}d\omega_d$ for $t \in [0, +\infty)$. Observe that $\tilde{\rho}$ is logarithmically concave, as g is convex. Therefore,

$$S(t) = \int_t^{+\infty} \tilde{\rho}(t) dt$$

is also logarithmically concave. Let r_ρ be the real number which solves the equation $2\kappa_\rho(r_\rho) = 1$, then r_ρ is also a solution for $2\kappa_{\tilde{\rho}}(r_\rho) = 1$ and we have

$$\frac{\ln(S(2r_\rho)) - \ln(S(r_\rho))}{2r_\rho - r_\rho} \leq \frac{\ln(S(r_\rho)) - \ln(S(0))}{r_\rho - 0} \Rightarrow S(2r_\rho) \leq \frac{1}{4} \quad (23)$$

where the implication comes from noting that $S(0) = 1$ and $S(r_\rho) = \frac{1}{2}$. Observing that $S(2r_\rho) = \inf_{x \in \mathbb{R}^d} \rho(B^c(x, 2r_\rho))$ it follows from (23) and Theorem 4.1 with $\delta = \frac{1}{4}$ that

$$\mathcal{C}(\rho) \geq \frac{1}{4} \mathcal{C}_\infty(\rho)$$

□

Example 4.10. Consider a probability measure in $\mathbb{R}$ distributed as

$$\rho_\nu(x) = K(\nu) \frac{1}{\left(1 + \frac{|x|^2}{\nu}\right)^{\frac{\nu+1}{2}}}$$

where $\nu \geq 1$ and $K(\nu)$ is a normalization constant. Notice that this is the radially symmetric Student-t distribution and for $\nu = 1$ one recovers the Cauchy distribution, however as $\nu \rightarrow +\infty$ one obtains the normal distribution. For $\nu = 1$ one may compute

$$\rho_\nu(B^c(0, 2r_\rho)) = 1 - 2 \frac{\arctan(2)}{\pi} > \frac{1}{4}$$

while, for ν large enough,

$$\rho_\nu(B^c(0, 2r_\rho)) \leq \frac{1}{4}.$$

4.2 The case of discrete measures

Theorem 4.11. Let $\rho \in \mathcal{P}(\mathbb{R}^d)$ be a discrete measure, i.e. $\rho(x) = \sum_{i=1}^M \rho_{x_i} \delta_{x_i}$ where $x_1, \dots, x_M$ are distinct points in $\mathbb{R}^d$ and ρ_{x_i} are positive real numbers satisfying $\sum_{i=1}^M \rho_{x_i} = 1$ and assume that $\rho_{x_i} < 1/2$ for every $1 \leq i \leq M$. For every $\delta > 0$ such that

$$\min_{1 \leq i \leq M} \rho_{x_i} \geq \delta, \quad (24)$$

then

$$\mathcal{C}(\rho) \geq \frac{\delta}{2} \mathcal{C}_\infty(\rho).$$

Proof. Define $r_\rho = \sup\{r > 0 : \kappa_\rho(r) \leq \frac{1}{2}\}$ and $x_\rho \in \mathbb{R}^d$ the point realizing the supremum of $\kappa_\rho(r_\rho)$. By Theorem 3.4 we have $\mathcal{C}_\infty(\rho) \leq \frac{1}{r_\rho}$ and hence

$$\frac{2}{\mathcal{C}_\infty(\rho)} \geq 2r_\rho.$$

Since r_ρ is the largest radius satisfying $\kappa_\rho(r_\rho) \leq \frac{1}{2}$, it follows that $\kappa_\rho(2r_\rho) > \frac{1}{2}$. That is,

$$0 < 2\kappa(2r_\rho) - 1 = 2\rho(\bar{B}(x_\rho, 2r_\rho) \setminus \bar{B}(x_\rho, r_\rho))$$

which means that, for some $i \in \{1, \dots, M\}$ we have $x_i \in \bar{B}(x_\rho, 2r_\rho) \setminus \bar{B}(x_\rho, r_\rho)$, thus

$$2\kappa(2r_\rho) - 1 \geq 2 \min_{1 \leq i \leq M} \rho_{x_i} \geq 2\delta.$$

Therefore, $\mathbf{m}(\frac{2}{\mathcal{C}_\infty(\rho)}) \geq 2\delta$ where $\mathbf{m}$ is the function defined in Theorem 3.6 and we conclude by taking $C := \frac{\delta}{2}$. $\square$

Remark 4.12. We highlight that the Theorem 4.11 is a Corollary of 4.1. In fact, condition (20) implies that $\rho(B(x_\rho, 2r_\rho) \setminus B(x_\rho, r_\rho)) \geq \delta - \frac{1}{2}$ which implies a lower bound on $\{\rho_{x_i}\}_{i=1}^M$. However, we preferred to keep it as an independent result.

In [1] the authors discuss estimates in the same fashion as [3] and [14] for general cost functions $c : X \times X \rightarrow \mathbb{R}$, X a Polish space, but restricted to marginals $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}(X)$ being discrete measures. Their strategy is as follows: Consider the set $\Pi_{\text{opt}}(\mu, \nu)$ of couplings $\pi \in \mathcal{P}(X \times X)$ between μ and ν which are optimal for the cost function c . Among these we call *trim* the one with the smallest cardinality, i.e. the *trim* plan π^* satisfies

$$\#\text{supp}(\pi^*) \leq \#\text{supp}(\pi), \quad \forall \pi \in \Pi_{\text{opt}}(\mu, \nu).$$

For a *trim* plan it is possible to find ([1], Theorem 3) a decomposition $(\mu^{(d)}, \mu^{(c)})$ of μ and a decomposition $(\nu^{(d)}, \nu^{(c)})$ of ν , and a pair of functions $(h^{(1)}, h^{(2)})$ such that

$$\mu = \mu^{(d)} + \mu^{(c)} \quad \text{and} \quad \nu = \nu^{(d)} + \nu^{(c)}, \quad (25)$$

$$\pi^* = (\text{Id}, h^{(1)})_{\#} \mu^{(d)} + (h^{(2)}, \text{Id})_{\#} \nu^{(d)}. \quad (26)$$

Such decomposition is called *diffusive model* associated with the *trim* minimizer π^* and a consequence of the possibility of this decomposition is the following

Theorem 4.13 ([1], Corollary 4). *Let $\mu, \nu \in \mathcal{P}(X)$ be positive discrete measure and $c : X \times X$ a cost function. If $\pi^* \in \Pi_{\text{opt}}$ is trim, then*

$$\int_{X \times X} c(x, y) d\pi^* \geq \alpha \min_{\pi \in \Pi(\mu, \nu)} \|c(x, y)\|_{L^\infty_\pi(X \times X)}$$

where

$$\alpha = \min_{x \in \text{supp}(\mu^{(d)}), y \in \text{supp}(\nu^{(d)})} \{\mu_x^{(d)}, \nu_y^{(d)}\}$$

In the next proposition we show that on a Polish space X , $\mu = \nu \in \mathcal{P}(X)$, X and $c(x, y) = d(x, y)^{-1}$ our function $\mathbf{m}$ obtained in Theorem 3.6 is bigger than α as defined Theorem 4.13.

Proposition 4.14. *Let $\rho \in \mathcal{P}(X)$ be a positive and discrete probability measure. Let $(\bar{\rho}^{(d)}, \bar{\rho}^{(c)})$ and $(\underline{\rho}^{(d)}, \underline{\rho}^{(c)})$ be any diffusive system associated with a trim solution π , then*

$$\mathbf{m}\left(\frac{2}{\mathcal{C}_\infty(\rho)}\right) \geq \alpha$$

where

$$\alpha = \min_{x \in \text{supp}(\mu^{(d)}), y \in \text{supp}(\nu^{(d)})} \{\mu_x^{(d)}, \nu_y^{(d)}\}$$

Proof. We have

$$\rho = \underline{\rho}^{(d)} + \underline{\rho}^{(c)} \text{ and } \rho = \bar{\rho}^{(d)} + \bar{\rho}^{(c)} \quad (27)$$

and $(h^{(1)}, h^{(2)})$ such that

$$\pi = (\text{Id}, h^{(1)})_{\#} \underline{\rho}^{(d)} + (h^{(2)}, \text{Id})_{\#} \bar{\rho}^{(d)}.$$

As a consequence the decomposition (27) we have

$$\min_{z \in \text{supp}(\rho)} \rho_z \geq \min_{x \in \text{supp}(\underline{\rho}^{(d)})} (\underline{\rho}^{(d)})_x \text{ and } \min_{z \in \text{supp}(\rho)} \rho_z \geq \min_{y \in \text{supp}(\bar{\rho}^{(d)})} (\bar{\rho}^{(d)})_y. \quad (28)$$

Let $r_\rho = \sup\{r > 0 \mid \kappa_\rho(r) \leq \frac{1}{2}\}$, then $\kappa_\rho(2r_\rho) > \frac{1}{2}$ which implies that there exists $z \in \text{supp}(\rho)$ such that $\rho(B(x_\rho, 2r_\rho) \setminus B(x_\rho, r_\rho)) \geq \rho_z$ where x_ρ is the point in X realizing the supremum in $\kappa_\rho(r_\rho)$. Thus,

$$\begin{aligned} \mathfrak{m}\left(\frac{2}{\mathcal{C}_\infty}\right) &\geq \kappa_\rho(2r_\rho) - \frac{1}{2} \\ &= \rho(\bar{B}(x_\rho, 2r_\rho) \setminus \bar{B}(x_\rho, r_\rho)) \\ &\geq \min_{z \in \text{supp}(\rho)} \rho_z \\ &\geq \min_{x \in \text{supp}(\underline{\rho}^{(d)}), y \in \text{supp}(\bar{\rho}^{(d)})} \{(\underline{\rho}^{(d)})_x, (\bar{\rho}^{(d)})_y\} = \alpha \end{aligned}$$

where the last inequality follows from (28). $\square$

Remark 4.15. Proposition 4.14 is not claiming that constant in Theorem 4.11 is better than the constant α in 4.13. The number $\delta/2$ could be smaller than α due to the $1/2$ factor. Instead, this Proposition only highlights how a more general strategy encompasses some particular cases already present in the literature.

Declarations

Conflict of Interest The authors declare that there is no Conflict of Interest.

Data Availability Data sharing is not applicable to this article, as no data sets were generated or analyzed during the current study.

References

- [1] Gennaro Auricchio and Marco Veneroni. On the structure of optimal transportation plans between discrete measures. *Applied Mathematics & Optimization*, 85(3):42, 2022.
- [2] Ugo Bindini. Marginals with finite repulsive cost. *Can. J. Math.*, 72(2):373–391, 2020.

- [3] Guy Bouchitte, Chloé Jimenez, and Rajesh Mahadevan. A new L^∞ estimate in optimal mass transport. *Proceedings of the American Mathematical Society*, 135(11):3525–3535, 2007.
- [4] Giuseppe Buttazzo, Thierry Champion, and Luigi De Pascale. Continuity and estimates for multimarginal optimal transportation problems with singular costs. *Applied Mathematics & Optimization*, 78:185–200, 2018.
- [5] Giuseppe Buttazzo, Luigi De Pascale, and Paola Gori-Giorgi. Optimal-transport formulation of electronic density-functional theory. *Physical Review A*, 85(6):062502, 2012.
- [6] Maria Colombo, Luigi De Pascale, and Simone Di Marino. Multimarginal optimal transport maps for one-dimensional repulsive costs. *Canadian Journal of Mathematics*, 67(2):350–368, 2015.
- [7] Maria Colombo, Simone Di Marino, and Federico Stra. Continuity of multimarginal optimal transport with repulsive cost. *SIAM Journal on Mathematical Analysis*, 51(4):2903–2926, 2019.
- [8] Maria Colombo and Federico Stra. Counterexamples in multimarginal optimal transport with Coulomb cost and spherically symmetric data. *Mathematical Models and Methods in Applied Sciences*, 26(06):1025–1049, 2016.
- [9] Codina Cotar, Gero Friesecke, and Claudia Klüppelberg. Density functional theory and optimal transportation with Coulomb cost. *Communications on Pure and Applied Mathematics*, 66(4):548–599, 2013.
- [10] Noemi David, Alpár R Mészáros, and Filippo Santambrogio. Improved convergence rates for the Hele-Shaw limit in the presence of confining potentials. *arXiv preprint arXiv:2405.07227*, 2024.
- [11] Luigi De Pascale. On c-cyclical monotonicity for optimal transport problem with Coulomb cost. *European Journal of Applied Mathematics*, 30(6):1210–1219, 2019.
- [12] Gero Friesecke and Tobias Ried. Partial regularity of optimal transport with Coulomb cost, 2025.
- [13] Cristian E Gutiérrez and Annamaria Montanari. L^∞ -estimates in optimal transport for non quadratic costs. *Calculus of Variations and Partial Differential Equations*, 61(5):163, 2022.
- [14] Heikki Jylhä and Tapio Rajala. L^∞ estimates in optimal mass transportation. *Journal of Functional Analysis*, 270(11):4297–4321, 2016.
- [15] Goldman Michael and Otto Felix. A variational proof of partial regularity for optimal transportation maps. *Annales scientifiques de l’Ecole Normale Supérieure*, 53(5):1209–1233, 2020.
- [16] Felix Otto, Maxime Prod’Homme, and Tobias Ried. Variational approach to regularity of optimal transport maps: general cost functions. *Annals of PDE*, 7(2):17, 2021.

- [17] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing, 2015.
- [18] Filippo Santambrogio and Gayrat Toshpulatov. Strong L^2H^2 convergence of the JKO scheme for the Fokker–Planck equation. *Archive for Rational Mechanics and Analysis*, 248(6):99, 2024.
- [19] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- [20] C. Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.