Counterexamples in multimarginal optimal transport with Coulomb cost and spherically symmetric data

Maria Colombo [*]   Federico Stra [†]

November 29, 2015

Abstract

We disprove a conjecture in Density Functional Theory, relative to multimarginal optimal transport maps with Coulomb cost. In the case of spherically symmetric data, which model for instance Litium and Berillium atoms, we show that some special maps, introduced by Seidl, Gori Giorgi and Savin [17, 18], are not always optimal in the corresponding transport problem. We also provide examples of maps satisfying optimality conditions for special classes of data.

1 Introduction

A natural problem in Quantum Physics consists in studying the behavior of N electrons subject to the interaction with some nuclei, their mutual interaction and the effect of an external potential. In this setting, a relevant quantity is the ground state energy of the system, which can be found by solving the Schrödinger equation. However, this procedure is computationally very costly even for a small number of electrons; Density Functional Theory proposes an alternative method to compute the ground state energy and was first introduced by Hohenberg and Kohn [12] and then by Kohn and Sham [13]. Because of its low computational cost and of its accuracy, it is considered the most popular method for electronic structure calculations in condensed matter physics and quantum chemistry.

In [3, 6] the authors present a mathematical model for the strong interaction limit of Density Functional Theory; they study the minimal interaction of N electrons and the semiclassical limit of DFT. They want to determine the ground state energy

$$E_0 = \min_{\psi}\{T[\psi] + V_{ee}[\psi] + V_{\text{ext}}[\psi]\}$$

[*]Scuola Normale Superiore, Pisa, `maria.colombo@sns.it`.

[†]Scuola Normale Superiore, Pisa, `federico.stra@sns.it`.

of a system of N electrons in an external potential (orbiting the nucleus of an atom, for instance), where $\psi \in H^1(\mathbb{R}^{3N}; \mathbb{C})$, $\|\psi\|_{L^2} = 1$, is the wave function of the system and the three terms appearing on the right are respectively the kinetic energy

$$T[\psi] = \frac{1}{2} \int_{\mathbb{R}^{3N}} |\nabla \psi(x_1, \ldots, x_N)|^2 \ dx_1 \cdots dx_N,$$

the electron-electron interaction energy

$$V_{ee}[\psi] = \sum_{1 \leq i < j \leq N} \int_{\mathbb{R}^{3N}} \frac{1}{|x_i - x_j|} |\psi(x_1, \ldots, x_N)|^2 \ dx_1 \cdots dx_N,$$

and the external energy

$$V_{\text{ext}}[\psi] = \sum_{i=1}^{N} \int_{\mathbb{R}^{3N}} v(x_i) |\psi(x_1, \ldots, x_N)|^2 \ dx_1 \cdots dx_N$$

induced by a given potential $v : \mathbb{R}^3 \to \mathbb{R} \cup \{\infty\}$. The idea is then to express everything in terms of the single electron density induced by ψ (denoted by $\psi \downarrow \rho$)

$$\rho(x) = \int_{\mathbb{R}^{3(N-1)}} |\psi(x, x_2, \ldots, x_N)|^2 \ dx_2 \cdots dx_N.$$

The external energy is easily rewritten as $V_{\text{ext}}[\psi] = N \int_{\mathbb{R}^3} v(x) \, d\rho(x)$. Hohenberg and Kohn then write

$$E_0 = \min_{\rho} \left\{ F_{HK}(\rho) + N \int_{\mathbb{R}^3} v(x) \, d\rho(x) \right\}$$

where

$$F_{HK}(\rho) = \min_{\psi \downarrow \rho} \{ T[\psi] + V_{ee}[\psi] \}$$

is the *universal Hohenberg-Kohn functional*. From the physical point of view, the difficulty lies in approximating this functional. The simplification introduced by Kohn and Sham consists in writing $F_{HK}(\rho) = F_{KS}(\rho) + F_{xc}(\rho)$ where

$$F_{KS}(\rho) = \min_{\psi \downarrow \rho} T[\psi]$$

is the *Kohn-Sham functional* and $F_{xc}(\rho)$ is the so called *exchange-correlation energy*, which takes into account the electron-electron interaction and needs to be accurately estimated. Clearly $F_{xc}(\rho) \geq \min_{\psi \downarrow \rho} V_{ee}[\psi]$ and there are situations in which this estimate provides a suitable approximation, that is we can assume

$$F_{HK}(\rho) \simeq F_{KS}(\rho) + \min_{\psi \downarrow \rho} V_{ee}[\psi],$$

for example when the electron-electron interaction is preponderant, the so called *strictly correlated electrons* regime.

It's at this point that the optimal transport theory comes into play: the minimization $\min_{\psi \downarrow \rho} V_{ee}[\psi]$ can be seen as an instance of the Monge multimarginal optimal transport problem.

This problem (for which we refer the reader to the recent survey [9], where the state of the art about it is described) consists in the minimization

$$(M) = \inf \left\{ \int_{\mathbb{R}^n} C(x, T_2(x), \dots, T_N(x)) \, d\rho(x) : T_2, \dots, T_N \in \mathcal{T}(\rho) \right\}, \qquad (1.1)$$

where $\rho \in \mathscr{P}(\mathbb{R}^n)$ is a given probability measure, $C : (\mathbb{R}^n)^N \to [0, \infty]$ is the Coulomb interaction

$$C(x_1, \dots, x_N) = \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|} \qquad \forall (x_1, \dots, x_N) \in (\mathbb{R}^n)^N, \qquad (1.2)$$

and $\mathcal{T}(\rho)$ is the set of admissible transport maps

$$\mathcal{T}(\rho) = \left\{ T : \mathbb{R}^n \to \mathbb{R}^n \text{ Borel} : T_\sharp \rho = \rho \right\}.$$

Since the cost is symmetric, a natural variant of the Monge problem allows only cyclical maps

$$(M_{\mathrm{cycl}}) = \inf \left\{ \int_{\mathbb{R}^n} C(x, T(x), \dots, T^{(N-1)}(x)) \, d\rho(x) : T \in \mathcal{T}(\rho),\, T^{(N)} = \mathrm{Id} \right\}$$

where with $T^{(k)}$ we denote the composition of T with itself for k times. Following the standard theory of optimal transport (see [19, 1]), we also introduce the Kantorovich problem

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

where $\Pi(\rho)$ is the set of transport plans

$$\Pi(\rho) = \left\{ \gamma \in \mathscr{P}(\mathbb{R}^{nN}) : \pi_\sharp^i \gamma = \rho,\, i = 1, \dots, N \right\}$$

and $\pi^i : (\mathbb{R}^n)^N \to \mathbb{R}^n$ are the projections on the i-th component for $i = 1, \dots, N$. To every $(N-1)$-uple of transport maps $T_2, \dots, T_N \in \mathcal{T}(\rho)$ we canonically associate the transport plan $\gamma = (Id, T_2, \dots, T_N)_\sharp \rho \in \Pi(\rho)$. As proved in [4], if ρ is non-atomic the values of the minimum problems coincide

$$(K) = (M) = (M_{\mathrm{cycl}}).$$

Existence of optimal transport plans in (K) follows from a standard compactness and lower semicontinuity argument. In turn, existence of optimal maps in (M) is largely open; it is understood only with $N = 2$ marginals in any dimension n and in dimension $n = 1$ with any number N of marginals (see [6] and [5] respectively). In a different context, optimal cyclical maps as in (M_{cycl}) appear in [11] for some particular costs generated by vector fields.

As regards uniqueness of optimal symmetric plans with Coulomb cost, it holds in dimension 1, but, as shown in [15], it fails in the same class already when we consider spherically symmetric densities in $\mathbb{R}^2$, for any N. On the other hand, the Kantorovich duality works also for this cost (see [16]) and the dual problem admits maximizers

(namely, Kantorovich's potentials), as shown by De Pascale [8]; moreover, in [7] the limit of symmetric optimal plans as $N \to \infty$ is shown to be the infinite product measure of ρ with itself.

Beyond the 1-dimensional case, which is well understood, a physically relevant case is given by spherically symmetric densities ρ in $\mathbb{R}^n$, with any number of marginals. In the physics literature, they appear in [17, 18] to study simple atoms like Helium ($N = 2$), Litium ($N = 3$), and Berillium ($N = 4$). In this case the problem reduces, thanks to the spherical symmetry, to a problem in 1-dimension, with a more complicated cost function (see [15], where this reduction is rigorously described). In the class of admissible transport maps for problem (M_{cycl}), Seidl, Gori Giorgi and Savin identified some particularly simple maps: roughly speaking, they divide $\mathbb{R}^n$ in N spherical shells, each containing one electron in average, and consider the transport maps which send each shell onto the next one by a monotonically increasing or decreasing map. They conjecture the optimality of one of these maps in (M_{cycl}).

In the following, we provide counterexamples to the conjecture showing that there are cases in which none of these maps are optimal in problem (M_{cycl}). On the other hand, we also point out situations where some of these maps satisfy optimality conditions, namely c-monotonicity. We deal for simplicity with radial measures in $\mathbb{R}^2$ with 3 marginals, although similar examples and computations can be carried out in any dimension and with any number of marginals. The result applies in the to the physically relevant cases described above and shows that the optimal maps for radially symmetric data exibit a richer structure than the one depicted in [17, 18]. In [10], currently in preparation, further numerical examples and considerations will be provided on the same topic.

The plan of the paper is the following. In Section 2 we present the problem with spherically symmetric data, we recall the notion of c-monotonicity and a few properties of optimal transport maps, and we give some examples and counterexamples. In Sections 3 and 4 we study the properties of the cost for close radii and for spread apart radii, respectively. In Section 5 we apply these properties to give rigorous proofs of the examples and counterexamples.

2 Examples and counterexamples

2.1 Monge and Kantorovich problems with radial densities

As we mentioned above, the transport problem (1.1) reduces to a 1-dimensional one (i.e., by proving that spheres get mapped to spheres), as rigorously done in [15]. Assuming from now on $N = 3$ and denoting $(0, \infty)$ by $\mathbb{R}_+$, given three radii $r_1, r_2, r_3 \in \mathbb{R}_+$, we consider the associated *exact cost* (see Figure 1)

$$c(r_1, r_2, r_3) = \min \left\{ \frac{1}{|v_2 - v_1|} + \frac{1}{|v_3 - v_2|} + \frac{1}{|v_1 - v_3|} : |v_i| = r_i, \ i = 1, 2, 3 \right\}, \quad (2.1)$$
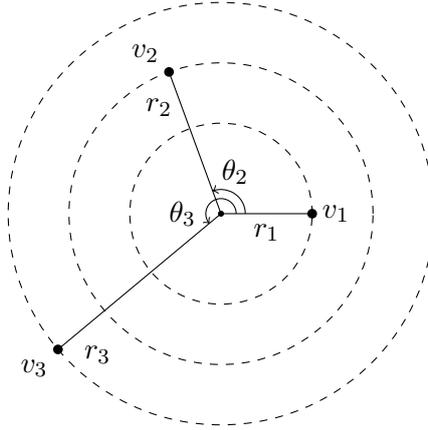
Figure 1: A configuration of three charges at distances r_1, r_2 and r_3 with angles θ_2 and θ_3.

which is a positive, symmetric, continuous function. Given a non-atomic probability measure $\rho \in \mathscr{P}(\mathbb{R}_+)$, the set of transport maps reads as

$$\mathcal{T}(\rho) = \{\, T : \mathbb{R}_+ \to \mathbb{R}_+ \text{ Borel} : T_\sharp \rho = \rho \,\},$$

and the cyclical Monge problem corresponding to (1.1) can be written as

$$(M_{\mathrm{cycl}}) = \inf \left\{ \int_{\mathbb{R}_+} c(x, T(x), T^{(2)}(x)) \, d\rho(x) : T \in \mathcal{T}(\rho),\, T^{(3)} = \mathrm{Id} \right\}. \tag{2.2}$$

We also introduce the set of transport plans

$$\Pi(\rho) = \{\, \gamma \in \mathscr{P}(\mathbb{R}_+^3) : \pi_\sharp^i \gamma = \rho,\, i = 1, 2, 3 \,\},$$

where $\pi^i : (\mathbb{R}_+)^3 \to \mathbb{R}_+$ are the projections on the i-th component for $i = 1, 2, 3$, and the Kantorovich multimarginal problem

$$(K) = \min \left\{ \int_{(\mathbb{R}_+)^3} c(r_1, r_2, r_3) \, d\gamma(r_1, r_2, r_3) : \gamma \in \Pi(\rho) \right\}. \tag{2.3}$$

2.2 Some special maps

In the following definition, we introduce some special transport maps, which were conjectured in [18] to be good candidates for optimality in problem (2.2).

Definition 2.1. Let $\rho \in \mathscr{M}(\mathbb{R}_+)$ be a non-atomic probability measure and let $d_1, d_2 \in \mathbb{R}_+$ such that $\rho([0, d_1]) = \rho([d_1, d_2]) = \rho([d_2, \infty]) = 1/3$. The *DDI*-map $T : \mathbb{R}_+ \to \mathbb{R}_+$ associated to ρ is the unique (up to ρ-negligible sets) map such that $T_\sharp \rho = \rho$ and

- T maps $(0, d_1)$ onto (d_1, d_2) decreasingly,

- T maps (d_1, d_2) onto (d_2, ∞) decreasingly,

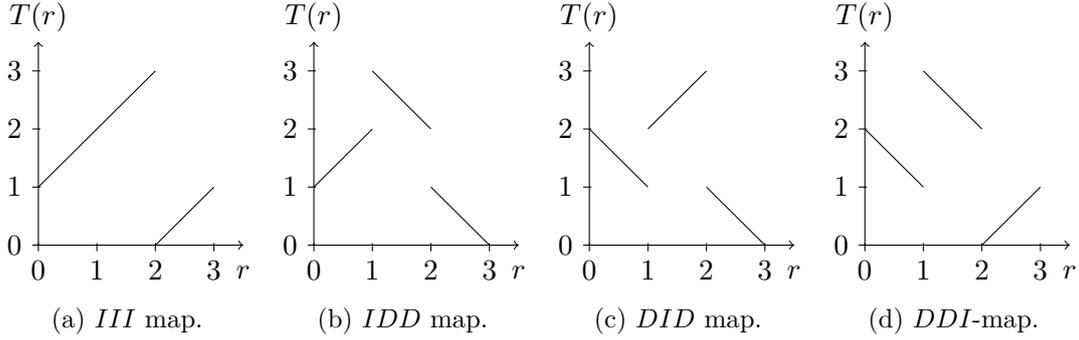- T maps (d_2, ∞) onto $(0, d_1)$ increasingly.

5

| (a) *III* map. | (b) *IDD* map. | (c) *DID* map. | (d) *DDI*-map. |

Figure 2: The four types of maps considered in the conjecture in the case of a uniform density on $[0, 3]$.

Similarly, we define, for instance, the *DID*-map mapping $(0, d_1)$ onto (d_1, d_2) decreasingly, (d_1, d_2) onto (d_2, ∞) increasingly and (d_2, ∞) onto $(0, d_1)$ decreasingly.

The $\{D, I\}^3$-class associated to ρ is composed by the maps with all the possible monotonicities, under the condition that $T^{(3)} = \mathrm{Id}$: therefore we have *III*, *IDD*, *DID* and *DDI*, (see Figure 2).

In the rest of the paper we answer the following question:

Question 2.2. Is the *DDI*-map associated to ρ optimal in problem (2.2) for every measure $\rho \in \mathscr{P}(\mathbb{R}_+)$? Is one of the maps in $\{D, I\}^3$-class associated to ρ optimal in problem (2.2) for every non-atomic probability measure $\rho \in \mathscr{P}(\mathbb{R}_+)$?

2.3 A necessary condition for optimality: *c*-monotonicity

Before presenting the examples and counterexamples, we recall a well-known optimality condition in optimal transport.

Definition 2.3. Let $c : (\mathbb{R}_+)^N \to [0, \infty]$ be a cost function. We say that a set $\Gamma \subset (\mathbb{R}_+)^N$ is *c*-monotone with respect to $p \subseteq \{1, \ldots, N\}$ if

$$c(x) + c(y) \leq c(X(x, y, p)) + c(Y(x, y, p)) \qquad \forall x, y \in \Gamma, \qquad (2.4)$$

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

$$X_i(x, y, p) = \begin{cases} x_i & \text{if } i \in p \\ y_i & \text{if } i \notin p \end{cases} \qquad Y_i(x, y, p) = \begin{cases} y_i & \text{if } i \in p \\ x_i & \text{if } i \notin p \end{cases} \qquad \forall i \in \{1, ..., N\}. \tag{2.5}$$

We say that $\Gamma \subset (\mathbb{R}_+)^N$ is *c*-monotone if (2.4) holds true for every $p \subseteq \{1, \ldots, N\}$.

Let $\gamma \in \Pi(\rho)$ be a transport plan. The following Proposition ([14, Lemma 2], see also [5, Proposition 2.2], where the result is used to describe optimal maps with Coulomb cost in 1 dimension) presents a necessary condition for optimality of γ.

Proposition 2.4. *Let $c : (\mathbb{R}_+)^3 \to [0, \infty]$ be a continuous cost and let ρ be a probability measure on $(\mathbb{R}_+)$. Let $\gamma \in \Pi(\rho)$ be an optimal transport plan for problem (2.3) and assume $(K) < \infty$ (therefore γ has finite cost). Then $\mathrm{supp}\,\gamma$ is c-monotone.*

Remark 2.5. Given an optimal plan γ, the support of γ is c-monotone even in a stronger sense than the one in Definition 2.3. More precisely, given two points x and y (for simplicity, assume that all their coordinates are distinct to avoid multiplicity issues), we have that

$$c(x) + c(y) \leq c(X) + c(Y) \tag{2.6}$$

for every choice of $X, Y \in (\mathbb{R}_+)^N$ such that the union of the coordinates of X and Y is the same as the union of the coordinates of x and y. Indeed, given any permutation σ of the coordinates of $(\mathbb{R}_+)^N$, we have that $\sigma(y)$ is in the support of the symmetrization of γ, which is still optimal because of the symmetry of the optimal plan. Hence, applying Proposition 2.4 to x and $\sigma(y)$, we obtain (2.6) for any X and Y.

2.4 Counterexamples

The first example shows that the *DDI*-map is not always optimal in problem (2.2), by taking as marginal a measure which is concentrated in a small neighborhood of the unit sphere.

Counterexample 2.6. *There exists $\varepsilon > 0$ such that, setting*

$$\rho_\varepsilon = \frac{1}{12\varepsilon} 1_{[1,1+12\varepsilon]} \, dr \in \mathscr{M}(\mathbb{R}_+),$$

the DDI-map associated to ρ_ε is not c-monotone and, therefore, not optimal in problem (2.2).

The proof is based on the analysis of c-monotonicity for similar radii, obtained by Taylor expanding the cost around the point $(1, 1, 1)$.

The next example modifies the previous one by sending $1/6$ of the total mass far away; in this way, the cost of the orbits of these points (which have two coordinates close to 1 and one large coordinate) can be easily computed. Thanks to this property, we can show that none of the maps in the $\{D, I\}^3$-class can be optimal, since their support is not c-monotone.

**Counterexample 2.7.** *There exist $M, \varepsilon > 0$ such that, setting*

$$\rho_{M,\varepsilon} = \Big(\frac{1}{6\varepsilon} 1_{[1,1+5\varepsilon]} + \frac{1}{6} 1_{[M,M+1]} \Big) \, dr \in \mathscr{M}(\mathbb{R}_+),$$

none of the maps in the $\{D, I\}^3$-class associated to $\rho_{M,\varepsilon}$ is optimal in problem (2.2).

Remark 2.8. In Remark 5.2 we will see a similar result for the problem with 4 marginals. However, we preferred to restrict the presentation to the case with 3 marginals since the ideas involved are the same, but the computations are easier.

There are particular measures ρ for which the *DDI*-map is c-monotone (whereas this property fails in Counterexample 2.6 and 2.7). For this reason one may expect that this map is also optimal in problem (2.2), but, to show this, sufficient conditions for optimality (stronger than c-monotonicity) would have to be identified.

Proposition 2.9 (Examples of c-monotone *DDI*-maps)**.** *There exists $M > 0$ such that for any probability measure ρ such that $\rho([1,2]) = \rho([3,4]) = \rho([M, \infty)) = 1/3$ the DDI-map is c-monotone (according to Definition 2.3).*
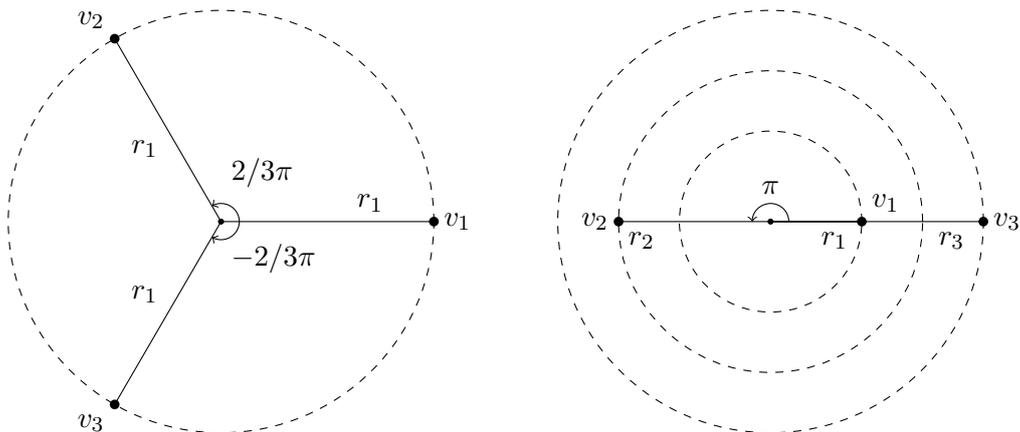
(a) A configuration of three charges at the same distance r_1 from the origin with angles $\theta_2 = 2/3\pi$ and $\theta_3 = -2/3\pi$.

(b) A configuration of three charges at distances r_1, r_2 and r_3 with angles $\theta_2 = \pi$ and $\theta_3 = 0$.

3 Taylor expansion of the cost at $r_1 = r_2 = r_3 = 1$

In this section we want to address the following problem: given three radii $r_1(t)$, $r_2(t)$ and $r_3(t)$ parametrized by $t \in \mathbb{R}$ and starting from the value 1 at $t = 0$, what is the expansion of $c(r_1(t), r_2(t), r_2(t))$ in powers of t at $t = 0$?

First, we notice that at $t = 0$ the optimal angles are $\pm 2/3\pi$ and $c(1, 1, 1) = \sqrt{3}$. Indeed, given three unitary vectors v_1, v_2, v_3, calling α_1 the angle between v_2 and v_3 (and cyclical), we have that $|v_1 - v_2| = 2\sin(\alpha_3/2)$ (and cyclical), therefore, by Jensen's inequality and by the convexity of $\alpha \mapsto [2\sin(\alpha/2)]^{-1}$ in $[0, 2\pi]$,

$$\frac{1}{|v_2 - v_1|} + \frac{1}{|v_3 - v_2|} + \frac{1}{|v_1 - v_3|} = \sum_{i=1}^{3} \frac{1}{2\sin(\alpha_i/2)} \geq \frac{3}{2\sin((\alpha_1 + \alpha_2 + \alpha_3)/6)} = \sqrt{3},$$

with equality if and only if the triangle is equilateral.

Taking the angles to be exactly $\pm 2/3\pi$ leads to the following cost

$$c_\triangle(r_1, r_2, r_3) := \frac{1}{\sqrt{r_1^2 + r_1 r_2 + r_2^2}} + \frac{1}{\sqrt{r_2^2 + r_2 r_3 + r_3^2}} + \frac{1}{\sqrt{r_1^2 + r_1 r_3 + r_3^2}} \geq c(r_1, r_2, r_3).$$

$$\tag{3.1}$$

However the inequality is strict as soon as the three radii are different and the approximation of c with $c_\triangle$ is too rough to deduce that they enjoy the same c-monotonicity structures. Therefore, we perform a finer analysis.

We want to take into account only the first order variation of the radii as functions of t, so it is natural to consider three linearly varying radii

$$r_1(t) = 1 + a_1 t, \qquad r_2(t) = 1 + a_2 t, \qquad r_3(t) = 1 + a_3 t$$

where $a_1, a_2, a_3 \in \mathbb{R}$ are some constants. To these radii we associate the exact cost

$$g(a_1, a_2, a_3, t) = c(1 + a_1 t, 1 + a_2 t, 1 + a_3 t), \tag{3.2}$$

and we study the expansion of this function near $t = 0$.

Lemma 3.1. *Let $a_1, a_2, a_3 \in \mathbb{R}$ and let g be as in (3.2). Then we have that*

$$g(a, b, c, 0) = \sqrt{3}.$$

$$\frac{\partial g}{\partial t}(a_1, a_2, a_3, 0) = -\frac{a_1 + a_2 + a_3}{\sqrt{3}},$$

$$\frac{\partial^2 g}{\partial t^2}(a_1, a_2, a_3, 0) = \frac{4(a_1^2 + a_2^2 + a_3^2) + 6(a_1 a_2 + a_2 a_3 + a_3 a_1)}{5\sqrt{3}},$$

$$\frac{\partial^3 g}{\partial t^3}(a_1, a_2, a_3, 0) = -\frac{308(a_1^3 + a_2^3 + a_3^3)}{375\sqrt{3}}$$
$$-\frac{888(a_1^2 a_2 + a_1 a_2^2 + a_2^2 a_3 + a_2 a_3^2 + a_3^2 a_1 + a_3 a_1^2) + 498 a_1 a_2 a_3}{375\sqrt{3}}.$$
$$\tag{3.3}$$

In the proof, we will write the Coulomb potential of three charges in terms of the distances from the origin and the angles between the charges. Given three radii r_1, r_2, r_3 and two angles θ_2 and θ_3, we define the *Coulomb potential* of the configuration of charges depicted in Figure 1:

$$C(r_1, r_2, r_3, \theta_2, \theta_3) = \frac{1}{|v_2 - v_1|} + \frac{1}{|v_3 - v_2|} + \frac{1}{|v_1 - v_3|} \tag{3.4}$$

where

$$v_1 = (r_1, 0), \qquad v_2 = r_2(\cos\theta_2, \sin\theta_2), \qquad v_3 = r_3(\cos\theta_3, \sin\theta_3).$$

By definition of c, we notice that

$$c(r_1, r_2, r_3) = \min_{\theta_2, \theta_3 \in \mathbb{R}} C(r_1, r_2, r_3, \theta_2, \theta_3). \tag{3.5}$$

Proof of Lemma 3.1. For $t \in \mathbb{R}$ and $\theta = (\theta_2, \theta_3) \in \mathbb{R}^2$ we define also the function

$$G(t, \theta) = C(1 + a_1 t, 1 + a_2 t, 1 + a_3 t, \theta_2, \theta_3).$$

Then $g(t) = G(t, \theta_0(t))$ where $\theta_0(t)$ is the pair of angles which minimizes (3.5). From this optimality condition we know that

$$G_\theta(t, \theta_0(t)) = 0.$$

We want to apply the implicit function theorem to find the behavior of $\theta_0(t)$. It's easy to check that $\theta_0(0) = (2/3\pi, -2/3\pi)$ and a direct computation shows that

$$G_{\theta\theta}(0, \theta_0(0)) = \frac{5}{6\sqrt{3}} \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix} \in \mathrm{Inv}(\mathbb{R}^2; \mathbb{R}^2).$$

Therefore $\theta_0 \in C^\infty((-\varepsilon, \varepsilon))$ for some $\varepsilon > 0$ and we can compute its derivatives in 0. In particular, we have that

$$\theta_0'(0) = G_{\theta\theta}^{-1} \cdot G_{t\theta}\Big|_{(0, \theta_0(0))} = \frac{1}{5\sqrt{3}} \begin{pmatrix} -a_1 - a_2 + 2a_3 \\ a_1 - 2a_2 + a_3 \end{pmatrix}. \tag{3.6}$$

The idea is now to consider the first order approximation

$$\bar{\theta}(t) = \theta_0(0) + \theta_0'(0)t = \begin{pmatrix} 2/3\pi \\ -2/3\pi \end{pmatrix} + \frac{1}{5\sqrt{3}} \begin{pmatrix} -a_1 - a_2 + 2a_3 \\ a_1 - 2a_2 + a_3 \end{pmatrix} t$$

and the perturbed cost

$$h(t) = G(t, \bar{\theta}(t)).$$

We claim that $h(t) = g(t) + o(t^3)$, namely

$$h(0) = g(0), \qquad h'(0) = g'(0), \qquad h''(0) = g''(0), \qquad h'''(0) = g'''(0).$$

The first two are clearly true, since $\bar{\theta}(0) = \theta_0(0)$ and $\bar{\theta}'(0) = \theta_0'(0)$ by definition. Now consider the function $t \mapsto G(t, \theta(t))$, where θ is either θ_0 or $\bar{\theta}$. To prove the claim, we show that its second and third derivatives at $t = 0$ depend only on $\theta'(0)$ and not on the second and third derivatives of θ.

As a matter of fact, we have

$$\left. \frac{d^2 G(t, \theta(t))}{dt^2} \right|_{t=0} = G_{tt} + 2G_{t\theta}\theta' + G_{\theta\theta}\theta'\theta' + G_\theta\theta'' \Big|_{t=0},$$

but $G_\theta(0, \theta(0)) = 0$, so the second derivative does not depend on $\theta''(0)$. In a similar fashion, we have

$$\left. \frac{d^3 G(t, \theta(t))}{dt^3} \right|_{t=0} = G_{ttt} + 3G_{tt\theta}\theta' + 3G_{t\theta\theta}(\theta')^2 + G_{\theta\theta\theta}(\theta')^3 + 3\left(G_{t\theta} + G_{\theta\theta}\theta'\right)\theta'' + G_\theta\theta''' \Big|_{t=0}.$$

Again, $G_\theta(0, \theta(0)) = 0$, therefore $\theta'''(0)$ doesn't contribute. Furthermore, we have $G_\theta(t, \theta_0(t)) = 0$, so that differentiating in t yields

$$G_{t\theta}(0, \theta_0(0)) + G_{\theta\theta}(0, \theta_0(0))\theta_0'(0) = 0.$$

But then also

$$G_{t\theta}(0, \bar{\theta}(0)) + G_{\theta\theta}(0, \bar{\theta}(0))\bar{\theta}'(0) = 0,$$

since $\bar{\theta}'(0) = \theta_0'(0)$. Therefore we see that in both cases the coefficient of θ'' vanishes. This concludes the proof of the claim because we have shown that the first three derivatives of h and g coincide at $t = 0$.

At this point the derivatives of h can be computed directly, since $h(a_1, a_2, a_3, \cdot)$ is an explicit function of the last variable. $\qquad \square$

In Lemma 3.1 we found the first nontrivial Taylor term in the expansion of $g(t)$. We employ this computation to obtain information on the c-monotonicity of points with linearly spaced radii close to $t = 0$.

Lemma 3.2. *For every $t > 0$, consider six linearly spaced radii*

$$(r_1, r_2, r_3, r_4, r_5, r_6) = (1, 1 + t, 1 + 2t, 1 + 3t, 1 + 4t, 1 + 5t). \tag{3.7}$$

Then there exists $t_0 > 0$ such that, for every $t \leq t_0$,

$$c(r_1, r_4, r_6) + c(r_2, r_3, r_5) < c(r_1, r_4, r_5) + c(r_2, r_3, r_6).$$

Proof. Let us define

$$F(t) = g(0, 3, 5, t) + g(1, 2, 4, t) - g(0, 3, 4, t) - g(1, 2, 5, t)$$

Applying Lemma 3.1 we can compute the derivatives of F and find that

$$F(0) = 0, \qquad F'(0) = 0, \qquad F''(0) = 0, \qquad F'''(0) = -\frac{284\sqrt{3}}{125} < 0;$$

this shows that $F(t) < 0$ for t sufficiently small and proves the lemma. $\qquad\square$

Remark 3.3. Considering $r_1, ..., r_6$ as in (3.7), one could prove that the choice 146-235 is optimal between all possible choices, namely

$$c(r_1, r_4, r_6) + c(r_2, r_3, r_5)$$
$$= \min\{ c(p_1, p_2, p_3) + c(p_4, p_5, p_6) : \{p_1, \ldots, p_6\} = \{r_1, \ldots, r_6\} \}, \quad (3.8)$$

for t small enough. Moreover, one could see that (3.8) holds also if we replace c with $c_\triangle$ defined in (3.1). This is, however, not needed for our counterexamples.

Remark 3.4 (Asymptotic expansion of the cost at infinity). Although they will not be used in the proofs of the main results, we report the following formulas since they might help in future studies to gain more insight into the structure of c-monotone sets. We are interested in the asymptotic expansion of the cost as some of the radii go to infinity and the others remain fixed.

For $(r_1, r_2, r_3) = (1, 1, r)$, the optimal angles are

$$\theta_2(r) = \pi - \frac{8}{r^2} + o\left(\frac{1}{r^3}\right), \qquad \theta_3(r) = -\frac{\pi}{2} - \frac{4}{r^2} + o\left(\frac{1}{r^3}\right).$$

In comparison to (3.6), this expansion is harder to justify (but can be easily verified numerically). However, from this fact it follows rigorously that the cost has the following asymptotic behaviour:

$$c(1, 1, r) = C(1, 1, r, \pi, -\pi/2) - \frac{4}{r^4} + o\left(\frac{1}{r^4}\right)$$
$$= \left(\frac{1}{2} + \frac{1}{\sqrt{1 + r^2}}\right) - \frac{4}{r^4} + o\left(\frac{1}{r^4}\right).$$

Similarly, for $(r_1, r_2, r_3) = (1, r, r)$, the optimal angles are

$$\theta_2(r) = \frac{\pi}{2} + \frac{4}{r} + o\left(\frac{1}{r^2}\right), \qquad \theta_3(r) = -\frac{\pi}{2} - \frac{4}{r} + o\left(\frac{1}{r^2}\right),$$

and the cost is

$$c(1, r, r) = C(1, r, r, \pi/2, -\pi/2) - \frac{4}{r^3} + o\left(\frac{1}{r^4}\right)$$
$$= \frac{1}{2r} + \frac{2}{\sqrt{1 + r^2}} - \frac{4}{r^3} + o\left(\frac{1}{r^4}\right).$$

Furthermore, one can verify that

$$c(1, r, r) = C\left(1, r, r, \frac{\pi}{2} + \frac{4}{r}, -\frac{\pi}{2} - \frac{4}{r}\right) - O\left(\frac{1}{r^7}\right).$$

4   Condition for $c = c_\pi$ and c_π-monotonicity

When the radii are spread apart, a reasonable approximate cost appears to be

$$c_\pi(r_1, r_2, r_3) = \frac{1}{r_1 + r_2} + \frac{1}{r_2 + r_3} + \frac{1}{r_3 - r_1},$$

which arises from collocating the charges at angles $\theta_2 = \pi$ and $\theta_3 = 0$ (see Figure 3b). In the first part of this section we want to study under which condition on the radii r_1, r_2 and r_3 we have

$$c(r_1, r_2, r_3) = c_\pi(r_1, r_2, r_3).$$

We start with a heuristic argument involving a necessary condition. Up to permutations, we may assume $r_1 \leq r_2 \leq r_3$. It is simple to check that

$$C_\theta(r_1, r_2, r_3, \pi, 0) = 0,$$

where C has been defined in (3.4), either by direct computation or by a symmetry argument.[1] If $(\theta_2, \theta_3) = (\pi, 0)$ must be a minimum, then a necessary condition is

$$C_{\theta\theta}(r_1, r_2, r_3, \pi, 0) \geq 0,$$

in the sense that the Hessian matrix is positive-definite. We have

$$C_{\theta\theta}(r_1, r_2, r_3, \pi, 0) = \begin{pmatrix} r_2 \left(\frac{r_1}{(r_1+r_2)^3} + \frac{r_3}{(r_2+r_3)^3} \right) & -\frac{r_2 r_3}{(r_2+r_3)^3} \\ -\frac{r_2 r_3}{(r_2+r_3)^3} & r_3 \left(\frac{r_2}{(r_2+r_3)^3} - \frac{r_1}{(r_3-r_1)^3} \right) \end{pmatrix}$$

since the first entry is positive, this 2×2 matrix is positive-definite if and only if the determinant is positive too, namely

$$\det C_{\theta\theta}(r_1, r_2, r_3, \pi, 0) = -\frac{r_1 r_2 r_3 [r_2 r_3 (r_2 - r_3) + r_1 (r_2^2 + 5 r_2 r_3 + r_3^2) + r_1^3]}{(r_1 + r_2)^3 (r_2 + r_3)^2 (r_3 - r_1)^3} \geq 0,$$

or equivalently

$$r_1(r_2^2 + 5 r_2 r_3 + r_3^2) + r_1^3 < r_2 r_3 (r_3 - r_2).$$

Figure 4 depicts the region where the Hessian is positive.

We partially justify the previous argument in the following lemma which, despite not being quantitative, will suffice for our purposes.

Lemma 4.1. *If $0 < r_1^- \leq r_1^+ < r_2^- \leq r_2^+$, then there exists $r_3^-(r_1^-, r_1^+, r_2^-, r_2^+)$ such that for every $r_1 \in [r_1^-, r_1^+]$, $r_2 \in [r_2^-, r_2^+]$ and $r_3 \geq r_3^-$ we have*

$$c(r_1, r_2, r_3) = c_\pi(r_1, r_2, r_3).$$

Proof. We denote by $\mathbb{T}^2$ the 2-dimensional torus $\mathbb{R}^2 / (2\pi\mathbb{Z})^2$. The idea of the proof is the following: we claim that for sufficiently large r_3 there are exactly four stationary points $(\theta_2, \theta_3) \in \mathbb{T}^2$ for $C(r_1, r_2, r_3, \theta_2, \theta_3)$, corresponding to $\theta_2, \theta_3 \in \{0, \pi\}$. Therefore $c(r_1, r_2, r_3)$ must coincide with the value achieved at one of them and by comparing the four values we arrive at the desired conclusion.

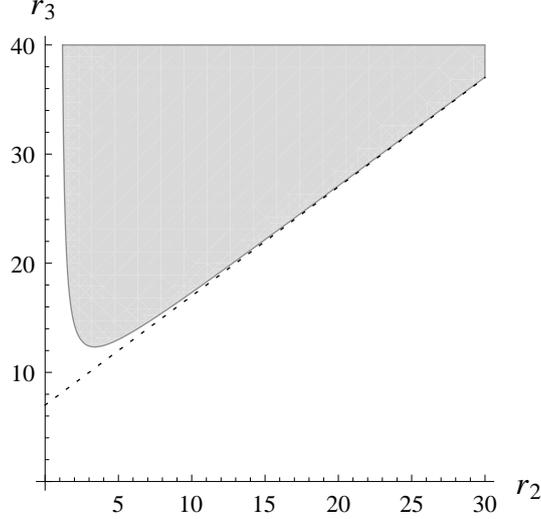[1]In fact, the four configurations with $\theta_2, \theta_3 \in \{0, \pi\}$ are always stationary.

Figure 4: The region in the (r_2, r_3) plane where $C_{\theta\theta}(r_1, r_2, r_3, \pi, 0) \geq 0$, with $r_1 = 1$. The dotted line is $r_3 = r_2 + 7$.

First of all, we compute the gradient

$$C_\theta(r_1, r_2, r_3, \theta_2, \theta_3) = \begin{pmatrix} -\dfrac{r_1 r_2 \sin(\theta_2)}{\left(r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta_2)\right)^{3/2}} - \dfrac{r_2 r_3 \sin(\theta_2 - \theta_3)}{\left(r_2^2 + r_3^2 - 2r_2 r_3 \cos(\theta_2 - \theta_3)\right)^{3/2}} \\ -\dfrac{r_1 r_3 \sin(\theta_3)}{\left(r_1^2 + r_3^2 - 2r_1 r_3 \cos(\theta_3)\right)^{3/2}} + \dfrac{r_2 r_3 \sin(\theta_2 - \theta_3)}{\left(r_2^2 + r_3^2 - 2r_2 r_3 \cos(\theta_2 - \theta_3)\right)^{3/2}} \end{pmatrix}.$$

The gradient vanishes if and only if the following equations are simultaneously satisfied:

$$\frac{r_1 r_2 \sin(\theta_2)}{\left(r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta_2)\right)^{3/2}} + \frac{r_1 r_3 \sin(\theta_3)}{\left(r_1^2 + r_3^2 - 2r_1 r_3 \cos(\theta_3)\right)^{3/2}} = 0, \qquad (4.1\text{a})$$

$$-\frac{r_1 r_3 \sin(\theta_3)}{\left(r_1^2 + r_3^2 - 2r_1 r_3 \cos(\theta_3)\right)^{3/2}} + \frac{r_2 r_3 \sin(\theta_2 - \theta_3)}{\left(r_2^2 + r_3^2 - 2r_2 r_3 \cos(\theta_2 - \theta_3)\right)^{3/2}} = 0. \qquad (4.1\text{b})$$

To show that there are exactly four stationary points, the idea is that, for r_3 sufficiently large, equations (4.1a) and (4.1b) define two pairs of closed curves on $\mathbb{T}^2$, of type $(0, 1)$ and $(1, 1)$ respectively, with the property that every curve from the first family intersects each curve of the second family in a single point. The situation is represented in Figure 5.

Step 1. Given r_1, r_2 and a sufficiently large r_3, we claim that for every $\theta_3 \in S^1$ there are exactly two values $\tilde{\theta}_2^0(\theta_3), \tilde{\theta}_2^\pi(\theta_3) \in S^1$ which satisfy (4.1a); moreover $\tilde{\theta}_2^0(\theta_3)$ and $\tilde{\theta}_2^\pi(\theta_3)$ are close to 0 and π respectively by less than $O(r_3^{-2})$, uniformly in θ_3, and their derivatives go to to zero uniformly in θ_3 for $r_3 \to \infty$.[2] These functions correspond to the solid, almost vertical, lines in Figure 5.

We begin by finding a useful bound on $|\sin(\theta_2)|$. The two terms of (4.1a) can be

---

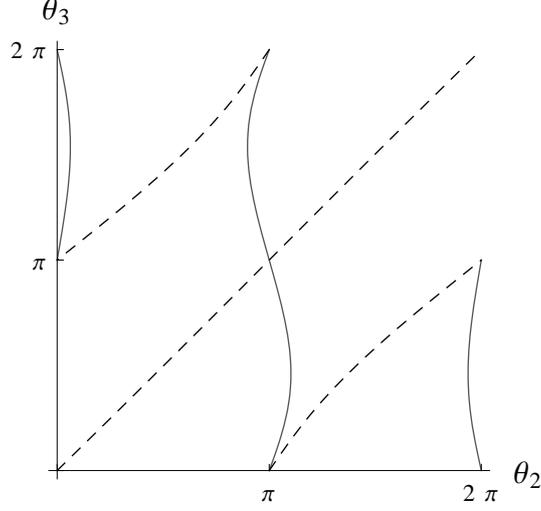[2]More precisely, they are close to zero by less than $O\left(r_3^{-2}\right)$, uniformly in θ_3.

Figure 5: The curves in $\mathbb{T}^2$ whose four intersections correspond to stationary points of $C(r_1, r_2, r_3, \theta_2, \theta_3)$. The two solid curves are defined by (4.1a). The dashed curves are defined by (4.1b).

estimated by

$$\left| \frac{r_1 r_2 \sin(\theta_2)}{\left(r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta_2)\right)^{3/2}} \right| \geq \frac{r_1^- r_2^- |\sin(\theta_2)|}{(r_1^+ + r_2^+)^3},$$

$$\left| \frac{r_1 r_3 \sin(\theta_3)}{\left(r_1^2 + r_3^2 - 2r_1 r_3 \cos(\theta_3)\right)^{3/2}} \right| \leq \frac{r_1^+ r_3}{(r_3 - r_1^+)^3},$$

therefore, in order to have equality (4.1a), it must be that

$$\frac{r_1^- r_2^- |\sin(\theta_2)|}{(r_1^+ + r_2^+)^3} \leq \frac{r_1^+ r_3}{(r_3 - r_1^+)^3},$$

that is

$$|\sin(\theta_2)| \leq \frac{r_1^+ (r_1^+ + r_2^+)^3}{r_1^- r_2^-} \cdot \frac{r_3}{(r_3 - r_1^+)^3} = O(r_3^{-2}) \tag{4.2}$$

as $r_3 \to \infty$, where the implied constant depends only on $r_1^\pm$ and $r_2^\pm$.

We have already discussed that, for every $\theta_3 \in S^1$, the second term in (4.1a) is smaller than $r_3(r_3 - r_1^+)^{-3}$ in magnitude. On the other hand, the first term vanishes for $\theta_2 = 0, \pi$ and is equal to $\pm r_1 r_2 (r_1^2 + r_2^2)^{3/2}$ for $\theta_2 = \pm \pi/2$. Therefore, by continuity, for r_3 large we have at least two solutions to (4.1a).

The estimate on $|\sin(\theta_2)|$ proves that the solutions must be located near 0 and π. Now we want to prove that there are exactly two of them. To do so, we verify that the partial derivative with respect to θ_2 of the first term in (4.1a) is different from zero for

14

θ_2 in the prescribed intervals around 0 and π. Indeed, the derivative is

$$\frac{\partial}{\partial\theta_2}\bigg|_{\theta_2=0}\left(\frac{r_1r_2\sin(\theta_2)}{\left(r_1^2+r_2^2-2r_1r_2\cos(\theta_2)\right)^{3/2}}\right)=\frac{r_1r_2}{(r_2-r_1)^3},$$

$$\frac{\partial}{\partial\theta_2}\bigg|_{\theta_2=\pi}\left(\frac{r_1r_2\sin(\theta_2)}{\left(r_1^2+r_2^2-2r_1r_2\cos(\theta_2)\right)^{3/2}}\right)=-\frac{r_2}{(r_1+r_2)^3},$$

therefore it is different from zero around the two points and the two solutions are simple.

The claim is almost entirely proved. We now have the two functions $\tilde{\theta}_2^0(\,\cdot\,),\tilde{\theta}_2^\pi(\,\cdot\,)$ and the last thing that we want to derive is the estimate of their first derivatives. Let $\theta_2(\,\cdot\,)$ be one of the two functions. Thanks to the implicit function theorem, we know that $\theta_2(\,\cdot\,)$ is at least C^1 and we can compute

$$\theta_2'(\theta_3)=-\frac{r_3}{r_2}\cdot\frac{2(r_1^2+r_3^2)\cos(\theta_3)+r_1r_3[-5+\cos(2\theta_3)]}{2(r_1^2+r_2^2)\cos(\theta_2(\theta_3))+r_1r_2[-5+\cos(2\theta_2(\theta_3))]}$$
$$\cdot\left(\frac{r_1^2+r_2^2-2r_1r_2\cos(\theta_2(\theta_3))}{r_1^2+r_3^2-2r_1r_3\cos(\theta_3)}\right)^{5/2}.$$

All the terms are fairly easy to deal with, apart from the denominator of the second fraction. However, we have that

$$2(r_1^2+r_2^2)\cos(\theta_2)+r_1r_2[-5+\cos(2\theta_2)]\big|_{\theta_2=0}=2(r_1^2-2r_1r_2+r_2^2)\geq 2(r_2^--r_1^+)^2,$$
$$-2(r_1^2+r_2^2)\cos(\theta_2)-r_1r_2[-5+\cos(2\theta_2)]\big|_{\theta_2=\pi}=2(r_1^2+2r_1r_2+r_2^2)\geq 2(r_2^-+r_1^-)^2,$$

therefore, by the continuity of the functions involved and by compactness, there exists a neighbourhood U of $\{0,\pi\}$ such that if $r_1\in[r_1^-,r_1^+]$, $r_2\in[r_2^-,r_2^+]$ and $\theta_2\in U$ then

$$\left|2(r_1^2+r_2^2)\cos(\theta_2)+r_1r_2[-5+\cos(2\theta_2)]\right|>(r_2^--r_1^+)^2.$$

From this and (4.2), which ensures that $\theta_2(\theta_3)\in U$, we deduce that for r_3 large

$$|\theta_2'(\theta_3)|\leq\frac{r_3}{r_2^-}\cdot\frac{2(r_1^+)^2+2r_3^2}{(r_2^--r_1^+)^2}\cdot\frac{(r_1^++r_2^+)^5}{(r_3-r_1^+)^5}=O\left(r_3^{-2}\right).$$

Step 2. Next we perform the same analysis for (4.1b). We prove that there exist two C^1 functions $\hat{\theta}_2^0(\theta_3)$ and $\hat{\theta}_2^\pi(\theta_3)$ which are the only solutions of (4.1b) when θ_3 is prescribed and that their derivatives are strictly positive. First of all, we introduce the new variable $\psi=\theta_2-\theta_3$. Equation (4.1b) reads as

$$-\frac{r_1\sin(\theta_3)}{\left(r_1^2+r_3^2-2r_1r_3\cos(\theta_3)\right)^{3/2}}+\frac{r_2\sin(\psi)}{\left(r_2^2+r_3^2-2r_2r_3\cos(\psi)\right)^{3/2}}=0. \qquad (4.3)$$

• **The solutions lie in two strips.** From equation (4.3) we get

$$\frac{r_1^+}{(r_3-r_1^+)^3}\geq\left|\frac{r_1\sin(\theta_3)}{\left(r_1^2+r_3^2-2r_1r_3\cos(\theta_3)\right)^{3/2}}\right|$$
$$=\left|\frac{r_2\sin(\psi)}{\left(r_2^2+r_3^2-2r_2r_3\cos(\psi)\right)^{3/2}}\right|\geq\frac{r_2^-|\sin(\psi)|}{(r_2^++r_3)^3}.$$

Therefore we have

$$|\sin(\psi)| \leq \left(\frac{r_3 + r_2^+}{r_3 - r_1^+}\right)^3 \frac{r_1^+}{r_2^-},$$

which, for r_3 sufficiently large, implies $|\sin(\psi)| < \eta$ for a fixed $\eta \in (r_1^+/r_2^-, 1)$.

- **There are at least two solutions.** The first term of (4.3) is bounded by

$$\left|\frac{r_1 \sin(\theta_3)}{(1 + r_3^2 - 2r_3 \cos(\theta_3))^{3/2}}\right| \leq \frac{r_1}{(r_3 - 1)^3}.$$

On the other hand, when $\psi = \pm\pi/2$ the second term equals

$$\pm \frac{r_2}{(r_2^2 + r_3^2)^{3/2}},$$

which is bigger for r_3 large enough. This tells us that for every θ_3 there are at least two distinct values of ψ which solve (4.3), because the second term is a continuous periodic function of ψ.

- **There are exactly two solutions.** The derivative of the second term is

$$\frac{\partial}{\partial \psi}\left(\frac{r_2 \sin(\psi)}{(r_2^2 + r_3^2 - 2r_2 r_3 \cos(\psi))^{3/2}}\right) = \frac{-3r_2^2 r_3 + (r_2^3 + r_2 r_3^2)\cos(\psi) + r_2^2 r_3 \cos(\psi)^2}{(r_2^2 + r_3^2 - 2r_2 r_3 \cos(\psi))^{5/2}}.$$

We observe that the denominator is always positive. We study the sign of the numerator. The equation

$$-3r_2^2 r_3 + (r_2^3 + r_2 r_3^2)t + r_2^2 r_3 t^2 = 0$$

for the unknown t has the two solutions

$$\frac{-r_2^2 - r_3^2 + \sqrt{r_2^4 + 14r_2^2 r_3^2 + r_3^4}}{2r_2 r_3}, \qquad \frac{-r_2^2 - r_3^2 - \sqrt{r_2^4 + 14r_2^2 r_3^2 + r_3^4}}{2r_2 r_3}.$$

However, only the first one lies in the range $[-1, 1]$, whereas the second is less than -2. In fact,

$$r_2^2 + r_3^2 + \sqrt{r_2^4 + 14r_2^2 r_3^2 + r_3^4} \geq r_2^2 + r_3^2 + \sqrt{r_2^4 + 2r_2^2 r_3^2 + r_3^4} = 2(r_2^2 + r_3^2) \geq 4r_2 r_3.$$

Therefore the function has exactly two stationary points and is monotone between them.

- **Derivative of the solutions.** At this point we know that there exist two functions $\psi_0(\theta_3)$ and $\psi_\pi(\theta_3)$ such that the corresponding $\hat{\theta}_2^0(\theta_3) = \psi_0(\theta_3) + \theta_3$ and $\hat{\theta}_2^\pi(\theta_3) = \psi_\pi(\theta_3) + \theta_3$ parametrize the solutions of (4.1b).

  The goal is to show that for r_3 sufficiently large we have $\theta_2'(\theta_3) \geq C > 0$ for some constant C independent of r_3, where $\theta_2(\cdot)$ is either $\hat{\theta}_2^0(\cdot)$ or $\hat{\theta}_2^\pi(\cdot)$. Thanks to

the implicit function theorem we can compute the derivative

$$\theta_2'(\theta_3) = \frac{(r_2^2 + r_3^2 - 2r_2r_3\cos(\psi))^{5/2}}{-3r_2^2r_3 + (r_2^3 + r_2r_3^2)\cos(\psi) + r_2^2r_3\cos(\psi)^2}$$
$$\cdot \left(\frac{r_1\cos(\theta_3)}{(r_1^2 + r_3^2 - 2r_1r_3\cos(\theta_3))^{3/2}} + \frac{r_2\cos(\psi)}{(r_2^2 + r_3^2 - 2r_2r_3\cos(\psi))^{3/2}} \right.$$
$$\left. - \frac{3r_1^2r_3\sin(\theta_3)^2}{(r_1^2 + r_3^2 - 2r_1r_3\cos(\theta_3))^{5/2}} - \frac{3r_2^2r_3\sin(\psi)^2}{(r_2^2 + r_3^2 - 2r_2r_3\cos(\psi))^{5/2}} \right),$$

where $\psi = \theta_2 - \theta_3$ as before. We introduce the parameter $\kappa = 1/r_3$ and write the derivative in terms of it. We have that

$$\theta_2'(\theta_3) = f(r_1, r_2, 1/r_3, \theta_2 - \theta_3, \theta_3)$$

where

$$f(r_1, r_2, \kappa, \psi, \theta_3) = \frac{(1 - 2r_2\kappa\cos(\psi) + r_2^2\kappa^2)^{5/2}}{-3r_2^2\kappa + (r_2^3\kappa^2 + r_2)\cos(\psi) + r_2^2\kappa\cos(\psi)^2}$$
$$\cdot \left(\frac{r_1\cos(\theta_3)}{(1 + r_1^2\kappa^2 - 2r_1\kappa\cos(\theta_3))^{3/2}} + \frac{r_2\cos(\psi)}{(1 + r_2^2\kappa^2 - 2r_2\kappa\cos(\psi))^{3/2}} \right.$$
$$\left. - \frac{3r_1^2\kappa\sin(\theta_3)^2}{(1 + r_1^2\kappa^2 - 2r_1\kappa\cos(\theta_3))^{5/2}} - \frac{3r_2^2\kappa\sin(\psi)^2}{(1 + r_2^2\kappa^2 - 2r_2\kappa\cos(\psi))^{5/2}} \right). \quad (4.4)$$

Observe that the only singularities are due to the denominator of the first fraction. However, the singular values of ψ lie outside the two intervals

$$S = [-\arcsin(\eta), \arcsin(\eta)] \cup [\pi - \arcsin(\eta), \pi + \arcsin(\eta)]$$

for κ sufficiently small (r_3 large enough), because they converge to $\pm\pi/2$. Therefore there exists $\kappa^+ > 0$ such that the function f is continuous in the domain

$$D = [r_1^-, r_1^+]_{r_1} \times [r_2^-, r_2^+]_{r_2} \times [0, \kappa^+]_\kappa \times S_\psi \times [0, 2\pi]_{\theta_3}.$$

- **Limit case.** We rewrite equation (4.3) in terms of κ as

$$-\frac{r_1\sin(\theta_3)}{(1 + r_1^2\kappa^2 - 2r_1\kappa\cos(\theta_3))^{3/2}} + \frac{r_2\sin(\psi)}{(1 + r_2^2\kappa^2 - 2r_2\kappa\cos(\psi))^{3/2}} = 0. \quad (4.5)$$

Let $\Gamma_{r_1, r_2, \kappa}$ denote the set of solutions $(\psi, \theta_3) \in S_\psi \times [0, 2\pi]_{\theta_3}$ to (4.5). By the continuity of (4.5) we know that

$$\Gamma = \bigcup_{r_1 \in [r_1^-, r_1^+]} \bigcup_{r_2 \in [r_2^-, r_2^+]} \bigcup_{\kappa \in [0, \kappa^+]} \Gamma_{r_1, r_2, \kappa} \subset D$$

is a closed set. Our ultimate goal is to show that f is positive on $\Gamma_{r_1, r_2, \kappa}$ when κ is small enough.

We start by studying the limit case $\kappa = 0$. The limit curve $\Gamma_{r_1,r_2,0}$ is given by the equation

$$r_1 \sin(\theta_3) = r_2 \sin(\psi). \tag{4.6}$$

For $\kappa = 0$, the function f equals

$$f(r_1, r_2, 0, \psi, \theta_3) = \frac{1}{r_2 \cos(\psi)} (r_1 \cos(\theta_3) + r_2 \cos(\psi)) = 1 + \frac{r_1 \cos(\theta_3)}{r_2 \cos(\psi)}.$$

We claim that this function is positive on the curve defined by (4.6). Indeed, positivity is guaranteed if we are able to prove that

$$\left| \frac{r_1 \cos(\theta_3)}{r_2 \cos(\psi)} \right| < 1.$$

But, by squaring, this is equivalent to

$$r_1^2 \cos(\theta_3)^2 < r_2 \cos(\psi)^2,$$

which, thanks to (4.6), reduces to the true inequality $r_1^2 < r_2^2$.

- **Conclusion.** Finally, we prove that $f \geq C > 0$ on $\Gamma_{r_1,r_2,\kappa}$ for κ close to zero, where C is a constant depending only on $r_1^{\pm}$ and $r_2^{\pm}$.

  We know that f is positive on the compact set

  $$K = \bigcup_{r_1 \in [r_1^-, r_1^+]} \bigcup_{r_2 \in [r_2^-, r_2^+]} \Gamma_{r_1,r_2,0}.$$

  Therefore there exists a positive constant C and an open neighbourhood U of K in D such that $f > C$ on U. Since Γ is closed, a compactness argument shows that $\Gamma_{r_1,r_2,\kappa} \subset U$ for κ close to zero and this concludes the proof.

Step 3. The previous steps tell us that (4.1a) defines two vertical curves and (4.1b) two diagonal curves. The estimates on the derivatives of such curves prove that the intersections are simple, therefore there are exactly four stationary points. But we already know four stationary points, namely

$$(\theta_2, \theta_3) = (0,0), \ (0,\pi), \ (\pi,0), \ (\pi,\pi).$$

To conclude, we can just compare the costs associated to each of them and pick the smallest one. It is easy to see that $(\theta_2, \theta_3) = (\pi, 0)$ is the optimal choice. In fact, $(0,0)$ is clearly the worst. Among the three cases left, we can say that $(\pi, 0)$ always beats (π, π), that is

$$C(r_1, r_2, r_3, \pi, \pi) - C(r_1, r_2, r_3, \pi, 0)$$
$$= \left(\frac{1}{r_3 - r_2} - \frac{1}{r_3 - r_1} \right) + \left(\frac{1}{r_2 + r_1} - \frac{1}{r_3 + r_2} \right) > 0,$$

as both the differences in parenthesis are positive. Finally, $(\pi, 0)$ beats $(0, \pi)$ too because

$$C(r_1, r_2, r_3, 0, \pi) - C(r_1, r_2, r_3, \pi, 0) = \frac{2r_1(r_3^2 - r_2^2)}{(r_2^2 - r_1^2)(r_3^2 - r_1^2)} > 0. \qquad \square$$

In the following lemma, we prove that, with the frozen cost c_π, given six increasing radii numbered $1, \ldots, 6$ the choice of two disjoint subsets of three elements which minimizes the cost is always given by 145 and 236. Actually, we prove only some comparisons that are enough for our examples, but one could show in general that

$$c_\pi(r_1, r_4, r_5) + c_\pi(r_2, r_3, r_6) =$$
$$= \min \left\{ c_\pi(p_1, p_2, p_3) + c_\pi(p_4, p_5, p_6) : \{p_1, \ldots, p_6\} = \{r_1, \ldots, r_6\} \right\}.$$

The proof of this fact reduces to the characterization of c-monotonicity with Coulomb cost performed in [5, Proposition 2.4].

Lemma 4.2. *Let $0 < r_1 < \cdots < r_6$. Then we have that*

$$c_\pi(r_1, r_4, r_5) + c_\pi(r_2, r_3, r_6) \leq \min \{c_\pi(r_1, r_4, r_6) + c_\pi(r_2, r_3, r_5),$$
$$c_\pi(r_1, r_3, r_6) + c_\pi(r_2, r_4, r_5), \ c_\pi(r_1, r_3, r_5) + c_\pi(r_2, r_4, r_6)\}. \tag{4.7}$$

Proof. Let us consider the one dimensional Coulomb cost defined in $\mathbb{R}$

$$\bar{c}(v_1, v_2, v_3) = \frac{1}{|v_2 - v_1|} + \frac{1}{|v_3 - v_2|} + \frac{1}{|v_1 - v_3|} \qquad \forall v_1, v_2, v_3 \in \mathbb{R}.$$

We notice that $c_\pi(r_1, r_4, r_5) = \bar{c}(r_1, -r_4, r_5)$ and, more in general, for all the 3-uples appearing in (4.7) the c_π-cost and the $\bar{c}$-cost satisfy the same relation. In [5, Proposition 2.4] it is proved that, given the six points $-r_4, -r_3, r_1, r_2, r_5, r_6$ the best way to choose two 3-uples to minimize the one dimensional Coulomb cost is to take the points in odd position and the points in even position; in particular, we have

$$\bar{c}(-r_4, r_1, r_5) + \bar{c}(-r_3, r_2, r_6) \leq \min \{\bar{c}(-r_4, r_1, r_6) + \bar{c}(-r_3, r_2, r_5),$$
$$\bar{c}(-r_3, r_1, r_6) + \bar{c}(-r_4, r_2, r_5), \ \bar{c}(-r_3, r_1, r_5) + \bar{c}(-r_4, r_2, r_6)\},$$

which proves (4.7). $\qquad\square$

Remark 4.3. The previous lemma allows to prove that, for the cost c_π, the symmetrized optimal plan for the problem (2.3) is unique and coincides with the symmetrization of the *DDI*-map.

5   Proofs of examples and counterexamples

Proof of Counterexample 2.6. Let t_0 be given by Lemma 3.2 and let us choose $\varepsilon \leq t_0/2$. If, by contradiction, the *DDI*-map T associated to ρ_ε is optimal, by Proposition 2.4 its support is c-monotone. Let us consider $1 + \varepsilon, 1 + 3\varepsilon$ and the images of these points through T and $T \circ T$:

$$T(1 + \varepsilon) = 1 + 7\varepsilon, \quad T \circ T(1 + \varepsilon) = 1 + 9\varepsilon,$$

$$T(1 + 3\varepsilon) = 1 + 5\varepsilon, \quad T \circ T(1 + 3\varepsilon) = 1 + 11\varepsilon,$$

We notice that these points

$$(r_1, \ldots, r_6) = (1 + \varepsilon, 1 + 3\varepsilon, 1 + 5\varepsilon, 1 + 7\varepsilon, 1 + 9\varepsilon, 1 + 11\varepsilon),$$

are equally spaced; hence, we can apply the scaling properties of the cost function and Lemma 3.2 with $t = 2\varepsilon/(1+\varepsilon) \leq t_0$ to deduce that,

$$
\begin{aligned}
c(r_1, r_4, r_6) + c(r_2, r_3, r_5) &= \frac{1}{1+\varepsilon}\left[c\Big(\frac{r_1}{1+\varepsilon}, \frac{r_4}{1+\varepsilon}, \frac{r_6}{1+\varepsilon}\Big) + c\Big(\frac{r_2}{1+\varepsilon}, \frac{r_3}{1+\varepsilon}, \frac{r_5}{1+\varepsilon}\Big) \right] \\
&< \frac{1}{1+\varepsilon}\left[c\Big(\frac{r_1}{1+\varepsilon}, \frac{r_4}{1+\varepsilon}, \frac{r_5}{1+\varepsilon}\Big) + c\Big(\frac{r_2}{1+\varepsilon}, \frac{r_3}{1+\varepsilon}, \frac{r_6}{1+\varepsilon}\Big) \right] \\
&= c(r_1, r_4, r_5) + c(r_2, r_3, r_6).
\end{aligned}
$$

This contradicts the c-monotonicity of the support by taking $p = \{3\}$. $\qquad\square$

Remark 5.1. In the paper in preparation [10] it is shown that, in the setting of Counterexample 2.6, none of the maps in the $\{D, I\}^3$-class associated to ρ_ε is optimal in problem (2.2), by considering the limit problem as $\varepsilon \to 0$ of a suitably rescaled problem and analyzing the optimal maps with the cost $c_0(x_1, x_2, x_3) = |x_1 + x_2 + x_3|^2$.

Proof of Counterexample 2.7. **Step 1.** By choosing ε sufficiently small (independently of M), we exclude that the DDI-map is optimal in problem (2.2) for every $M > 2$.

Let T be the piecewise continuous DDI-map. Consider the following two points in the support of the plan associated to T (recall that the support is a closed set):

$$
\left(1 + \frac{\varepsilon}{2}, T\left(1 + \frac{\varepsilon}{2}\right), T^{(2)}\left(1 + \frac{\varepsilon}{2}\right)\right) = \left(1 + \frac{\varepsilon}{2}, 1 + \frac{7\varepsilon}{2}, 1 + \frac{9\varepsilon}{2}\right),
$$
$$
\lim_{r \to 1+\varepsilon^-} \left(r, T(r), T^{(2)}(r)\right) = (1 + \varepsilon, 1 + 3\varepsilon, 1 + 5\varepsilon).
$$

We claim that they violate the c-monotonicity property (Proposition 2.4) with $p = \{3\}$, namely

$$
\begin{aligned}
f(\varepsilon) &= c\left(1 + \frac{\varepsilon}{2}, 1 + \frac{7\varepsilon}{2}, 1 + \frac{9\varepsilon}{2}\right) + c(1 + \varepsilon, 1 + 3\varepsilon, 1 + 5\varepsilon) \\
&\quad - \left[c\left(1 + \frac{\varepsilon}{2}, 1 + \frac{7\varepsilon}{2}, 1 + 5\varepsilon\right) + c\left(1 + \varepsilon, 1 + 3\varepsilon, 1 + \frac{9\varepsilon}{2}\right) \right] > 0
\end{aligned}
$$

for ε sufficiently small. The proof is similar to that of Lemma 3.2. Using the formulas obtained in Lemma 3.1 we just compute the derivatives

$$
f(0) = f'(0) = f''(0) = 0,
$$
$$
f'''(0) = \frac{71\sqrt{3}}{100} > 0.
$$

Step 2. We exclude that the maps DID, IDD, III in the $\{D, I\}^3$-class are optimal in problem (2.2) for M large enough.

We present the argument to exclude the DID-map, the others being similar. Let us fix $x, y \in (M + 1/4, M + 3/4)$, $x < y$, and let us consider their orbits through T, that is $T(x), T(y) \in (1, 1 + \varepsilon_0)$ and $T^{(2)}(x), T^{(2)}(y) \in (1 + 3\varepsilon_0, 1 + 4\varepsilon_0)$. Let us consider the increasingly ordered points

$$
(r_1, ..., r_6) = \left(T(y), T(x), T^{(2)}(x), T^{(2)}(y), x, y\right);
$$

the couples of points (r_1, r_4, r_6) and (r_2, r_3, r_5) belong to the support of the plan associated to the DID-map. By Lemma 4.1, we can choose M sufficiently large so that the previous points, as well as the points (r_1, r_4, r_5) and (r_2, r_3, r_6), have the same c and c_π cost. By Lemma 4.2, which describes the c_π monotonicity, we have

$$
\begin{aligned}
c(r_1, r_4, r_5) + c(r_2, r_3, r_6) &= c_\pi(r_1, r_4, r_5) + c_\pi(r_2, r_3, r_6) \\
&\leq c_\pi(r_1, r_4, r_6) + c_\pi(r_2, r_3, r_5) \\
&= c(r_1, r_4, r_6) + c(r_2, r_3, r_5).
\end{aligned}
$$

This shows, by Proposition 2.4, that the DID-map cannot be optimal. $\qquad\square$

Remark 5.2. Our method can be applied to the 4-marginal problem to show that there exists $\varepsilon > 0$ such that, setting

$$
\rho_\varepsilon = \frac{1}{16\varepsilon} 1_{[1, 1+16\varepsilon]} \, dr \in \mathcal{M}(\mathbb{R}_+),
$$

any map in the $\{D, I\}^4$-class associated to ρ_ε is not optimal in problem (2.2). Indeed, let T be any such map. Pick two points in $[1, 1+16\varepsilon]$ such that the union of their two orbits is

$$
\{r_1, \ldots, r_8\} = \{1 + \varepsilon, 1 + 3\varepsilon, 1 + 5\varepsilon, 1 + 7\varepsilon, 1 + 9\varepsilon, 1 + 11\varepsilon, 1 + 13\varepsilon, 1 + 15\varepsilon\}.
$$

We claim that T is not c-monotone because the partitioning of $\{r_1, \ldots, r_8\}$ into two quartets that minimizes

$$
c(r_{i_1}, r_{i_2}, r_{i_3}, r_{i_4}) + c(r_{i_5}, r_{i_6}, r_{i_7}, r_{i_8})
$$

is $\{(r_1, r_5, r_6, r_7), (r_2, r_3, r_4, r_8)\}$ and such partition doesn't correspond to any of the maps in the $\{D, I\}^4$-class.

The way to see this is to extend the results of Section 3 to the 4-marginal case. Consider four radii

$$
(r_1, r_2, r_3, r_4) = (1 + a_1 t, 1 + a_2 t, 1 + a_3 t, 1 + a_4 t).
$$

Following the same derivation, we find that the angles that give the cost c are

$$
\begin{pmatrix} \theta_2(t) \\ \theta_3(t) \\ \theta_4(t) \end{pmatrix} = \begin{pmatrix} \pi/2 \\ \pi \\ 3/3\pi \end{pmatrix} + \frac{6 - \sqrt{2}}{34} \begin{pmatrix} -a_1 - a_2 + a_3 + a_4 \\ 2a_4 - 2a_2 \\ a_1 - a_2 - a_3 + a_4 \end{pmatrix} t + o(t).
$$

In turn, this provides the expansion of the cost up to the third order and this information can be used to verify the asymptotic optimality of any given partition. We omit the formulas, since this computations are better performed with the aid of a computer algebra system.

Proof of Proposition 2.9. Let M be chosen, thanks to Lemma 4.1, so that

$$
c(r_1, r_2, r_3) = c_\pi(r_1, r_2, r_3) \qquad \text{for every } r_1 \in [1, 2], \ r_2 \in [3, 4], \ r_3 \in [M, \infty). \qquad (5.1)
$$

In order to prove the c-monotonicity property, since the map T is cyclical and since its orbits take exactly one point in each interval $[1,2]$, $[3,4]$, and $[M,\infty)$, it is enough to show that, given $x,y \in [1,2]$, $x < y$, we have

$$c(x,T(x),T^{(2)}(x)) + c(y,T(y),T^{(2)}(y)) \leq c(x,A,B) + c(y,C,D) \qquad (5.2)$$

for every possible choice of A,B,C,D such that $\{A,C\} = \{T(x),T(y)\}$ and $\{B,D\} = \{T^{(2)}(x),T^{(2)}(y)\}$. By definition, we have that

$$1 \leq x < y \leq 2 \leq 3 \leq T(y) < T(x) \leq 4 \leq M \leq T^{(2)}(x) < T^{(2)}(y);$$

hence by (5.1) we have that $c(x,T(x),T^{(2)}(x)) = c_\pi(x,T(x),T^{(2)}(x))$ (and similarly for y and for the other 3-uples) and by Lemma 4.2 we have that

$$\begin{aligned}
c(x,T(x),T^{(2)}(x)) + c(y,T(y),T^{(2)}(y)) &= c_\pi(x,T(x),T^{(2)}(x)) + c_\pi(y,T(y),T^{(2)}(y)) \\
&\leq c_\pi(x,A,B) + c_\pi(y,C,D) = c(x,A,B) + c(y,C,D),
\end{aligned}$$

for every possible choice of A,B,C,D such that $\{A,C\} = \{T(x),T(y)\}$ and $\{B,D\} = \{T^{(2)}(x),T^{(2)}(y)\}$; this proves (5.2). $\qquad\square$

References

[1] L. Ambrosio, *Lecture notes on optimal transport problems*, Lecture Notes in Math., **1812**, Springer, Berlin, 2003, 1–52.

[2] L. Ambrosio, N. Gigli, G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics, ETH Zürich, Birkhäuser (2008).

[3] G. Buttazzo, L. De Pascale, P. Gori-Giorgi, *Optimal transport formulation of electronic density-functional theory*, Phys. Rev. A, **85** (2012), 062502.

[4] M. Colombo, S. Di Marino, *Equality between Monge and Kantorovich multimarginal problems with Coulomb cost*, Ann. Mat. Pura Appl., **194** (2015), 307–320.

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

[6] C. Cotar, G. Friesecke, C. Klüppelberg, *Density functional theory and optimal transportation with Coulomb cost*, Comm. Pure Appl. Math., **66** (2013), 548–599.

[7] C. Cotar, G. Friesecke, B. Pass, *Infinite-body optimal transport with Coulomb Cost*, to appear on Calc. Var. Partial Differential Equations, DOI 10.1007/s00526-014-0803-0.

[8] L. De Pascale, *Optimal transport with Coulomb cost: approximation and duality*, preprint 2015.

[9] S. Di Marino, A. Gerolin, L. Nenna, *Optimal transportation theory for repulsive costs*, preprint 2015.

[10] S. Di Marino, L. Nenna, P. Gori-Giorgi, *The strictly-correlated electron functional for spherically symmetric systems revisited*, in preparation.

[11] N. Ghoussoub, A. Moameni, *Symmetric Monge-Kantorovich problems and polar decompositions of vector fields*, Geom. Funct. Anal. **24** (2014), 1129–1166.

[12] P. Hohenberg, W. Kohn, *Inhomogeneous electron gas*, Physical review, **136** (1964), B864.

[13] W. Kohn, L.J. Sham, *Self-consistent equations including exchange and correlation effects*, Physical Review, **140** (1965), A1133.

[14] B. Pass, *On the local structure of optimal measures in the multi-marginal optimal transportation problem*, Calc. Var. Partial Differential Equations, **43** (2012), 529–536.

[15] B. Pass, *Remarks on the semi-classical Hohenberg-Kohn functional*, Nonlinearity, **26** (2013), 2731–2744.

[16] S. Rachev, L. Rüschendorf, *Mass Transportation Problems*, Springer-Verlag, New York (1998).

[17] M. Seidl, *Strong-interaction limit of density-functional theory*, Phys. Rev. A, **60** (1999), 4387.

[18] M. Seidl, P. Gori-Giorgi, A. Savin, *Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities*, Phys. Rev. A, **75** (2007), 042511.

[19] C. Villani, *Optimal transport. Old and new*, Grundlehren der Mathematischen Wissenschaften, **338**, Springer-Verlag, Berlin (2009).