SCUOLA NORMALE SUPERIORE

Classe di Scienze Matematiche, Fisiche e Naturali

Multi-marginal Optimal Transport: theory and applications

Tesi di Perfezionamento in Matematica

Candidato

Ugo Bindini

Relatore

Prof. **Luigi De Pascale**

Anno accademico 2020/2021

Introduction

In Mathematics, the theory of Optimal Transport has a long and interesting history. It dates back to the work of G. Monge "Sur la théorie des déblais et des remblais" (1781), where he investigated the cheapest way to transfer mass from a source place to a target. Crucial contributions were then given in the 20th century by A. N. Tolstoi, N. Kantorovich and may others. In the last years, this branch is getting increasing popularity, due to the high number of applications in different areas of mathematical analysis, physics, chemistry. Also, a great boost to the subject was provided by the interest of Fields' medallists C. Villani and A. Figalli. For general monographs we refer to the unsurpassed works of C. Villani [41, 42].

We will concentrate our attention to the problem of Optimal Transport in the Euclidean space $\mathbb{R}^d$. A great progress has also been made in more general settings, but we prefer to treat problems which are open already in the Euclidean setting and to keep an eye on applications (see for instance chapter 5).

We start by recalling a modern formulation of the classical Monge's Optimal Transport problem. Given a cost function $c\colon \mathbb{R}^d \times \mathbb{R}^d \to \overline{\mathbb{R}}$, and given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ probability measures, we state the minimization problem

$$(\mathcal{M}_2) = \inf \left\{ \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x) \mid T_{\#}\mu = \nu \right\}. \tag{0.1}$$

Here $T_{\#}\mu$ denotes the push-forward of a measure, given by $T_{\#}\mu(E) = \mu(T^{-1}(E))$ for every $E \subseteq \mathbb{R}^d$ μ-measurable (when using this symbol we assume that the map T is μ-measurable).

This formulation respects very well the application to a real-world transport of mass from a source distribution to a target, as the transport map T prescribes deterministically the destination of every source point. However, when looking at this problem from a mathematical viewpoint, it was soon realized that the space of admissible maps T does not enjoy a good structure in order to treat a variational problem. For instance, given a minimizing sequence of maps $(T_n)_{n\in\mathbb{N}}$, in general there is no way to obtain a suitable limit $T = \lim_{n\to\infty} T_n$, and the infimum in (0.1) is in general not a minimum.

In view of this, it is often convenient to consider the Kantorovich relaxed formulation of the same problem, described by

$$(\mathcal{K}_2) = \min\left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)d\gamma(x,y) \mid \gamma\in\Pi(\mu,\nu) \right\}; \tag{0.2}$$

here $\Pi(\mu,\nu)$ is the set of admissible transport plans, given by

$$\Pi(\mu,\nu) = \left\{ \gamma\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d) \mid \pi^1_{\#}\gamma = \mu, \pi^2_{\#}\gamma = \nu \right\},$$

where π^1, π^2 denote the projection $\pi^j\colon \mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^d$ on the first and second component respectively. In other words, a transport plan γ should satisfy

$$\int \phi(x)d\gamma(x,y) = \int \phi(x)d\mu(x) \quad \text{and} \quad \int \phi(y)d\gamma(x,y) = \int \phi(y)d\nu(y)$$

for every $\phi\in C_b(\mathbb{R}^d)$.

This problem is much more stable from a variational viewpoint. The set of transport plans $\Pi(\mu,\nu)$ is convex and tight. Moreover, the problem is stated as a minimum of a linear functional over a convex set, and this is a well-understood class of problems in convex analysis.

Observe that (0.2) is indeed a relaxation of the problem (0.1), since every map T corresponds to a plan $\gamma = (\mathrm{id}\times T)_{\#}\mu$. As a consequence, $(\mathcal{K}_2)\le(\mathcal{M}_2)$. The converse is of course not true: not every transport plan is induced by a map.

The question whether the equality $(\mathcal{K}_2) = (\mathcal{M}_2)$ holds is much more involved. As shown by A. Pratelli in [39], the equality holds for a continuous cost function.

This does not say, however, that there exists always an optimal transport map T which realizes the infimum in (0.1). A general positive result for the existence of an optimal transport map is given by the following

Theorem 0.0.1. *Suppose that $c(x,y) = h(x-y)$ with h strictly convex, and μ is absolutely continuous. Then there exists a unique optimal transport map T such that*

$$(\mathcal{M}_2) = (\mathcal{K}_2) = \int_{\mathbb{R}^d} c(x,T(x))\,\mathrm{d}\mu(x).$$

In the case $h = |x|^2$, Theorem 0.0.1 is often referred to as "Brenier's theorem", since it was first proven by Y. Brenier in [8].

Multi-marginal optimal transport The Kantorovich formulation of the optimal transport problem admits a natural generalization to the so-called *multi-marginal optimal transport problem*: given a cost function $c\colon(\mathbb{R}^d)^N\to\overline{\mathbb{R}}$, and given $\rho_1,\dots,\rho_N\in\mathcal{P}(\mathbb{R}^d)$, find

$$(\mathcal{K}_N) = \min\left\{ \int c(x_1,\dots,x_N)d\gamma(x_1,\dots,x_N) \mid \gamma\in\Pi(\rho_1,\dots,\rho_N) \right\}; \tag{0.3}$$

here $\Pi(\rho_1, \dots, \rho_N)$ denotes the set of admissible transport plans given by

$$\Pi(\rho_1, \dots, \rho_N) = \left\{ \gamma \in \mathcal{P}((\mathbb{R}^d)^N) \mid \pi^j_\# \gamma = \rho_j \ \forall j = 1, \dots, N \right\},$$

where $\pi^j : (\mathbb{R}^d)^N \to \mathbb{R}^d$ denotes the projection on the j-th component.

This minimization problem shares the same good structure as the classical Kantorovich formulation with two marginals (linearity of the functional, convexity and tightness of the set).

A first interesting question, which arises also in the classical 2-marginal case, is the following. Suppose that the cost function c is not bounded: for which marginals $\rho_1, \dots, \rho_N$ is the Kantorovich minimum (0.3) finite? This is not a merely theoretical problem: for instance, in many applications, the Riesz cost

$$c_s(x_1, \dots, x_N) = \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|^s}$$

(for some exponent $s > 0$) is of interest (see for instance [20, 38]), and in general the theory for repulsive cost functions is well studied ([13, 23, 25, 16]). We will study this question in Chapter 1, providing a sharp sufficient condition for the finiteness of (0.3) in the case when all the marginals are equal.

The Monge formulation can be extended as well to the multi-marginal case, by letting

$$(\mathcal{M}_N) = \inf \left\{ \int_{(\mathbb{R}^d)^N} c(x, T_2(x), \dots, T_N(x)) \, \mathrm{d}\rho_1(x) \mid (T_j)_\# \rho_1 = \rho_j \right\}, \quad (0.4)$$

where instead of searching for a single optimal map T as in (0.1), we search for $N - 1$ maps, each transporting the first marginal ρ_1 to one of the others.

Recalling the result of Pratelli [39] for two marginals, one could easily get that $(\mathcal{M}_N) = (\mathcal{K}_N)$ in the case of a continuous cost function and a non-atomic first marginal ρ_1. Indeed, as already pointed out in [15, Remark 1.3], the problem $(\mathcal{M}_N)$ can be seen as a 2-marginal optimal transport problem between $\mathbb{R}^d$ and $(\mathbb{R}^d)^{N-1}$.

This suggests that, even when the marginals $\rho_1, \dots, \rho_N$ enjoy some regularity assumptions (*e.g.* they are absolutely continuous w.r.t. the Lebesgue measure), the support of optimal transport plans tends to concentrate on sets of zero Lebesgue measure. In the extreme case of a Monge solution, the support is concentrated on the graph of a function $T : \mathbb{R}^d \to (\mathbb{R}^d)^{N-1}$. However, for many application to physics, it is important to exhibit a diffused transport plan γ which is "almost" optimal for (0.3), in the sense that $\int c(X) \, \mathrm{d}\gamma(X) \le (\mathcal{K}_N) + \epsilon$. We address this problem in Chapter 4, and we will show how this has been applied to quantum physical systems in Chapter 5.

In many applications, included the latter to quantum systems, the multi-marginal optimal transport problem is posed with N equal marginals $\rho_1 = \dots = \rho_N = \rho \in \mathcal{P}(\mathbb{R}^d)$ and a symmetric cost function c, *i.e.*, $c(x_1, \dots, x_N) =$

$c(x_{\sigma(1)}, \ldots, x_{\sigma(N)})$ for every permutation $\sigma \in \mathfrak{S}_N$. In this setting, an alternative definition of the multi-marginal Monge problem can be given by the following

$$(\mathcal{M}_{cyc}) = \inf \left\{ \int_{(\mathbb{R}^d)^N} c(x, T(x), \ldots, T^{N-1}(x))\, \mathrm{d}\rho(x) \mid T_{\#}\rho = \rho, T^N = id \right\}, \tag{0.5}$$

where T^j denotes the composition of T with itself j times. In [15], M. Colombo and S. Di Marino proved that $(\mathcal{M}_{cyc}) = (\mathcal{K}_N)$ in the case of a continuous cost function $c\colon (\mathbb{R}^d)^N \to [0, +\infty]$.

The existence of an optimal transport map T that realizes the infimum in $(\mathcal{M}_{cyc})$ is much more difficult. Some positive results (see [14]) are known in dimension $d = 1$ for the Coulomb cost function

$$c(x_1, \ldots, x_N) = \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|}.$$

For particular classes of cost functions generated by vector fields, the existence of optimal maps was proven by N. Ghoussoub et al. in [30, 29]. In general, the problem of finding optimal transport maps has been solved only under special assumptions on the local behaviour of the cost (see [28, 35, 31]).

In chapter 2, inspired by [17], we will give an original contribute to the "Monge question" in multi-marginal optimal transport, in the case of the Coulomb cost function in dimension $d = 2$, for a spherically symmetric marginal ρ. This is part of a work in collaboration with L. De PAscale and A. Kausamo. As we already pointed out, the choice of the Coulomb cost is due to the large number of applications of multi-marginal Optimal Transport to physical systems of charged particles (see [10, 21, 33]).

Remaining in this applied setting with equal marginals, one is often interested in the limit $N \to +\infty$ of a large number of particles. Some recent works [36, 21] present some progress in this direction. However, the former treats only the case of a positive-definite cost function, while the latter treats the pointwise limit, obtaining a very sharp result. In chapter 3 we will study the Γ-convergence of the multi-marginal OT functional in the case of a general pairwise cost function. This will extend the result of B. Pass et al. [36].

Chapter 1

Marginals with finite repulsive cost

In this chapter we consider a multi-marginal optimal transport problem with a pairwise repulsive cost function $c\colon (\mathbb{R}^d)^N \to \mathbb{R}$, *i.e.*,

$$c(X) = \sum_{1\leq i<j\leq N} \frac{1}{\omega(|x_i - x_j|)}, \tag{1.1}$$

where $X = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ and $\omega\colon \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, strictly increasing, differentiable on $(0, +\infty)$, with $\omega(0) = 0$. The problem we want to address is the following: for which probabilities $\rho \in \mathcal{P}(\mathbb{R}^d)$ is the transport cost

$$C(\rho) = \inf \left\{ \int c(X)\,\mathrm{d}\gamma(X) \mid \gamma \in \Pi(\rho, \dots, \rho) \right\}$$

finite?

Due to the fact that the cost function diverges on the diagonal

$$D = \left\{ (x_1, \dots, x_N) \in (\mathbb{R}^d)^N \mid x_i = x_j \text{ for some } i \neq j \right\}, \tag{1.2}$$

the question is not trivial. For instance, if $\rho = \delta_{x_0}$, then the set of transport plans $\Pi(\rho, \dots, \rho)$ consists only of the element $\gamma = \rho \otimes \cdots \otimes \rho$, and $C(\rho) = \int c(X)\,\mathrm{d}\gamma(X) = +\infty$.

It turns out that the right quantity to consider is given by the following

Definition 1. If $\rho \in \mathcal{P}(\mathbb{R}^d)$, the *concentration* of ρ is $\kappa(\rho) = \sup_{x\in\mathbb{R}^d} \rho(\{x\})$.

Let us state the main result of this chapter.

Theorem 1.0.1. *Let c be a repulsive cost function, and $\rho \in \mathcal{P}(\mathbb{R}^d)$ with*

$$\kappa(\rho) < \frac{1}{N}. \tag{1.3}$$

Then $C(\rho)$ is finite.

In an independent work by F. Stra, S. Di Marino and M. Colombo [16], appeared after the publication of [2], the same result and some finer study of the problem is obtained via a different technique, closer in the approach to some arguments in [9].

Structure of the chapter The material for this chapter comes mainly from [2]. In Section 1.1 we give some notation, and regroup some definitions, constructions and results to be used later. In particular, we state and prove a simple but useful result about partitioning $\mathbb{R}^d$ into measurable sets with prescribed mass.

We then show in Section 1.2 that the condition (1.3) is sharp, *i.e.*, given any repulsive cost function, there exists $\rho \in \mathcal{P}(\mathbb{R}^d)$ with $\kappa(\rho) = 1/N$, and $C(\rho) = \infty$. The construction of this counterexample is explicit, but it is important to note that the marginal ρ depends on the given cost function.

Finally we devote Sections 1.3 to 1.5 to the proof of Theorem 1.0.1. The construction is universal, in the following sense: given $\rho \in \mathcal{P}(\mathbb{R}^d)$ such that (1.3) holds, we exhibit a symmetric transport plan γ which has support outside the enlarged diagonal

$$D_\alpha = \Big\{(x_1, \dots, x_N) \in (\mathbb{R}^d)^N \mid \exists i \neq j \text{ with } |x_i - x_j| < \alpha\Big\}$$

for some $\alpha > 0$. This implies that $\int c(X)\,\mathrm{d}\gamma(X)$ is finite for any repulsive cost function.

1.1 Notation and preliminary results

We will denote by $B(x_j, r)$ a ball with center $x_j \in \mathbb{R}^d$ and radius $r > 0$. Where it is not specified, the integrals are extended to all the space; if τ is a positive measure over $\mathbb{R}^d$, we denote by $|\tau|$ its total mass, *i.e.*,

$$|\tau| = \int_{\mathbb{R}^d} \mathrm{d}\tau.$$

If $\mu \in \mathcal{M}((\mathbb{R}^d)^N)$ is any measure, we define

$$\mu_{sym} = \frac{1}{N!} \sum_{s \in \mathfrak{S}_N} \phi^s_{\#}\mu,$$

where $\mathfrak{S}_N$ is the premutation group of $\{1, \dots, N\}$, and $\phi^s \colon (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N$ is the function $\phi^s(x_1, \dots, x_N) = (x_{s(1)}, \dots, x_{s(N)})$. Note that μ_{sym} is a symmetric measure; moreover, if μ is a probability measure, then also μ_{sym} is a probability measure.

Lemma 1.1.1. *Let $\mu \in \mathcal{M}((\mathbb{R}^d)^N)$. Then μ_{sym} has marginals equal to*

$$\frac{1}{N} \sum_{j=1}^{N} \pi^j_{\#}\mu$$

Proof. Since μ_{sym} is symmetric, me may calculate its first marginal:

$$\pi^1_\# \mu_{sym} = \pi^1_\# \left(\frac{1}{N!} \sum_{s \in \mathfrak{S}_N} \phi^s_\# \mu \right) = \frac{1}{N!} \sum_{s \in \mathfrak{S}_N} \pi^1_\# (\phi^s_\# \mu)$$
$$= \frac{1}{N!} \sum_{s \in \mathfrak{S}_N} \pi^{s(1)}_\# \mu = \frac{1}{N} \sum_{j=1}^N \pi^j_\# \mu,$$

where the last equality is due to the fact that for every $j = 1, \dots, N$ there are exactly $(N-1)!$ permutations in $\mathfrak{S}_N$ such that $s(1) = j$. □

For a symmetric probability $\gamma \in \mathcal{P}((\mathbb{R}^d)^N)$ we will use the shortened notation $\pi(\gamma)$ to denote its marginals $\pi^j_\# \gamma$, which are all equal.

If $\sigma_1, \dots, \sigma_N \in \mathcal{M}(\mathbb{R}^d)$, we define $\sigma_1 \otimes \cdots \otimes \sigma_N \in \mathcal{M}((\mathbb{R}^d)^N)$ as the usual product measure. In similar fashion, if $Q \in \mathcal{M}((\mathbb{R}^d)^{N-1})$, $\sigma \in \mathcal{M}(\mathbb{R}^d)$ and $1 \leq j \leq N$, we define the measure $Q \otimes_j \sigma \in \mathcal{M}((\mathbb{R}^d)^N)$ as

$$\int_{(\mathbb{R}^d)^N} f \, \mathrm{d}(Q \otimes_j \sigma) = \int_{(\mathbb{R}^d)^N} f(x_1, \dots, x_N) \, \mathrm{d}\sigma(x_j) \, \mathrm{d}Q(x_1, \dots, \hat{x}_j, \dots, x_N) \tag{1.4}$$

for every $f \in C_b((\mathbb{R}^d)^N)$.

Partitions of non-atomic measures

Let $\sigma \in \mathcal{M}(\mathbb{R}^d)$ be a finite non-atomic measure, and $b_1, \dots, b_k$ real positive numbers such that $b_1 + \cdots + b_k = |\sigma|$. We may want to write

$$\mathbb{R}^d = \bigcup_{j=1}^k E_j,$$

where the E_j's are disjoint measurable sets with $\sigma(E_j) = b_j$. This is trivial if $d = 1$, since the cumulative distribution function $\phi_\sigma(t) = \sigma((-\infty, t))$ is continuous, and one may find the E_j's as intervals. However, in higher dimension, the measure σ might concentrate over $(d-1)$-dimensional surfaces, which makes the problem slightly more difficult. Therefore we present the following

Proposition 1.1.2. *Let $\sigma \in \mathcal{M}(\mathbb{R}^d)$ be a finite non-atomic measure. Then there exists a direction $y \in \mathbb{R}^d \setminus \{0\}$ such that $\sigma(H) = 0$ for all the affine hyperplanes H such that $H \perp y$.*

In order to prove Proposition 1.1.2, it is useful to present the following

Lemma 1.1.3. *Let (X, μ) be a measure space, with $\mu(X) < \infty$, and $\{E_i\}_{i \in I}$ a collection of measurable sets such that*

1. *$\mu(E_i) > 0$ for every $i \in I$;*

2. *$\mu(E_i \cap E_j) = 0$ for every $i \neq j$.*

Then I is at most countable.

Proof. Let $i_1, \dots, i_n$ be a finite set of indices. Using the monotonicity of μ and the fact that $\mu(E_i \cap E_j) = 0$ if $i \neq j$,

$$\mu(X) \geq \mu\left(\bigcup_{k=1}^{n} E_{i_k}\right) = \sum_{k=1}^{n} \mu(E_{i_k}).$$

Hence we have that

$$\sup\left\{\sum_{j\in J} \mu(E_j) \mid J \subset I, J \text{ finite}\right\} \leq \mu(X) < \infty.$$

Since all the $\mu(E_i)$ are strictly positive numbers, this is possible only if I is countable (at most). □

Now we present the proof of Proposition 1.1.2.

Proof. For $k = 0, 1, \dots, d-1$ we recall the definitions of the Grassmannian

$$\mathrm{Gr}(k, \mathbb{R}^d) = \left\{v \text{ linear subspace of } \mathbb{R}^d \mid \dim v = k\right\}$$

and the affine Grassmannian

$$\mathrm{Graff}(k, \mathbb{R}^d) = \left\{w \text{ affine subspace of } \mathbb{R}^d \mid \dim w = k\right\}.$$

Given $w \in \mathrm{Graff}(k, \mathbb{R}^d)$, we denote by $[w]$ the unique element of $\mathrm{Gr}(k, \mathbb{R}^d)$ parallel to w.

If $S \subseteq \mathrm{Graff}(k, \mathbb{R}^d)$, we say that S is *full* if for every $v \in \mathrm{Gr}(k, \mathbb{R}^d)$ there exists $w \in S$ such that $[w] = v$. For every $k = 1, 2, \dots, d-1$ let $S^k \subseteq \mathrm{Graff}(k, \mathbb{R}^d)$ be the set

$$S^k = \left\{w \in \mathrm{Graff}(k, \mathbb{R}^d) \mid \sigma(w) > 0\right\}.$$

The goal is to prove that S^{d-1} is *not* full, while by hypothesis we know that $S^0 = \emptyset$, since σ is non-atomic.

The following key Lemma leads to the proof in a finite number of steps:

Lemma 1.1.4. *Let $1 \leq k \leq d-1$. If S^{k-1} is not full, then S^k is not full.*

Proof. Let $v \in \mathrm{Gr}(k-1, \mathbb{R}^d)$, such that for every $v' \in \mathrm{Graff}(k-1, \mathbb{R}^d)$ with $[v'] = v$ it holds $\sigma(v') = 0$. Consider the collection

$$W_v = \left\{ w \in \mathrm{Graff}(k, \mathbb{R}^d) \mid v \subseteq [w] \right\}.$$

If $w, w' \in W_v$ are distinct, then $w \cap w' \subseteq v'$ for some $v' \in \mathrm{Graff}(k-1, \mathbb{R}^d)$ with $[v'] = v$, thus $\sigma(w \cap w') = 0$. Since the measure σ is finite, because of Lemma 1.1.3 at most countably many elements $w \in W_v$ may have positive measure, which implies that S^k is not full. □

□

Corollary 1.1.5. *Given $b_1, \dots, b_k$ real positive numbers with $b_1 + \cdots + b_k = |\sigma|$, there exist measurable sets $E_1, \dots, E_k \subseteq \mathbb{R}^d$ such that*

(i) The E_j's form a partition of $\mathbb{R}^d$, i.e.,

$$\mathbb{R}^d = \bigcup_{j=1}^{k} E_j, \qquad E_i \cap E_j = \emptyset \text{ if } i \neq j;$$

(ii) $\sigma(E_j) = b_j$ for every $j = 1, \dots, k$.

Proof. Let $y \in \mathbb{R}^d \setminus \{0\}$ given by Proposition 1.1.2, and observe that the cumulative distribution function

$$F(t) = \sigma\left(\left\{ x \in \mathbb{R}^d \mid x \cdot y < t \right\} \right)$$

is continuous. Hence we may find $E_1, \dots, E_k$ each of the form

$$E_j = \left\{ x \in \mathbb{R}^d \mid t_j < x \cdot y \leq t_{j+1} \right\}$$

for suitable $-\infty = t_1 < t_2 < \cdots < t_k < t_{k+1} = +\infty$, such that $\sigma(E_j) = b_j$. □

Corollary 1.1.6. *Given $b_1, \dots, b_k$ non-negative numbers with $b_1 + \cdots + b_k < |\sigma|$, there exists measurable sets $E_0, E_1, \dots, E_k \subseteq \mathbb{R}^d$ such that*

(i) the E_j's form a partition of $\mathbb{R}^d$;

(ii) $\sigma(E_j) = b_j$ for every $j = 1, \dots, k$;

(iii) the distance between E_i and E_j is strictly positive if $i, j \geq 1$, $i \neq j$.

Proof. If $k = 1$ the results follows trivially by Corollary 1.1.5 applied to $b_1, |\sigma| - b_1$. If $k \geq 2$, define

$$\epsilon = \frac{|\sigma| - b_1 - \cdots - b_k}{k-1} > 0.$$

As before, letting $y \in \mathbb{R}^d \setminus \{0\}$ given by Proposition 1.1.2 and considering the corresponding cumulative distribution function, we may find $F_1, \ldots, F_{2k-1}$ each of the form

$$F_j = \left\{ x \in \mathbb{R}^d \mid t_j < x \cdot y \leq t_{j+1} \right\}$$

for suitable $-\infty = t_1 < t_2 < \cdots < t_{2k-1} < t_{2k} = +\infty$, such that

$$\begin{aligned} \sigma(F_{2j-1}) &= b_j \quad \forall j = 1, \ldots, k \\ \sigma(F_{2j}) &= \epsilon \quad \forall j = 1, \ldots, k-1. \end{aligned}$$

Finally we define

$$\begin{aligned} E_j &= F_{2j-1} \quad \forall j = 1, \ldots, k \\ E_0 &= \bigcup_{j=1}^{k-1} F_{2j}. \end{aligned}$$

The properties (i), (ii) are immediate to check, while the distance between E_i and E_j, for $i, j \geq 1$, $i \neq j$, is uniformly bounded from below by

$$\min \{ t_{2j+1} - t_{2j} \mid 1 \leq j \leq k-1 \} > 0. \qquad \square$$

1.2 The condition (1.3) is sharp

In this section we prove that the condition (1.3) is sharp, in the sense that given any repulsive cost function there exists $\rho \in \mathcal{P}(\mathbb{R}^d)$ with $\mu(\rho) = 1/N$ such that $C(\rho) = \infty$.

Fix a repulsive cost as in (1.1), and set

$$k = \int_{B(0,1)} \frac{\omega'(|y|)}{|y|^{d-1}} \, \mathrm{d}y.$$

Note that k is a positive finite constant, depending only on ω and the dimension d. In fact, integrating in spherical coordinates,

$$k = \int_0^1 \frac{\omega'(r)}{r^{d-1}} \alpha_d r^{d-1} \, \mathrm{d}r = \alpha_d \omega(1),$$

where α_d is the d-dimensional volume of the unit ball $B(0,1) \subseteq \mathbb{R}^d$.

Now define a probability measure $\rho \in \mathcal{P}(\mathbb{R}^d)$ as

$$\int_{\mathbb{R}^d} f \, \mathrm{d}\rho := \frac{1}{N} f(0) + \frac{N-1}{N} \int_{B(0,1)} f(x) \frac{\omega'(|x|)}{k \, |x|^{d-1}} \, \mathrm{d}x \quad \forall f \in C_b(\mathbb{R}^d). \tag{1.5}$$

This measure has an atom of mass $1/N$ in the origin, and is absolutely continuous on $\mathbb{R}^d \setminus \{0\}$. Hence the concentration of ρ is equal to $1/N$, even if for every ball B around the origin one has $\rho(B) > 1/N$.

We want to prove that any symmetric transport plan with marginals ρ has infinite cost. Let us consider, by contradiction, a symmetric plan γ, with $\pi(\gamma) = \rho$, such that

$$\int \sum_{1 \le i < j \le N} \frac{1}{\omega(|x_i - x_j|)} \, \mathrm{d}\gamma(X) < \infty$$

Then one would have the following geometric properties.

Lemma 1.2.1. *(i) $\gamma(D) = 0$, where D is the diagonal defined in* (1.2)*;*

(ii) γ is concentrated over the N coordinate hyperplanes $\{x_j = 0\}$, $j = 1, \ldots, N$, i.e.,

$$\mathrm{supp}(\gamma) \subseteq E := \bigcup_{j=1}^{N} \{x_j = 0\} \,.$$

Proof. (i) Since $\omega(0) = 0$, recalling Definition 1.1, the cost function is identically equal to $+\infty$ on the diagonal D. Therefore, since by assumption the cost of γ is finite, it must be

$$\gamma(D) = 0.$$

(ii) Define

$$\begin{aligned} p_1 &= \gamma(\{x_1 = 0\}) \\ p_2 &= \gamma(\{x_1 = 0\} \cap \{x_2 = 0\}) \\ &\vdots \\ p_N &= \gamma((0, \ldots, 0)). \end{aligned}$$

Note that $p_1 = \gamma(\{x_1 = 0\}) = \pi(\gamma)(\{0\}) = \rho(\{0\}) = 1/N$. We claim that $p_2 = \cdots = p_N = 0$. Indeed, $\{x_1 = 0\} \cap \{x_2 = 0\} \subset D$, and hence $p_2 = 0$; by monotonicity of the measure γ we have $p_2 \ge p_3 \ge \cdots \ge p_N$.

Now by inclusion-exclusion we have

$$P(E) = \sum_{j=1}^{N} (-1)^{j+1} \binom{N}{j} p_j = N p_1 = 1,$$

and hence γ is concentrated over E. □

In view of Lemma 1.2.1, letting $H_j = \{x_j = 0\}$ for $j = 1, \ldots, N$,

$$\gamma = \sum_{j=1}^{N} \gamma|_{H_j}.$$

For every $j = 1, \ldots, N$ there exists a unique measure Q_j over $\mathbb{R}^{(N-1)d}$ such that, recalling equation (1.4), $\gamma|_{H_j} = Q_j \otimes_j \delta_0$, with $Q_j(\mathbb{R}^{(N-1)d}) = \frac{1}{N}$. Since

γ is symmetric, considering a permutation $s \in \mathfrak{S}_N$ with $s(j) = j$, it follows that Q_j is symmetric; then, considering any permutation in $\mathfrak{S}_N$ we see that there exists a symmetric probability Q over $\mathbb{R}^{(N-1)d}$ such that $Q_j = \frac{1}{N} Q$ for every $j = 1, \dots, N$, *i.e.*,

$$\gamma = \frac{1}{N} \sum_{j=1}^{N} Q \otimes_j \delta_0.$$

Projecting γ to its one-particle marginal and using the definition of ρ in (1.5), we get that $\pi(Q)$ is absolutely continuous w.r.t. the Lebesgue measure, with

$$\frac{\mathrm{d}\pi(Q)}{\mathrm{d}\mathcal{L}^d} = \frac{\chi_{B(0,1)}(x)\omega'(x)}{k\,|x|^{d-1}}.$$

Here we get the contradiction, because

$$\begin{aligned}\int c(X) d\gamma(X) &\geq \frac{1}{N} \int \frac{1}{\omega(|x_1 - x_2|)} \delta_0(x_1)\, \mathrm{d}x_1\, \mathrm{d}Q(x_2, \dots, x_N) \\ &= \frac{1}{N} \int \frac{1}{\omega(|x_2|)}\, \mathrm{d}Q(x_2, \dots, x_N) = \frac{1}{N} \int_{\mathbb{R}^d} \frac{1}{\omega(|x|)}\, \mathrm{d}\pi(Q)(x) \\ &= \frac{1}{N} \int_{B(0,1)} \frac{\omega'(|x|)}{\omega(|x|)} \frac{1}{k\,|x|^{d-1}}\, \mathrm{d}x = \frac{1}{N} \frac{\alpha_d}{k} \int_0^1 \frac{\omega'(r)}{\omega(r)}\, \mathrm{d}r = +\infty.\end{aligned}$$

1.3 Non-atomic marginals

This short section deals with the case where ρ is non atomic, *i.e.*, $\kappa(\rho) = 0$. In this case the transport plan is given by an optimal transport map in Monge's fashion, which we proceed to construct.

Using Corollary 1.1.5, let $E_1, \dots, E_{2N}$ be a partition of $\mathbb{R}^d$ such that

$$\rho(E_j) = \frac{1}{2N} \qquad \forall j = 1, \dots, 2N.$$

Next we take a measurable function $\phi \colon \mathbb{R}^d \to \mathbb{R}^d$, preserving the measure ρ and defined locally such that

$$\begin{aligned}\phi(E_j) &= E_{j+2} \qquad \forall j = 1, \dots, N-2 \\ \phi(E_{2N-1}) &= E_1 \\ \phi(E_{2N}) &= E_2.\end{aligned}$$

The behaviour of ϕ on the hyperplanes which separate the E_j's is arbitrary, since they form a ρ-null set. Note that $|x - \phi(x)|$ is uniformy bounded from below by some constant $\beta > 0$, as is clear by the construction of the E_j's (see the proof of Corollary 1.1.5). A transport plan γ of finite cost is now defined for every $f \in C_b((\mathbb{R}^d)^N)$ by

$$\int_{(\mathbb{R}^d)^N} f(X)\, \mathrm{d}\gamma(X) = \int_{(\mathbb{R}^d)^N} f(x, \phi(x), \dots, \phi^{N-1}(x))\, \mathrm{d}\rho(x),$$

since

$$\int_{(\mathbb{R}^d)^N} c(X)\,\mathrm{d}\gamma(X) = \binom{N}{2} \int_{\mathbb{R}^d} \frac{1}{\omega(|x-\phi(x)|)}\,\mathrm{d}\rho(x) \le \binom{N}{2} \frac{1}{\omega(\gamma)}.$$

1.4 Marginals with a finite number of atoms

This section constitutes the core of the proof, as we deal with measures of general form with an arbitrary (but finite) number of atoms. Throughout this and the next Section we assume that the marginal ρ fulfills the condition (1.3).

The number of atoms is less than or equal to N

Note that, if the number of atoms is at most N, then ρ must have a non-atomic part σ, due to the condition (1.3). From here on we consider

$$\rho = \sigma + \sum_{i=1}^{k} b_i \delta_{x_i},$$

where $b_1 \ge b_2 \ge \dots \ge b_k > 0$.

We begin with the following

Definition 2. A *partition* of σ of level $k \le N$ subordinate to $(x_1, \dots, x_k; b_1, \dots, b_k)$ is

$$\sigma = \tau + \sum_{i=1}^{k} \sum_{h=i+1}^{N} \sigma_h^i,$$

where:

(i) τ, σ_h^i are non-atomic measures;

(ii) for every i and every $h \neq k$, the distance between $\operatorname{supp}\sigma_h^i$ and $\operatorname{supp}\sigma_k^i$ is strictly positive;

(iii) for every i, h, if $j \le i$ then x_j has a strictly positive distance from $\operatorname{supp}\sigma_h^i$;

(iv) for every i, h, $\left|\sigma_h^i\right| = b_i$, and $|\tau| > 0$.

Note that such a partition may only exists if

$$|\sigma| > \sum_{i=1}^{k} (N-i) b_i. \tag{1.6}$$

On the other hand, the following Lemma proves that the condition (1.6) is also sufficient to get a partition of σ.

Lemma 1.4.1. *Let* $(b_1, \dots, b_k)$ *with* $k \le N$*, and*

$$|\sigma| > \sum_{i=1}^{k} (N-i) b_i .$$

Then there exists a partition of σ *subordinate to* $(x_1, \dots, x_k; b_1, \dots, b_k)$*.*

Proof. Fix $(x_1, \dots, x_k)$ and for every $\varepsilon > 0$ define

$$A_\varepsilon = \bigcup_{j=1}^{k} B(x_j, \varepsilon).$$

and $\sigma_\varepsilon = \sigma \chi_{A_\varepsilon}$. Then take ε small enough such that

$$|\sigma - \sigma_\varepsilon| > \sum_{i=1}^{k} (N-i) b_i, \tag{1.7}$$

which is possibile because $\mu(\sigma) = 0$ (σ has concentration zero), and hence $|\sigma_\varepsilon| \to 0$ as $\varepsilon \to 0$. Due to Corollary 1.1.6, the set $\mathbb{R}^d \setminus A_\varepsilon$ may be partitioned as

$$\mathbb{R}^d \setminus A_\varepsilon = \left(\bigcup_{i=1}^{k} \bigcup_{h=i+1}^{N} E_h^i \right) \cup E,$$

with $\sigma(E_h^i) = b_i$, and $\mathrm{dist}(E_h^i, E_k^i)$ is uniformly bounded from below.

Finally define $\sigma_h^i = \sigma \chi_{E_h^i}$, $\tau = \sigma_\varepsilon + \sigma \chi_E$.

□

Proposition 1.4.2. *Suppose that* $k \le N$ *and* $(b_1, \dots, b_k)$ *are such that*

$$|\sigma| > N b_1 - \sum_{j=1}^{k} b_j . \tag{1.8}$$

Then there exists a transport plan of finite cost with marginals

$$\sigma + \sum_{j=1}^{k} b_j \delta_{x_j} .$$

Proof. In order to simplify the notation, set $b_{k+1} = 0$. First of all we shall fix a partition of σ subordinate to $(x_1, \dots, x_k; b_1 - b_2, \dots, b_{k-1} - b_k, b_k)$. To do this we apply Lemma 1.6, since

$$\sum_{i=1}^{k-1} (N-i)(b_i - b_{i+1}) + (N-k) b_k = (N-1) b_1 - \sum_{i=2}^{k} b_i < |\sigma| .$$

Next we define the measures $\lambda_i = \delta_{x_1} \otimes \cdots \otimes \delta_{x_i} \otimes \sigma^i_{i+1} \otimes \cdots \otimes \sigma^i_N \in \mathcal{M}((\mathbb{R}^d)^N)$. Let us calculate the marginals of λ_i: since $\left|\sigma^i_h\right| = b_i - b_{i+1}$ for all $h = i+1, \dots, N$, we get

$$\pi^j_\# \lambda_i = \begin{cases} (b_i - b_{i+1})^{N-i} \delta_{x_j} & \text{if } 0 \le j \le i \\ (b_i - b_{i+1})^{N-i-1} \sigma^i_j & \text{if } i+1 \le j \le N. \end{cases}$$

Let us define, for $i = 1, \dots, k$, the measure

$$P_i = \frac{N}{(b_i - b_{i+1})^{N-i-1}} (\lambda_i)_{sym},$$

where $P_i = 0$ if $b_i = b_{i+1}$. By Lemma 1.1.1, the marginals of P_i are equal to

$$\pi(P_i) = \frac{1}{(b_i - b_{i+1})^{N-i-1}} \sum_{j=0}^{N} \pi^j_\# \lambda_i = \sum_{j=1}^{i} (b_i - b_{i+1}) \delta_{x_j} + \sum_{h=i+1}^{N} \sigma^i_h,$$

so that

$$\sum_{i=1}^{k} \pi(P_i) = \sum_{j=1}^{k} b_j \delta_{x_j} + \sum_{i=1}^{k} \sum_{h=i+1}^{N} \sigma^i_h$$

It suffices now to take any symmetric transport plan P_τ of finite cost with marginals τ, given by the result of Section 1.3, and finally set

$$\gamma = P_\tau + \sum_{i=1}^{k} P_i. \qquad \square$$

As a corollary we obtain

Theorem 1.4.3. *If ρ has $k \le N$ atoms, then there exists a transport plan of finite cost.*

Proof. Let

$$\rho = \sigma + \sum_{j=1}^{k} b_j \delta_{x_j}.$$

Note that, since $b_1 < 1/N$,

$$|\sigma| = 1 - \sum_{j=1}^{k} b_j > N b_1 - \sum_{j=1}^{k} b_j,$$

hence we may apply Proposition 1.4.2 to conclude. $\square$

The number of atoms is greater than N

Here we deal with the much more difficult situation in which ρ has $N+1$ or more atoms, *i.e.*,

$$\rho = \sigma + \sum_{j=1}^{k} b_j \delta_{x_j}$$

with $k \geq N+1$ and as before $b_1 \geq b_2 \geq \cdots \geq b_k > 0$. Note that in this case it might happen that $\sigma = 0$.

The main point is to use a double induction on the dimension N and the number of atoms k, as will be clear in Proposition 1.4.5. The following lemma is a nice numerical trick needed for the inductive step in Proposition 1.4.5.

Lemma 1.4.4. *Let $(b_1, \dots, b_k)$ with $k \geq N+2$ and*

$$(N-1)b_1 \leq \sum_{j=2}^{k} b_j. \tag{1.9}$$

Then there exist $t_2, \dots, t_k$ such that

(i) $t_2 + \cdots + t_k = (N-1)b_1$;

(ii) for every $j = 2, \dots, k$, $0 \leq t_j \leq b_j$, and moreover

$$t_2 \geq \cdots \geq t_k.$$

$$b_2 - t_2 \geq b_3 - t_3 \geq \cdots \geq b_k - t_k,$$

(iii)

$$(N-2)t_2 \leq \sum_{j=3}^{k} t_j;$$

(iv)

$$(N-1)(b_2 - t_2) \leq \sum_{j=3}^{k} (b_j - t_j).$$

Proof. For $j = 2, \dots, k$ define

$$p_j = \sum_{h=j}^{k} b_j,$$

and let $\bar{\jmath}$ be the least $j \geq 2$ such that $(N-j+2)b_j \leq p_j$; note that $j = N+2$ works — hence $\bar{\jmath} \leq N+2$. Define

$$t_j = b_j - \frac{p_2 - (N-1)b_1}{N} \qquad \text{for } j = 2, \dots, \bar{\jmath}-1,$$

$$t_j = b_j - \frac{b_j}{p_{\bar{\jmath}}} \frac{p_2 - (N-1)b_1}{N} (N - \bar{\jmath} + 2) \qquad \text{for } j = \bar{\jmath}, \dots, k.$$

Next we prove that this choice fulfills the conditions (i)-(iv).

Proof of (i)

$$\begin{aligned}\sum_{j=2}^{k} t_j &= p_2 - \frac{p_2-(N-1)b_1}{N}(\bar{\jmath}-2) - \frac{p_2-(N-1)b_1}{N}(N-\bar{\jmath}+2)\\ &= p_2\left(1-\frac{\bar{\jmath}-2}{N}-\frac{N-\bar{\jmath}+2}{N}\right)+(N-1)b_1\left(\frac{\bar{\jmath}-2}{N}+\frac{N-\bar{\jmath}+2}{N}\right)\\ &= (N-1)b_1.\end{aligned}$$

Proof of (ii) In view of the fact that $(N-1)b_1 \le p_2$ and $\bar{\jmath} \le N+2$, it is clear that $t_j \le b_j$. If $j < \bar{\jmath}$ we have $(N-j+2)b_j > p_j$, and hence

$$p_2 = b_2 + \cdots + b_{j-1} + p_j < (j-2)b_1 + (N-j+2)b_j.$$

Thus, since $2 \le j \le N+1$,

$$\begin{aligned}t_j &= \frac{Nb_j - p_2 + (N-1)b_1}{N} > \frac{Nb_j - (N-j+2)b_j - (j-2)b_1 + (N-1)b_1}{N}\\ &= \frac{(j-2)b_j + (N-j+1)b_1}{N} \ge 0.\end{aligned}$$

To show that $t_j \ge 0$ for $j \ge \bar{\jmath}$, we must prove $[p_2-(N-1)b_1](N-\bar{\jmath}+2) \le Np_{\bar{\jmath}}$, which is trivial if $\bar{\jmath} = N-2$. Otherwise, it is equivalent to

$$-(\bar{\jmath}-2)[p_2-(N-1)b_1] + N[b_2+\cdots+b_{\bar{\jmath}-1}-(N-1)b_1] \le 0.$$

Since $2 \le \bar{\jmath} \le N+1$, the first term is negative and $b_2+\cdots+b_{\bar{\jmath}-1}-(N-1)b_1 \le -(N-\bar{\jmath}+1)b_1 \le 0$.

Using the fact that $b_2 \ge \cdots \ge b_k$, it is easy to see that $t_2 \ge \cdots \ge \cdots t_{\bar{\jmath}-1}$ and $t_{\bar{\jmath}} \ge \cdots \ge t_k$ — note that for $j \ge \bar{\jmath}$ we have $t_j = \alpha b_j$, for some $0 \le \alpha \le 1$. As for the remaining inequality,

$$t_{\bar{\jmath}-1} \ge t_{\bar{\jmath}} \iff b_{\bar{\jmath}-1} - b_{\bar{\jmath}} \ge \frac{p_2-(N-1)b_1}{Np_{\bar{\jmath}}}[p_{\bar{\jmath}} - (N-\bar{\jmath}+2)b_{\bar{\jmath}}],$$

we already proved

$$\frac{p_2-(N-1)b_1}{Np_{\bar{\jmath}}} \le \frac{1}{N-\bar{\jmath}+2};$$

moreover, by definition of $\bar{\jmath}$, we have $(N-\bar{\jmath}+3)b_{\bar{\jmath}-1} > p_{\bar{\jmath}-1}$, or equivalently $(N-\bar{\jmath}+2)b_{\bar{\jmath}-1} > p_{\bar{\jmath}}$. Thus

$$\frac{p_2-(N-1)b_1}{Np_{\bar{\jmath}}}[p_{\bar{\jmath}} - (N-\bar{\jmath}+2)b_{\bar{\jmath}}] \le \frac{p_{\bar{\jmath}}}{N-\bar{\jmath}+2} - b_{\bar{\jmath}} < b_{\bar{\jmath}-1} - b_{\bar{\jmath}},$$

as wanted.

It is left to show that $b_2 - t_2 \geq \cdots \geq b_k - t_k$. It is trivial to check that $b_2 - t_2 = \cdots = b_{\bar{\jmath}-1} - t_{\bar{\jmath}-1}$, and $b_{\bar{\jmath}} - t_{\bar{\jmath}} \geq \cdots \geq b_k - t_k$ using $b_{\bar{\jmath}} \geq \cdots \geq b_k$ as before. Finally,

$$b_{\bar{\jmath}-1} - t_{\bar{\jmath}-1} \geq b_{\bar{\jmath}} - t_{\bar{\jmath}} \iff \frac{p_2 - (N-1)b_1}{N} \geq \frac{b_{\bar{\jmath}}}{p_{\bar{\jmath}}} \frac{p_2 - (N-1)b_1}{N}(N - \bar{\jmath} + 2),$$

which is true since $(N - \bar{\jmath} + 2)b_{\bar{\jmath}} \leq p_{\bar{\jmath}}$ and $p_2 - (N-1)b_1 \geq 0$.

Proof of (iii) The thesis is equivalent to

$$(N-1)t_2 \leq \sum_{j=2}^{k} t_j \iff (N-1)t_2 \leq (N-1)b_1,$$

and this is implied by $t_2 \leq b_2 \leq b_1$.

Proof of (iv) The thesis is equivalent to

$$N(b_2 - t_2) \leq p_2 - (N-1)b_1,$$

which is in fact an equality (see the definition of t_2).

□

We are ready to present the main result of this Section, which provides a transport plan of finite cost under an additional hypothesis on the tuple $(b_1, \ldots, b_k)$. The result is peculiar for the fact that it does not involve the non-atomic part of the measure – it is in fact a general discrete construction to get a purely atomic symmetric measure having fixed purely atomic marginals.

Proposition 1.4.5. *Let $k > N$ and $(b_1, \ldots, b_k)$ with*

$$(N-1)b_1 \leq b_2 + \cdots + b_k. \tag{1.10}$$

Then for every $x_1, \ldots, x_k \in \mathbb{R}^d$ distinct, there exists a symmetric transport plan of finite cost with marginals $\rho = b_1\delta_{x_1} + \cdots + b_k\delta_{x_k}$.

Proof. For every pair of positive integers (N, k), with $k > N$, let $\mathfrak{P}(N, k)$ be the following proposition:

> Let $(x_1, \ldots, x_k; b_1, \ldots, b_k)$ with $(N-1)b_1 \leq b_2 + \cdots + b_k$. Then for every $(x_1, \ldots, x_k)$ there exists a symmetric N-transport plan of finite cost with marginals $b_1\delta_{x_1} + \cdots + b_k\delta_{x_k}$.

We will prove $\mathfrak{P}(N, k)$ by double induction, in the following way: first we prove $\mathfrak{P}(1, k)$ for every k and $\mathfrak{P}(N, N+1)$ for every N. Then we prove

$$\mathfrak{P}(N-1, k) \wedge \mathfrak{P}(N, k-1) \implies \mathfrak{P}(N, k).$$

Proof of $\mathfrak{P}(1,k)$ This is trivial: simply take $b_1\delta_{x_1}+\cdots+b_k\delta_{x_k}$ as a "transport plan".

Proof of $\mathfrak{P}(N,N+1)$ Let us denote by A_N the $(N+1)\times(N+1)$ matrix

$$A_N = \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & & \ddots & \\ 1 & \cdots & 1 & 0 \end{pmatrix},$$

whose inverse is

$$A_N^{-1} = \frac{1}{N}\begin{pmatrix} -(N-1) & 1 & \cdots & 1 \\ 1 & -(N-1) & \cdots & 1 \\ \vdots & & \ddots & \\ 1 & \cdots & 1 & -(N-1) \end{pmatrix}$$

Define also the following $(N+1)\times N$ matrix, with elements in $\mathbb{R}^d$:

$$(x_{ij}) = \begin{pmatrix} x_2 & x_3 & \cdots & x_{N+1} \\ x_1 & x_3 & \cdots & x_{N+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & \cdots & x_N \end{pmatrix},$$

where the i-th row is $(x_1,\dots,x_{i-1},x_{i+1},\dots,x_{N+1})$. We want to construct a transport plan of the form

$$P = N\sum_{i=1}^{N+1} a_i(\delta_{x_{i1}}\otimes\cdots\otimes\delta_{x_{iN}})_{sym},$$

where $a_i \geq 0$. Note that, by Lemma 1.1.1, the marginals of P are equal to

$$\pi(P) = \sum_{j=1}^{N+1}\left(\sum_{\substack{i=1\\ i\neq j}}^{N+1} a_i\right)\delta_{x_j}.$$

Thus, the condition on the a_i's to have $\pi(P)=\rho$ is

$$A_N\begin{pmatrix} a_1 \\ \vdots \\ a_{N+1}\end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_{N+1}\end{pmatrix},$$

i.e.,

$$\begin{pmatrix} a_1 \\ \vdots \\ a_{N+1}\end{pmatrix} = A_N^{-1}\begin{pmatrix} b_1 \\ \vdots \\ b_{N+1}\end{pmatrix}.$$

Finally, observe that the condition (1.9) implies that $a_1 \geq 0$, while the fact that $b_1 \geq b_2 \geq \cdots \geq b_{N+1}$ leads to $a_1 \leq a_2 \leq \cdots \leq a_{N+1}$, and hence $a_i \geq 0$ for every i and we are done.

Inductive step Let $(b_1, \dots, b_k)$ satisfying (1.9), with $k \geq N+2$ (otherwise we are in the case $\mathfrak{P}(N, N+1)$, already proved). Take $t_2, \dots, t_k$ given by Lemma 1.4.4, and apply the inductive hypotheses to find

- a symmetric transport plan Q_1 of finite cost in $(N-1)$ variables, with marginals
$$\pi(Q_1) = \sum_{j=2}^{k} t_j \delta_{x_j};$$
- a symmetric transport plan R of finite cost in N variables, with marginals
$$\pi(R) = \sum_{j=2}^{k} (b_j - t_j) \delta_{x_j}.$$

Define
$$Q = \frac{1}{N-1} \sum_{j=1}^{N} (Q_1 \otimes_j \delta_{x_1}).$$
Since Q_1 is symmetric, Q is symmetric. Moreover, using Lemma 1.4.4 (i),
$$\pi(Q) = \frac{1}{N-1} \delta_{x_1} \sum_{j=2}^{k} t_j + \sum_{j=2}^{k} t_j \delta_{x_j} = b_1 \delta_{x_1} + \sum_{j=2}^{k} t_j \delta_{x_j}.$$

The transport plan $P = Q + R$ is symmetric, with marginals $\pi(P) = b_1 \delta_{x_1} + \cdots + b_k \delta_{x_k}$.

□

In order to conclude the proof of this Section, we must now deal not only with the non-atomic part of ρ, but also with the additional hypothesis of Proposition 1.4.5. Indeed, the presence of a non-atomic part will fix the atomic mass exceeding the inequality (1.10), as will be seen soon.

Definition 3. Given N, we say that the tuple $(b_1, \dots, b_\ell)$ is *fast decreasing* if
$$(N-j) b_j > \sum_{i=j+1}^{\ell} b_i \qquad \forall j = 1, \dots, \ell - 1.$$

Remark 1. Note that if $(b_1, \dots, b_\ell)$ is fast decreasing, then necessarily $\ell < N$. As a consequence, given any sequence $(b_1, b_2, \dots)$, even infinite, we may select its maximal fast decreasing initial tuple $(b_1, \dots, b_\ell)$ (which might be empty, *i.e.*, $\ell = 0$).

Theorem 1.4.6. *If ρ is such that*

$$\rho = \sigma + \sum_{j=1}^{k} b_j \delta_{x_j}$$

with $k > N$ atoms, then there exists a transport plan of finite cost.

Proof. Consider $(b_1, \ldots, b_k)$ and use the Remark 1 to select its maximal fast decreasing initial tuple $(b_1, \ldots, b_\ell)$, $\ell < N$. Thanks to Proposition 1.4.5, we may construct a transport plan $P_{\ell+1}$ over $\mathbb{R}^{(N-\ell)d}$ with marginals $b_{\ell+1}\delta_{x_{\ell+1}} + \cdots + b_k\delta_{x_k}$, since

$$(N-\ell-1)b_{\ell+1} \leq \sum_{j=\ell+2}^{k} b_j$$

by maximality of $(b_1, \ldots, b_\ell)$ — and this is condition (1.9) in this case. We extend step by step $P_{\ell+1}$ to an N-transport plan, letting

$$P_j = \frac{1}{N-j} \sum_{i=j}^{N} (P_{j+1} \otimes_i \delta_{x_j}),$$

for $j = \ell, \ell-1, \ldots, 1$.

Let $p_\ell = b_{\ell+1} + \cdots + b_k$, and $q_\ell = \frac{p_\ell}{N-\ell}$. We claim that $|P_j| = (N-j+1)q_\ell$. In fact, by construction $|P_{\ell+1}| = p_\ell$, and inductively

$$|P_j| = \frac{1}{N-j} \sum_{i=j-1}^{N} |P_{j+1}| = \frac{N-j+1}{N-j}(N-j)q_\ell = (N-j+1)q_\ell.$$

Moreover,

$$\pi(P_j) = \sum_{i=j}^{k} q_\ell \delta_{x_i} + \sum_{i=\ell+1}^{k} b_i \delta_{x_i}.$$

This is true by construction in the case $j = \ell+1$, and inductively

$$\pi(P_j) = \frac{1}{N-j}\delta_{x_j} |P_{j+1}| + \frac{N-j}{N-j}\pi(P_{j+1}) = \sum_{i=j}^{\ell} q_\ell \delta_{x_i} + \sum_{i=\ell+1}^{k} b_i \delta_{x_i}.$$

Note that, for every $i = 1, \ldots, \ell$, $b_i \geq b_\ell > q_\ell$. We shall find, using Proposition 1.4.2, a transport plan of finite cost with marginals

$$\sigma + \sum_{i=1}^{\ell} (b_i - q_\ell)\delta_{x_i},$$

since the condition (1.8) reads

$$N(b_1 - q_\ell) - \sum_{i=1}^{\ell}(b_i - q_\ell) = Nb_1 - \sum_{i=1}^{\ell} b_i - (N-\ell)q_\ell < 1 - \sum_{i=1}^{k} b_i = |\sigma|. \quad \square$$

1.5 Marginals with countably many atoms

In this Section we finally deal with the case of an infinite number of atoms, *i.e.*,

$$\rho = \sigma + \sum_{j=1}^{\infty} b_j \delta_{x_j}$$

with $b_j > 0$, $b_{j+1} \leq b_j$ for every $j \geq 1$.

The main issue is of topological nature: if the atoms x_j are too close each other (for example, if they form a dense subset of $\mathbb{R}^d$) and the growth of b_j for $j \to \infty$ is too slow, the cost might diverge. With this in mind, we begin with an elementary topological result, in order to separate the atoms in N groups, with controlled minimal distance from each other.

Lemma 1.5.1. *There exists a partition $\mathbb{R}^d = E_2 \sqcup \cdots \sqcup E_{N+1}$ such that:*

(i) for every $j = 2, \ldots, N+1$, $x_j \in \mathring{E}_j$;

(ii) for every $j = 2, \ldots, N+1$, ∂E_j does not contain any x_i.

Proof. For $j = 3, \ldots, N+1$ let $r_j > 0$ small enough such that

$$x_i \notin B(x_j, r_j) \quad \text{for every } i = 1, \ldots, N,\ i \neq j.$$

Fixed any $j = 3, \ldots, N+1$, by a cardinality argument there must be a positive real t_j with $0 < t_j < r_j$ and $\partial B(x_j, t_j)$ not containing any x_i, $i \geq 1$. We take $E_j = B(x_j, t_j)$ for $j = 3, \ldots, N+1$. Note that this choice fullfills the conditions (i), (ii) for $j = 3, \ldots, N+1$. Finally, we take

$$E_2 = \mathbb{R}^d \setminus \left(\bigcup_{j=3}^{N+1} E_j \right)$$

Clearly $x_2 \in \mathring{E}_2$, and moreover the condition (ii) is satisfied, since

$$\partial E_2 = \bigcup_{j=3}^{N+1} \partial E_j. \qquad \square$$

Consider the partition given by Lemma 1.5.1, and define the corresponding partition of $\mathbb{N}$ given by $\mathbb{N} = A_2 \cup \cdots \cup A_{N+1}$, where

$$A_j = \{ i \in \mathbb{N} \mid x_i \in E_j \} .$$

Next we consider, for every $j = 2, \ldots, N+1$ a threshold $n_j \geq 2$ large enough such that, defining

$$\epsilon_j = \sum_{\substack{i \geq n_j \\ i \in A_j}} b_i,$$

then

$$\epsilon_2 + \cdots + \epsilon_{N+1} < \min\left\{b_{N+1}, \frac{1}{N} - b_1\right\}. \tag{1.11}$$

This may be done since the series $\sum b_i$ converges, and hence for every $j = 2, \dots, N+1$ the series

$$\sum_{i \in A_j} b_i$$

is convergent.

For every $j = 2, \dots, N+1$ define the following transport plan:

$$P_j = N\left[\left(\sum_{i \in A_j, i \geq n_j} b_i \delta_{x_i}\right) \otimes \delta_{x_2} \otimes \cdots \otimes \hat{\delta}_{x_j} \otimes \cdots \otimes \delta_{x_{N+1}}\right]_{sym},$$

and note that, by Lemma 1.1.1,

$$\pi(P_j) = \epsilon_j \sum_{\substack{h=2 \\ h \neq j}}^{N+1} \delta_{x_h} + \sum_{\substack{i \geq n_j \\ i \in A_j}} b_i \delta_{x_i}.$$

Then let

$$P_\infty = \sum_{j=2}^{N+1} P_j,$$

and observe that

$$\pi(P_\infty) = \sum_{j=2}^{N+1}\left(\sum_{\substack{i=2 \\ i \neq j}}^{N+1} \epsilon_i\right)\delta_{x_j} + \sum_{j=2}^{N+1} \sum_{\substack{i \geq n_j \\ i \in A_j}} b_i \delta_{x_i}.$$

Let now

$$\tilde{b}_i = \begin{cases} b_i - \displaystyle\sum_{\substack{h=2 \\ h \neq i}}^{N+1} \epsilon_h & \text{if } 2 \leq i \leq N+1 \\ 0 & \text{if } i \geq n_j \text{ and } i \in A_j \text{ for some } j = 2, \dots, N+1 \\ b_i & \text{otherwise.} \end{cases}$$

We are left to find a transport plan of finite cost with marginals

$$\sigma + \sum_{i=1}^{\infty} \tilde{b}_i \delta_{x_i},$$

which has indeed a finite number of atoms. Note that $\tilde{b}_i \geq 0$ for every i, thanks to condition (1.11). Moreover, since $\tilde{b}_1 = b_1$ and $\tilde{b}_j \leq b_j$, then $\tilde{b}_1 \geq \tilde{b}_j$ for every $j \in \mathbb{N}$, as is used in what follows. If

$$(N-1)\tilde{b}_1 \leq \sum_{i=2}^{\infty} \tilde{b}_i$$

we may conclude using Proposition 1.4.5. Otherwise, we proceed like in the proof of Theorem 1.4.6, with $\{\tilde{b}_j\}$ replacing $\{b_j\}$. At the final stage, it is left to check that

$$N(\tilde{b}_1 - \tilde{q}_{k+1}) - \sum_{i=1}^{k} (\tilde{b}_i - \tilde{q}_{k+1}) < 1 - \sum_{i=1}^{\infty} b_i = |\sigma| \, .$$

Indeed this is true, since using the condition (1.11) one gets

$$N(\tilde{b}_1 - \tilde{q}_{k+1}) - \sum_{i=1}^{k} (\tilde{b}_i - \tilde{q}_{k+1}) = N b_1 - \sum_{i=1}^{\infty} b_i + N(\epsilon_2 + \ldots + \epsilon_{N+1}) < 1 - \sum_{i=1}^{\infty} b_i.$$

Chapter 2

Monge solutions for spherical densities

We consider a multi-marginal optimal transport problem with Coulomb cost in the following setting: let $d = 2$, $N = 3$ and take $\tilde{\rho} \in \mathcal{P}(\mathbb{R}^2)$ radially distributed. To be formally precise, this means that there exists a positive measure $\rho \in \mathcal{P}([0, \infty))$ such that

$$\int_{\mathbb{R}^2} \phi(x, y) d\tilde{\rho}(x, y) = \int_0^{+\infty} \left(\frac{1}{2\pi} \int_0^{2\pi} \phi(r, \theta) d\theta \right) d\rho(r)$$

for every $\phi \in C_b(\mathbb{R}^2)$. If $\tilde{\rho}$ is absolutely continuous, this means that the density of $\tilde{\rho}$ is a radial function.

As well discussed in [37], in this case the multi-marginal optimal transport problem reduces to a one-dimensional one. To make this notion precise, we define the *radial cost* $c\colon (\mathbb{R}_+)^3 \to \mathbb{R} \cup \{+\infty\}$,

$$c(r_1, r_2, r_3) = \min \{\tilde{c}(v_1, v_2, v_3) \mid |v_i| = r_i \text{ for } i = 1, 2, 3\} \qquad \text{for all } (r_1, r_2, r_3) \in (\mathbb{R}_+)^3.$$

For a given triple (r_1, r_2, r_3) there exist many differently-oriented vectors (v_1, v_2, v_3) that realize the above minimum. Once a triple of minimizers (v_1, v_2, v_3) has been fixed, the optimal configuration can be characterized by giving the radii and the angles between them. We may always assume that the vector v_1 lies along the positive x-axis, as the Coulomb cost function is invariant under the action of SO(2) on $\mathbb{R}^2$, *i.e.*,

$$c(x_1, x_2, x_3) = c(A(x_1), A(x_2), A(x_3)) \quad \forall A \in \mathrm{SO}(2).$$

With this choice in mind we denote by θ_2 the angle between v_1 and v_2 and by θ_3 the angle between v_1 and v_3. For this radial and angular data that corresponds to the triple of vectors $(v_1, v_2, v_3) \in (\mathbb{R}^2)^3$ we will sometimes use the notation $C(r_1, r_2, r_3, \theta_2, \theta_3)$ for the Coulomb cost $\tilde{c}(v_1, v_2, v_2)$. This allows to rewrite the radial cost function c as

$$c(r_1, r_2, r_3) = \min_{(\theta_2,\theta_3)\in\mathbb{T}^2} C(r_1, r_2, r_3, \theta_2, \theta_3) \tag{2.1}$$

Now solving the (MK) problem for the Coulomb cost and the marginal measure $\tilde{\rho}$ is equivalent to solving the one-dimensional (MK) problem in the class $\Pi_3(\rho)$ for the radial density ρ and the radial cost c, as will be made more rigorous in the next theorem, first proven by Pass (see [37]).

Theorem 2.0.1. *The full (MK) problem for the Coulomb cost*

$$\min\left\{\int_{(\mathbb{R}^2)^3} \tilde{c}(v_1, v_2, v_3)\, d\tilde{\gamma}(v_1, v_2, v_3) \mid \tilde{\gamma} \in \Pi_3(\tilde{\rho})\right\} \tag{2.2}$$

and the corresponding radial problem

$$\min\left\{\int_{(\mathbb{R}_+)^3} c(r_1, r_2, r_3)\, d\gamma \mid \gamma \in \Pi_3(\rho)\right\} \tag{2.3}$$

are equivalent in the following sense: the measure $\gamma \in \Pi_3(\rho)$ is optimal for the problem (2.3) if and only if the measure

$$\tilde{\gamma} := \gamma(r_1, r_2, r_3) \otimes \mu^{r_1,r_2,r_3}$$

is optimal for the problem (2.2). Above, μ^{r_1,r_2,r_3} is the singular probability measure on the 3-dimensional torus defined by

$$\mu^{r_1,r_2,r_3} = \frac{1}{2\pi}\int_0^{2\pi} \delta_t \delta_{\theta_2+t}\delta_{\theta_3+t}\, dt,$$

where (θ_2, θ_3) are minimizing angles $\theta_2 = \angle(v_1, v_2)$, $\theta_3 = \angle(v_1, v_3)$ for

$$c(r_1, r_2, r_3) = \min\left\{\tilde{c}(v_1, v_2, v_3) \mid |v_i| = r_i \text{ for } i = 1, 2, 3\right\}.$$

In [40] the authors conjectured the solution to the radial problem (2.3). The conjecture is stated for all d and N but for the sake of clarity we formulate it here for $N = 3$.

Conjecture 2.0.2 (Seidl)**.** *Let $\tilde{\rho} \in \mathcal{P}(\mathbb{R}^d)$ be radially-symmetric with the radial density ρ. Let s_1 and s_2 be such that*

$$\rho([0, s_1)) = \rho([s_1, s_2)) = \rho([s_2, \infty)) = \frac{1}{3}.$$

We define the map $T : [0, \infty)$ to be the unique map that sends, in the way that preserves the density ρ,

- *the interval* $[0, s_1)$ *to the interval* $[s_1, s_2)$ *decreasingly,*
- *the interval* $[s_1, s_2)$ *to the half-line* $[s_2, \infty)$ *decreasingly, and*
- *the half-line* $[s_2, \infty)$ *to the interval* $[0, s_1)$ *increasingly.*

More formally, this map is defined as

$$T(x) = \begin{cases} F^{-1}\left(\frac{2}{3} - F(x)\right) & \textit{when } x \in [0, s_1) \\ F^{-1}\left(\frac{4}{3} - F(x)\right) & \textit{when } x \in [s_1, s_2) \\ F^{-1}\left(1 - F(x)\right) & \textit{when } x \in [s_2, \infty) \end{cases}$$

where F is the cumulative distribution function of ρ, that is, $F(r) = \rho([0, r))$.

Then the map T is optimal for the radial problem (2.3).

The map introduced in Conjecture 2.0.2 is also called "The Seidl map" or "the DDI map" where the letters DDI stand for *Decreasing, Decreasing, Increasing*, identifying the monotonicities in which the first interval is mapped on the second, the second on the third, and finally the third back on the first. In an analogous manner one can define maps with different monotonicities: III, IID, DDI and so on. Since the marginals of our MOT problem are all the same and equal to ρ, the only maps T that make sense satisfy $T^3 = Id$, which leads us to the so-called $\mathcal{T} := \{I, D\}^3$ class, first introduced by Colombo and Stra in [17]:

$$\mathcal{T} := \{III, DDI, DID, IDD\}.$$

In [17] the authors were the first to disprove the Seidl conjecture. They showed that for $N = 3$ and $d = 2$ the DDI map fails to be optimal if the marginal measure is concentrated on a very thin annulus. They also provided a positive example for the optimality of the DDI map: they constructed a density, concentrated on a union of three disjointed intervals the last of which is very far from the first two, so that the support of the transport plan given by the DDI map is c-cyclically monotone. On the other hand, in [24] De Pascale proved that also for the Coulomb cost the c-cyclical monotonicity implies optimality: this implication had been previously proven only for cost functions that can be bounded from above by a sum of ρ-integrable functions. Using these results and making the necessary passage between the radial problem (2.3) and the full problem (2.2) one gets the optimality of the DDI map for the example of Colombo and Stra. In [17] the authors also provided a counterexample for the non-optimality of all transport maps in the class $\mathcal{T}$.

We will address the connection between the density ρ and the optimality or non-optimality of the Seidl map for $d = 2$ and $N = 3$. Our main results are the following:

Theorem 2.0.3. *Let $\rho \in \mathcal{P}(\mathbb{R}_+)$ such that*

$$r_2(r_3 - r_1)^3 - r_1(r_3 + r_2)^3 - r_3(r_1 + r_2)^3 \geq 0 \tag{2.4}$$

for ρ-a.e. $(r_1, r_2, r_3) \in [0, s_1] \times [s_1, s_2] \times [s_2, s_3]$. Then the DDI map T provides an optimal Monge solution $\gamma = (Id, T, T^2)_\#(\rho)$ to the problem (2.3).

This theorem makes more quantitative the positive result of Colombo and Stra (see Remarks 3 and 4 for a more detailed description). Its proof also gives a necessary and sufficient condition for the radial Coulomb cost to coincide with a much simpler cost that corresponds to the situation where all three particles are aligned. More precisely, we show that

Theorem 2.0.4. *Let $0 < r_1 < r_2 < r_3$. Then $(\theta_2, \theta_3) = (\pi, 0)$ is optimal in (2.1) if and only if*

$$r_2(r_3 - r_1)^3 - r_1(r_3 + r_2)^3 - r_3(r_1 + r_2)^3 \geq 0.$$

Moreover, if (2.4) holds, $(\theta_2, \theta_3) = (\pi, 0)$ is the unique minimum point.

We continue by using this new condition to construct a wide class of counterexamples for the optimality of the Seidl map. This class contains densities that are rather physical, such as positive, continuous and differentiable.

Theorem 2.0.5. *Let $\rho \in \mathcal{P}(\mathbb{R}_+)$ positive everywhere such that $\frac{s_1}{s_2} > \frac{1+2\sqrt{3}}{5}$ and*

$$T(x)(T^2(x) - x)^3 - x(T^2(x) + T(x))^3 - T^2(x)(x + T(x))^3 \geq 0 \tag{2.5}$$

for ρ-a.e. $x \in (0, s_1)$, where T is the DDI map. Then the DDI maps does not *provide an optimal Monge solution $\gamma = (Id, S, S^2)_\#(\rho)$ to the problem (2.3). Moreover, there exist smooth counterexample densities.*

Structure of the chapter In section 2.1 we will present the proof of Theorem 2.0.4, which is quite technical and long. This will require a careful study of the stationary points for the radial cost $C(r_1, r_2, r_3, \theta_2, \theta_3)$ for fixed radii and variable angles $(\theta_2, \theta_3) \in \mathbb{T}^2$.

In section 2.2 we will prove the main Theorems Theorem 2.0.3 and Theorem 2.0.5, by applying the result of Theorem 2.0.4 and other techniques, both original and derived from the literature concerning the multi-marginal Optimal Transport with Coulomb cost in dimension 1.

2.1 Proof of Theorem 2.0.4

Let $0 < r_1 < r_2 < r_3$ be fixed. In order to lighten the notation, we will omit the dependence on the radii when possible. We will also introduce the following functions for $i, j \in \{1, 2, 3\}$ and $\theta \in \mathbb{T}^1$:

$$D_{ij}(\theta) = r_i^2 + r_j^2 - 2r_i r_j \cos\theta, \quad F_{ij}(\theta) = \frac{1}{D_{ij}(\theta)^{1/2}}$$

It will be useful to compute the derivatives of F_{ij}, so we do it now:

$$\begin{aligned} F'_{ij}(\theta) &= -\frac{r_i r_j \sin\theta}{D_{ij}^{3/2}} \\ F''_{ij}(\theta) &= -\frac{r_i r_j \cos\theta}{D_{ij}^{3/2}} + \frac{3}{2}\frac{2r_i^2 r_j^2 \sin^2\theta}{D_{ij}^{5/2}} \\ &= -\frac{r_i r_j (r_i r_j \cos^2\theta + (r_i^2 + r_j^2)\cos\theta - 3r_i r_j)}{D_{ij}(\theta)^{5/2}} \end{aligned}$$

In order to simplify the notation even more, we denote

$$Q_{ij}(t) = r_i r_j t^2 + (r_i^2 + r_j^2)t - 3r_i r_j, \quad t \in [-1, 1],$$

so that

$$F''_{ij}(\theta) = -\frac{r_i r_j Q_{ij}(\cos\theta)}{D_{ij}(\theta)^{5/2}}.$$

Observe that $Q_{ij}(-1) = -(r_i + r_j)^2$ and $Q_{ij}(1) = (r_i - r_j)^2$, so that

$$F''_{ij}(0) = -\frac{r_i r_j}{|r_i - r_j|^3} \quad \text{and} \quad F''_{ij}(\pi) = \frac{r_i r_j}{(r_i + r_j)^3} \tag{2.6}$$

First we prove that if $(\alpha, \beta) = (\pi, 0)$ is optimal in (2.1), then (2.4) holds. Recall that the function to minimize is

$$f(\alpha, \beta) = F_{12}(\alpha) + F_{13}(\beta) + F_{23}(\alpha - \beta)$$

and notice that $f \in C^\infty(\mathbb{T}^2)$. Thus, if $(\pi, 0)$ is minimal, it must be a stationary point with positive-definite Hessian. Let us compute the gradient and the Hessian of f:

$$\begin{aligned} \nabla f(\alpha, \beta) &= (F'_{12}(\alpha) + F'_{23}(\alpha - \beta), F'_{13}(\beta) - F'_{23}(\alpha - \beta), \\ Hf(\alpha, \beta) &= \begin{pmatrix} F''_{12}(\alpha) + F''_{23}(\alpha - \beta) & -F''_{23}(\alpha - \beta) \\ -F''_{23}(\alpha - \beta) & F''_{13}(\beta) + F''_{23}(\alpha - \beta) \end{pmatrix}. \end{aligned}$$

Using (2.6), we have

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

and

$$\begin{aligned} \det Hf(\pi, 0) &= -\frac{r_1^2 r_2 r_3}{(r_1 + r_2)^3 (r_3 - r_1)^3} - \frac{r_1 r_2 r_3^2}{(r_2 + r_3)^3 (r_3 - r_1)^3} \\ &\quad + \frac{r_1 r_2^2 r_3}{(r_1 + r_2)^3 (r_2 + r_3)^3} \\ &= \frac{r_1 r_2 r_3 [r_2 (r_3 - r_1)^3 - r_1 (r_3 + r_2)^3 - r_3 (r_1 + r_2)^3]}{(r_1 + r_2)^3 (r_2 + r_3)^3 (r_3 - r_1)^3}. \end{aligned}$$

The positivity of $\det Hf(\pi, 0)$ implies the condition (2.4), which proves the first part of 2.0.4.

Now we assume that (2.4) holds, and we want to get that $(\pi, 0)$ is the unique minimum point. The first (and most challenging) step is given by the following

Proposition 2.1.1. *Suppose that $0 < r_1 < r_2 < r_3$ satisfy (2.4). Then $(0,0), (0,\pi), (\pi,0), (\pi,\pi)$ are the only stationary points of $f(\alpha, \beta)$.*

The proof of 2.1.1 is quite technical and long. For the sake of clarity we postpone it to the end of this section, in order to keep focusing on the main result.

Since $\{0, \pi\}^2$ are the only stationary points, the global minimum of f must be between them. By direct comparison of the values $f(0,0)$, $f(0,\pi)$, $f(\pi,0)$, $f(\pi,\pi)$ we will conclude that $(\pi,0)$ is the unique minimum point.

We compute

$$\begin{aligned} f(0,0) &= \frac{1}{r_2 - r_1} + \frac{1}{r_3 - r_2} + \frac{1}{r_3 - r_1} \\ f(0,\pi) &= \frac{1}{r_2 - r_1} + \frac{1}{r_3 + r_2} + \frac{1}{r_3 + r_1} \\ f(\pi,0) &= \frac{1}{r_2 + r_1} + \frac{1}{r_3 + r_2} + \frac{1}{r_3 - r_1} \\ f(\pi,\pi) &= \frac{1}{r_2 + r_1} + \frac{1}{r_3 - r_2} + \frac{1}{r_3 + r_1}, \end{aligned}$$

and observe that clearly $f(0,0) > f(0,\pi)$. To deduce the other inequalities we notice that the function

$$h(x,y) = \frac{1}{x-y} - \frac{1}{x+y} \quad \text{for } 0 < y < x.$$

is decreasing in x and increasing in y, so $h(r_3, r_1) < h(r_2, r_1) \Rightarrow f(\pi, 0) < f(0, \pi)$ and $h(r_3, r_1) < h(r_3, r_2) \Rightarrow f(\pi, 0) < f(\pi, \pi)$, as wanted.

Proof of 2.1.1. A stationary point (α, β) must solve $\nabla f = 0$, *i.e.*,

$$\begin{cases} -\dfrac{r_1 r_2 \sin\alpha}{D_{12}(\alpha)^{3/2}} - \dfrac{r_2 r_3 \sin(\alpha - \beta)}{D_{23}(\alpha - \beta)^{3/2}} = 0 \\ -\dfrac{r_1 r_3 \sin\beta}{D_{13}(\beta)^{3/2}} + \dfrac{r_2 r_3 \sin(\alpha - \beta)}{D_{23}(\alpha - \beta)^{3/2}} = 0 \end{cases}$$

which we rewrite as

$$\begin{cases} \dfrac{r_1 r_2 \sin\alpha}{D_{12}(\alpha)^{3/2}} + \dfrac{r_1 r_3 \sin\beta}{D_{13}(\beta)^{3/2}} = 0 \\ \dfrac{r_1 r_3 \sin\beta}{D_{13}(\beta)^{3/2}} - \dfrac{r_2 r_3 \sin(\alpha - \beta)}{D_{23}(\alpha - \beta)^{3/2}} = 0. \end{cases} \tag{2.7}$$

Observe that the four points $(\alpha, \beta) \in \{0, \pi\}^2$ are always solutions for (2.7). We will study this system in detail for $\beta \in [0, \pi]$. The conclusions can then be derived for $\beta \in [-\pi, 0]$ by making use of the change of variables $\tilde{\alpha} = -\alpha$, $\tilde{\beta} = -\beta$. To proceed in the computations, we perform a finer study of the function

$$g_{ij}(\theta) = -F'_{ij}(\theta) = \frac{r_i r_j \sin\theta}{D_{ij}(\theta)^{3/2}}\,,$$

so that the optimality conditions (2.7) will be rewritten in the form

$$\begin{cases} g_{12}(\alpha) = -g_{13}(\beta) \\ g_{13}(\beta) = g_{23}(\alpha - \beta). \end{cases} \tag{2.8}$$

We now prove that for every β in $[0, \pi]$ there exists at least one and at most two α's such that each of the two equations is satisfied.

The derivative of g_{ij} is

$$g'_{ij}(\theta) = r_i r_j \frac{Q_{ij}(\cos\theta)}{D_{ij}(\theta)^{5/2}}$$

and it vanishes for

$$Q_{ij}(\cos\theta_{ij}) = 0 \Rightarrow \cos\theta_{ij} = \frac{-r_i^2 - r_j^2 + \sqrt{r_i^4 + 14 r_i^2 r_j^2 + r_j^4}}{2 r_i r_j} \in (0, 1).$$

By looking at the sign of the second degree polynomial Q_{ij}, we conclude that $g_{ij}(\theta)$ is increasing from 0 to its maximum on $[0, \theta_{ij}]$ and decreasing to 0 on $[\theta_{ij}, \pi]$

Lemma 2.1.2. *For every $\theta \in [0, \pi]$, $g_{13}(\theta) \leq g_{12}(\theta)$ and $g_{13}(\theta) \leq g_{23}(\theta)$. (See Fig. 2.1.)*

Proof. We claim that $0 \leq g'_{13}(0) \leq g'_{12}(0)$ and $g'_{13}(\pi) \geq g'_{12}(\pi) \geq 0$. Indeed, using 2.6,

$$g'_{13}(0) = -F''_{13}(0) = \frac{r_1 r_3}{(r_3 - r_1)^3} \geq 0, \quad \text{and} \quad g'_{12}(0) = \frac{r_1 r_2}{(r_2 - r_1)^3},$$

thus

$$g'_{13}(0) \leq g'_{12}(0) \iff r_3 (r_2 - r_1)^3 \leq r_2 (r_3 - r_1)^3$$

which is weaker than (2.4).

On the other hand,

$$g'_{13}(\pi) = -F''_{13}(\pi) = -\frac{r_1 r_3}{(r_3 + r_1)^3} \leq 0, \quad \text{and} \quad g'_{12}(\pi) = -\frac{r_1 r_2}{(r_1 + r_2)^3},$$

thus

$$g'_{13}(\pi) \geq g'_{12}(\pi) \iff r_3 (r_1 + r_2)^3 \leq r_2 (r_3 + r_1)^3$$

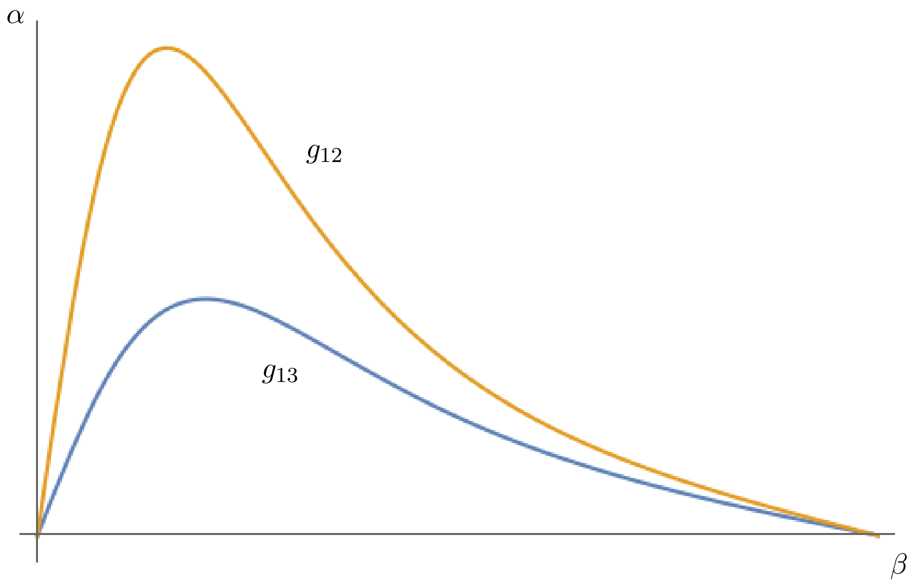

Figure 2.1: The relative position of the graphs of g_{12} and g_{13} on the interval $[0, \pi]$. However the strict inequality betwen the two maximal values is not proved. See Lemma 2.1.2

which is once again weaker than (2.4).

Moreover, the equation $g_{13}(\theta) = g_{12}(\theta)$ has at most one solution in $(0, \pi)$, since we have the following chain of equivalent equalities:

$$g_{13}(\theta) = g_{12}(\theta)$$
$$\frac{r_1 r_3}{D_{13}(\theta)^{3/2}} = \frac{r_1 r_2}{D_{12}(\theta)^{3/2}}$$
$$r_3^{2/3}(r_1^2 + r_2^2 - 2r_2 r_3 \cos\theta) = r_2^{2/3}(r_1^2 + r_3^2 - 2r_1 r_3 \cos\theta)$$

$$\cos\theta = \frac{r_2^{2/3} r_3^{2/3}(r_3^{4/3} - r_2^{4/3}) - r_1^2(r_3^{2/3} - r_2^{2/3})}{2r_1 r_2^{2/3} r_3^{2/3}(r_3^{1/3} - r_2^{1/3})}.$$

Recalling that both g_{13} and g_{12} vanish at the endpoints of $[0, \pi]$, we get the thesis. An analogous argument applies to the comparison between g_{13} and g_{23}. □

Remark 2. It follows from the Lemma above that for every value of g_{13}, and so for every fixed β, there exists at least one α where $g_{12}(\alpha)$ takes the same value. If the value of g_{13} is not the maximal one then there are exactly two different α's such that the value is achieved. The same holds for $g_{23}(\alpha - \beta)$. See figure below.

Lemma 2.1.3. *If* $\cos\theta \in (\cos\theta_{ij}, 1)$ *then*

$$g'_{ij}(\theta) < g'_{ij}(0)\frac{\cos\theta - \cos\theta_{ij}}{1-\cos\theta_{ij}};$$

if $\cos\theta \in (-1, \cos\theta_{ij})$ *then*

$$g'_{ij}(\theta) < g'_{ij}(\pi)\frac{\cos\theta - \cos\theta_{ij}}{-1-\cos\theta_{ij}}.$$

Proof. We omit for simplicity of notation the indices ij. Recall that

$$g'(\theta) = \frac{r_i r_j Q_{ij}(\cos\theta)}{(r_i^2 + r_j^2 - 2r_i r_j\cos\theta)} = h(\cos\theta)$$

, where $h\colon [-1,1] \to \mathbb{R}$, $h(t) = \frac{r_i r_j Q_{ij}(t)}{(r_i^2+r_j^2-2r_ir_jt)}$.

The thesis is a weak version of the convexity of h: if h is convex, then the inequalities hold by applying the Jensen's inequality separately in the intervals $[-1, \cos\theta_{ij}]$ and $[\cos\theta_{ij}, 1]$. It could happen, however, that h has a concave part between -1 and a certain threshold ξ, and then it is convex. In this case we prove the following:

- h_{ij} is decreasing between -1 and a certain threshold σ, where it reaches the minimum;
- $\xi < \sigma$, i.e., in the interval $[\sigma, 1]$ the function is convex.

Then we deduce that, for $-1 \le t \le \sigma$,

$$h_{ij}(t) \le h_{ij}(-1) \le h_{ij}(-1)\frac{t-\cos\theta_{ij}}{-1-\cos\theta_{ij}}$$

(recall that $h(-1)$ is negative).

On the other hand, for $\sigma \le t \le \cos\theta_{ij}$,

$$\begin{aligned} h(t) &\le \text{line joining } (\sigma, h(\sigma)) \text{ and } (\cos\theta_{ij}, 0)\\ &\le \text{line joining } (-1, h(-1)) \text{ and } (\cos\theta_{ij}, 0) \end{aligned}$$

since σ is a minimum point. See Figure 2.2 for a clearer graphical meaning of the proof.

Here come the computations:

$$h'(t) = \frac{r_i^2 r_j^2 t^2 + 5r_i r_j(r_i^2 + r_j^2)t + r_i^4 - 13r_i^2 r_j^2 + r_j^4}{(r_i^2 + r_j^2 - 2r_i r_j t)^{7/2}}.$$

We have that $h'(t) = 0$ for

$$t = \frac{-5(r_i^2 + r_j^2) \pm \sqrt{21r_i^4 + 102r_i^2 r_j^2 + 21r_i^4}}{2r_i r_j}.$$

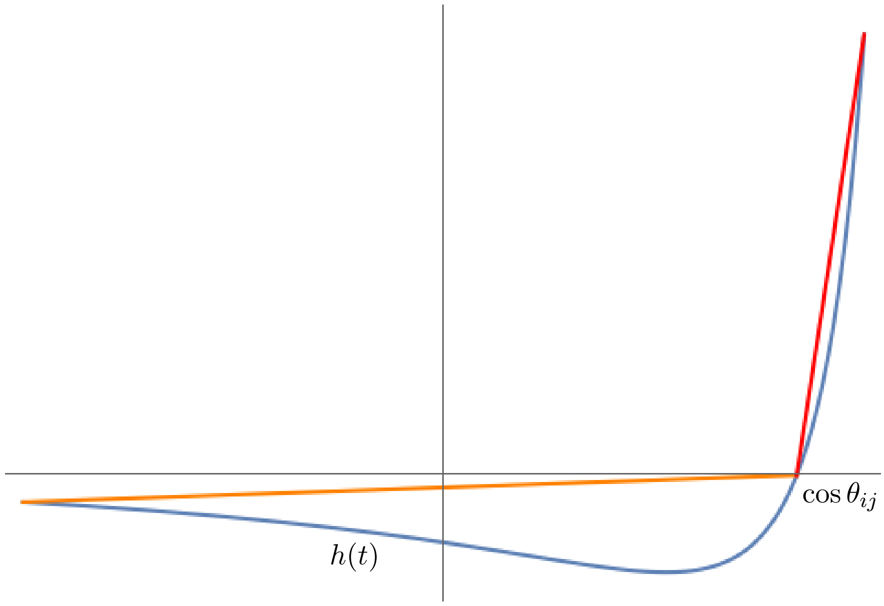

Figure 2.2: A graphical understanding of Lemma 2.1.3: the function $h(t)$ stays below two segments.

Observe that the smaller solution is always outside the interval $[-1, 1]$, since

$$-5(r_i^2 + r_j^2) - \sqrt{21r_i^4 + 102r_i^2r_j^2 + 21r_i^4} < -\sqrt{102}r_ir_j < -2r_ir_j.$$

Denote by σ the bigger root.

We move on to the second derivative:

$$h''(t) = 3r_ir_j\frac{r_i^2r_j^2t^2 + 9r_ir_j(r_i^2 + r_j^2)t + 4r_i^4 - 27r_i^2r_j^2 + 4r_j^4}{(r_i^2 + r_j^2 - 2r_ir_jt)^{9/2}}$$

We have that $h''(t) = 0$ for

$$t = \frac{-9(r_i^2 + r_j^2) \pm \sqrt{65r_i^4 + 270r_i^2r_j^2 + 65r_j^4}}{2r_ir_j}.$$

As above, the smaller root always lies outside the interval $[-1, 1]$. Denote by ξ the bigger root. Now we prove that $\sigma > \xi$ for any choice of the values $0 < r_i < r_j$. By homogeneity, denoting by $u = r_i^2/r_j^2 \in (0, 1)$, it suffices to prove that

$$-5(1 + u) + \sqrt{21u^2 + 102u + 21} > -9(1 + u) + \sqrt{65u^2 + 270u + 65},$$

i.e.,

$$\begin{gathered}
4(1+u)+\sqrt{21u^2+102u+21} > \sqrt{65u^2+270u+65} \\
37u^2+134u+37u+8(1+u)\sqrt{21u^2+102u+21} > 65u^2+270u+65 \\
8(1+u)\sqrt{3}\sqrt{7u^2+34u+7} > 28u^2+136u+28 \\
2\sqrt{3}(1+u) > \sqrt{7u^2+34u+7} \\
12(1+u)^2 > 7u^2+34u+7 \\
5(1-u)^2 > 0,
\end{gathered}$$

as wanted. □

Now the idea is the following: in view of Lemma 2.1.2, the first equation of (2.7) implicitly defines two C^∞ functions $\alpha_0(\beta)$ and $\alpha_\pi(\beta)$ such that $\alpha_0(0)=0, \alpha_\pi(0)=\pi$. Analogously, the second equation implicitly defines two functions $\widehat{\alpha}_0(\beta)$ and $\widehat{\alpha}_\pi(\beta)$ such that $\widehat{\alpha}_0(0)=0, \widehat{\alpha}_\pi(0)=\pi$.

We want to prove that each curve $\alpha_{0,\pi}$ intersects each curve $\widehat{\alpha}_{0,\pi}$ only in 0 or π. By sign considerations, we notice that the first equation implies $\alpha(\beta)\in[\pi,2\pi]$ and the second equation implies $\widehat{\alpha}(\beta)\in[\beta,\pi+\beta]$. Hence, the possible solutions lie in the region $\pi\le\alpha\le\pi+\beta$, and when considering the whole torus $\mathbb{T}^2$ the region has a "butterfly" shape.

This already shows that the curves $\alpha_0(\beta)$ and $\widehat{\alpha}_0(\beta)$ do not produce solutions, since we have that $\beta-\pi\le\alpha_0(\beta)\le 0$ and $0\le\widehat{\alpha}_0(\beta)\le\pi$. Thus we can concentrate our attention on the curves α_π and $\widehat{\alpha}_\pi$.

The key observation lies in the fact that

$$\pi\le\alpha_\pi(\beta)\le\pi+\alpha'_\pi(0)\beta,$$

i.e., the function $\alpha_\pi(\beta)$ stays below its tangent line at $\beta=0$ (see Picture 2.4). Likewise, the function $\widehat{\alpha}_\pi(\beta)$ stays above its tangent line at $\beta=0$. This allows us to conclude that they do not intersect since, as we will see, the condition (2.4) is equivalent to $\alpha'_\pi(0)\le\widehat{\alpha}'_\pi(0)$.

Lemma 2.1.4. *For $\beta\in(0,\pi)$ let $\alpha(\beta)$ be the solution of*

$$\begin{cases} g_{13}(\beta)+g_{ij}(\alpha)=0 \\ \alpha(0)=\alpha(\pi)=\pi. \end{cases}$$

Then

$$\pi\le\alpha(\beta)<\pi+\alpha'(0)\beta.$$

(See Figure 2.4 for a graphical understanding.)

Proof. Differentiating in β we get

$$g'_{13}(\beta)+\alpha'(\beta)g'_{ij}(\alpha(\beta))=0 \implies \alpha'(\beta)=\frac{g'_{13}(\beta)}{-g'_{ij}(\alpha(\beta))}.$$

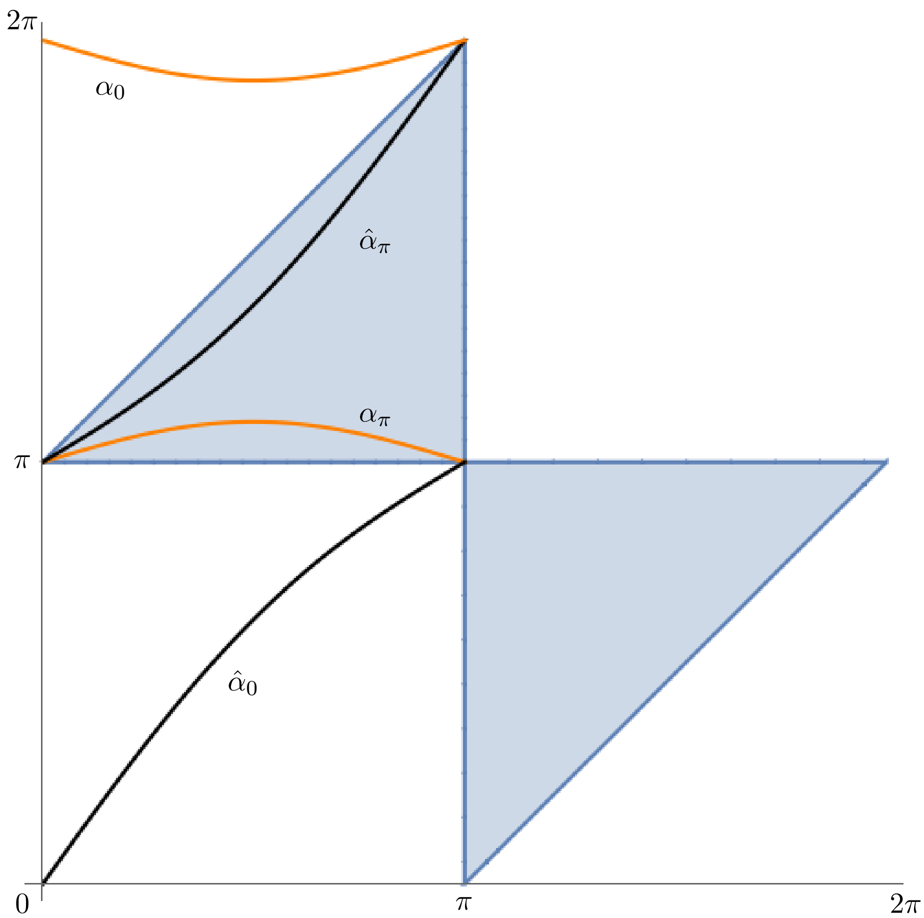

Figure 2.3: In blue, the "butterfly" region of admissible solutions to optimality conditions (2.7). In black and orange, a plot of the curves $\alpha_{0,\pi}$ and $\hat{\alpha}_{0,\pi}$ in the region $0 \leq \beta \leq \pi$.

Take $\beta \in (0, \theta_{13})$, where θ_{13} is the critical value of g_{13}, so that $\cos\beta > \cos\theta_{13}$. By Lemma 2.1.2 we have that $\alpha \in [\pi, 2\pi - \theta_{ij}]$, because the equation $g_{13}(\beta) + g_{ij}(\alpha) = 0$ has two solutions in the interval $[\pi, 2\pi]$ and by definition α is the leftmost one. Using Lemma 2.1.3 we have

$$\alpha'(\beta) \leq \frac{g'_{13}(0)}{-g'_{ij}(\pi)} \frac{\cos\beta - \cos\theta_{13}}{1 - \cos\theta_{13}} \frac{-1 - \cos\theta_{ij}}{\cos\alpha(\beta) - \cos\theta_{ij}}.$$

Since $\frac{g'_{13}(0)}{-g'_{ij}(\pi)} = \alpha'(0) \geq 0$, it suffices to show that

$$\frac{\cos\beta - \cos\theta_{13}}{1 - \cos\theta_{13}} \frac{-1 - \cos\theta_{ij}}{\cos\alpha(\beta) - \cos\theta_{ij}} \leq 1.$$

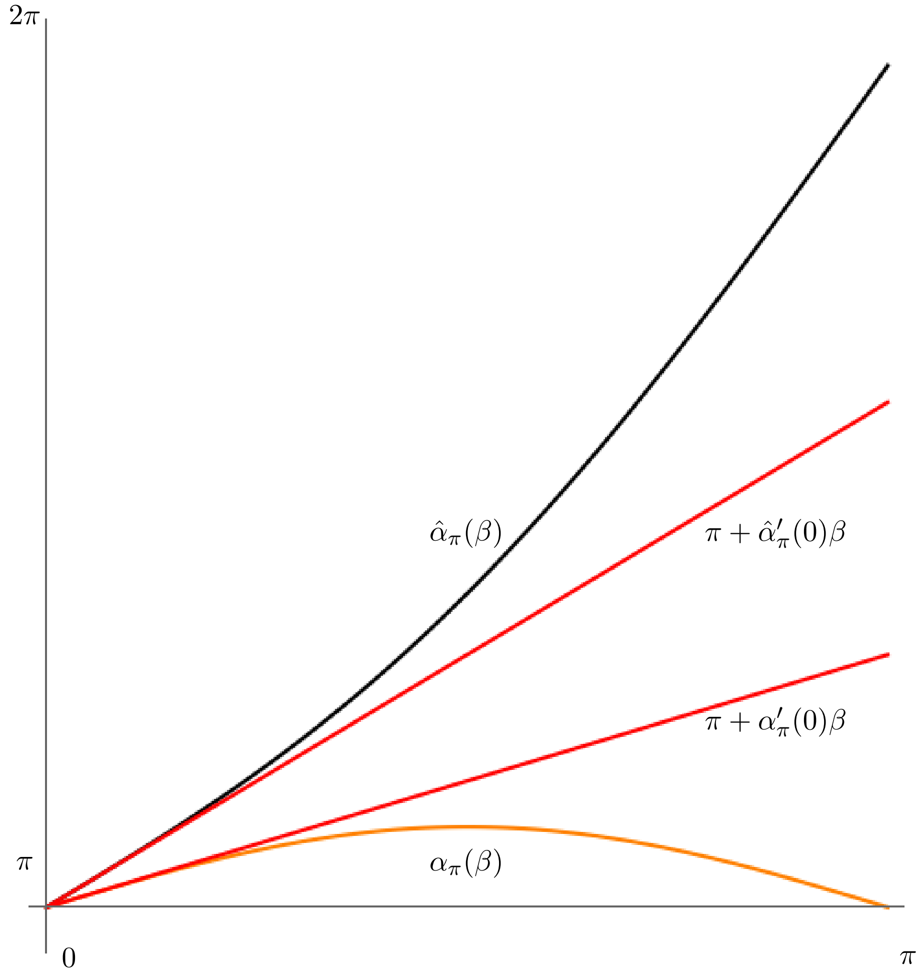

Figure 2.4: A graphical understanding of Lemma 2.1.4: the function $\alpha_\pi(\beta)$ is confined by $\pi \leq \alpha_\pi(\beta) \leq \pi+\alpha'_\pi(0)\beta$, and similarly $\pi+\widehat{\alpha}'_\pi(0)\beta \leq \widehat{\alpha}_\pi(\beta) \leq \pi+\beta$. This implies that the intersection between α_π and $\widehat{\alpha}_\pi$ is only at $\beta = 0$.

Let $\tilde{\alpha} = \alpha(\beta) - \pi$, so that $0 \le \tilde{\alpha} \le \beta$. We must prove

$$\begin{aligned}
&\frac{\cos\beta - \cos\theta_{13}}{1-\cos\theta_{13}} \frac{1+\cos\theta_{ij}}{\cos\tilde{\alpha} + \cos\theta_{ij}} \le 1. \\
&(1+\cos\theta_{ij})(\cos\beta - \cos\theta_{13}) \le (1-\cos\theta_{13})(\cos\tilde{\alpha} + \cos\theta_{ij}) \\
&(\cos\beta - \cos\tilde{\alpha}) + \cos\theta_{ij}\cos\beta + \cos\theta_{13}\cos\tilde{\alpha} \le \cos\theta_{ij} + \cos\theta_{13}.
\end{aligned}$$

But this is true, since $\tilde{\alpha} \le \beta \implies \cos\beta - \cos\tilde{\alpha} \ge 0$ and clearly

$$\cos\theta_{ij}\cos\beta + \cos\theta_{13}\cos\tilde{\alpha} \le \cos\theta_{ij} + \cos\theta_{13}.$$

We got the desired inequality for $\beta \in (0, \theta_{13})$. However, for $\beta \ge \theta_{13}$ we have $\alpha'(\beta) \le 0$, hence the line $\alpha'(0)\beta$ is increasing and the function $\alpha(\beta)$ is decreasing, giving the inequality for every β. □

By Lemma 2.1.4, we obtain that the function $\alpha_\pi(\beta)$ lies between the horizontal line $\alpha = \pi$ and the line $\alpha = \pi + \alpha'_\pi(0)\beta$ (strictly for $\beta > 0$). Recall that the function $\widehat{\alpha}_\pi(\beta)$ satisfies the second equation of the stationary system ((2.7))

$$\frac{r_1 r_3 \sin\beta}{D_{13}(\beta)^{3/2}} - \frac{r_2 r_3 \sin(\widehat{\alpha} - \beta)}{D_{23}(\widehat{\alpha} - \beta)^{3/2}} = 0$$

with $\widehat{\alpha}_\pi(0) = \pi$, $\widehat{\alpha}_\pi(\pi) = 2\pi$.

By a change of variables $\tilde{\alpha}(\beta) = 2\pi + \beta - \widehat{\alpha}_\pi(\beta)$, we get that $\tilde{\alpha}$ satisfies

$$\begin{cases} g_{13}(\beta) + g_{23}(\tilde{\alpha}) = 0 \\ \tilde{\alpha}(0) = \tilde{\alpha}(\pi) = 0, \end{cases}$$

hence $\pi < \tilde{\alpha}(\beta) < \pi + \tilde{\alpha}'(0)\beta$, i.e.,

$$\pi + \widehat{\alpha}'_\pi(0)\beta < \widehat{\alpha}_\pi(\beta) < \pi + \beta$$

for $\beta > 0$. So the idea is that the two lines provide a separation of the curves, so that no intersection can happen except at the starting point.

We conclude by observing that the condition (2.4) is equivalent to $\widehat{\alpha}'_\pi(0) \ge \alpha'_\pi(0)$: indeed we have

$$\begin{aligned}
&g_{13}(\beta) - g_{23}(\widehat{\alpha}_\pi(\beta) - \beta) = 0 \implies \\
&\widehat{\alpha}'_\pi(0) = 1 + \frac{g'_{13}(0)}{g'_{23}(\pi)} = 1 - \frac{r_1 r_3}{(r_3 - r_1)^3} \frac{(r_2 + r_3)^3}{r_2 r_3} = \frac{r_2(r_3 - r_1)^3 - r_1(r_2 + r_3)^3}{r_2(r_3 - r_1^3)}
\end{aligned}$$

and

$$\alpha'_\pi(0) = \frac{g'_{13}(0)}{-g'_{12}(\pi)} = \frac{r_1 r_3}{(r_3 - r_1)^3} \frac{(r_1 + r_2)^3}{r_1 r_2} = \frac{r_3(r_1 + r_2)^3}{r_2(r_3 - r_1)^3}.$$

□

Before coming to the consequences and proving the main results of this Chapter, let us present a couple of useful remarks.

Remark 3. We recall the polynomial condition (2.4):

$$r_2(r_3 - r_1)^3 - r_1(r_3 + r_2)^3 - r_3(r_1 + r_2)^3 \geq 0\,.$$

For fixed r_1 and r_2, the cubic polynomial in r_3 that appears on the left-hand side of (2.4) has three real roots. They are given by the following expressions:

$$-r_2, \qquad \frac{5r_1r_2 + r_2^2 \pm (r_1 + r_2)\sqrt{r_2^2 + 12r_1r_2 - 4r_1^2}}{2(r_2 - r_1)}.$$

Since we are only interested in the region where $r_3 > 0$ and since

$$\frac{5r_1r_2 + r_2^2 + (r_1 + r_2)\sqrt{r_2^2 + 12r_1r_2 - 4r_1^2}}{2(r_2 - r_1)}$$

is the only positive root for every value of $0 < r_1 < r_2$, the condition (2.4) can be rewritten as

$$\varphi(r_1, r_2) := \frac{5r_1r_2 + r_2^2 + (r_1 + r_2)\sqrt{r_2^2 + 12r_1r_2 - 4r_1^2}}{2(r_2 - r_1)} \leq r_3. \tag{2.9}$$

Remark 4. In [17], a crucial role was played by Lemma 4.1. In our framework this lemma can be obtained as a consequence of Theorem 2.0.4 by choosing (following the notation of [17])

$$r_3^- > \max_{[r_1^-, r_1^+]\times[r_2^-, r_2^+]} \varphi(r_1, r_2)\,.$$

If $r_1^+ < r_2^-$, as assumed by the authors in [17], then the maximum above is a real number and the threshold r_3^- can be fixed. Thus our result gives a quantitative optimal version of their choice. Moreover, Theorem 2.0.4 allows us to deal with the case in which there is no gap between r_1^+ and r_2^-, since we have an explicit control of the growth of $\varphi(r_1, r_2)$ as $r_1 \to r_2$.

2.2 Proofs of the main theorems

When ρ satisfies the assumptions of Theorem 2.0.4, we know that

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

for $(\rho\otimes\rho\otimes\rho)$-a.e. $(r_1, r_2, r_3) \in [0, s_1]\times[s_1, s_2]\times[s_2, +\infty)$. The key observation lies in the fact that this can be viewed as a 1-dimensional Coulomb cost for points $-r_2, r_1, r_3 \in \mathbb{R}$. We can now rely on a somewhat well-established theory for the Coulomb cost in dimension $d = 1$: see for instance [15, 14, 23, 25].

This allows to prove Theorem 2.0.3.

Proof of Theorem 2.0.3. This is a direct consequence of [14, Theorem 1.1]. Indeed, we can consider $\tilde{\rho} \in \mathcal{P}(\mathbb{R})$ the absolutely continuous measure defined by[1]

$$\tilde{\rho}(x) = \begin{cases} \rho(x) & x \in [0, s_1] \cup [s_2, +\infty) \\ \rho(-x) & x \in [-s_2, -s_1] \\ 0 & \text{otherwise} \end{cases}$$

and observe that the DDI map T for ρ corresponds to the optimal increasing map S defined in [14, Theorem 1.1].

The optimality follows from the fact that

$$c_r(x, T(x), T^2(x)) = c(y, S(y), S^2(y))$$

for ρ-a.e. $x \in [0, s_1]$ and $\tilde{\rho}$-a.e. $y \in [-s_2, -s_1]$, where c is the Coulomb cost, as observed above. □

The idea for the proof of Theorem 2.0.5 is to show that, on the support of the DDI map, the c_r-cyclical monotonicity is violated. We prepare a couple of technical results.

Lemma 2.2.1. *Let $\frac{s_1}{s_2} > \frac{1+2\sqrt{3}}{5}$. Then there exist $\epsilon, M > 0$ such that*

$$\frac{2}{s_2 + \epsilon} + \frac{1}{2s_2 + \epsilon} + \frac{1}{2s_1 + \epsilon} > \frac{\sqrt{3}}{s_1 - \epsilon} + \frac{1}{s_1} + \frac{1}{M - \epsilon}. \tag{2.10}$$

Proof. When $\epsilon = 0$ and $M = +\infty$, the inequality (2.10) reads

$$\frac{2}{s_2} + \frac{1}{2s_2} + \frac{1}{2s_1} > \frac{\sqrt{3}}{s_1} + \frac{1}{s_1},$$

which is equivalent to $\frac{s_1}{s_2} > \frac{2\sqrt{3}+1}{5}$.

By continuity, there is a small ϵ such that

$$\frac{2}{s_2 + \epsilon} + \frac{1}{2s_2 + \epsilon} + \frac{1}{2s_1 + \epsilon} > \frac{\sqrt{3}}{s_1 - \epsilon} + \frac{1}{s_1}.$$

Now choose M big enough such that the desired inequality (2.10) holds. □

Lemma 2.2.2. *Let s_1, s_2, ε and M as in Lemma 2.2.1, and let $(r_1, r_2, r_3) \in (0, \epsilon) \times (s_2 - \epsilon, s_2) \times (s_2, s_2 + \epsilon)$ and $(\ell_1, \ell_2, \ell_3) \in (s_1 - \epsilon, s_1) \times (s_1, s_1 + \epsilon) \times (M, +\infty)$. Suppose that the condition (2.9) is satisfied by both (r_1, r_2, r_3) and (ℓ_1, ℓ_2, ℓ_3). Then*

$$c_r(r_1, r_2, r_3) + c_r(\ell_1, \ell_2, \ell_3) > c_r(\ell_1, r_2, r_3) + c_r(r_1, \ell_2, \ell_3).$$

[1] for simplicity we denote again by $\rho(x)$ the density of the measure ρ.

Proof. Since the condition (2.9) is satisfied, we have

$$\begin{aligned} c_r(r_1,r_2,r_3) = c_\pi(r_1,r_2,r_3) &= \frac{1}{r_1+r_2} + \frac{1}{r_2+r_3} + \frac{1}{r_3-r_1} \\ &\geq \frac{1}{s_2+\epsilon} + \frac{1}{2s_2+\epsilon} + \frac{1}{s_2+\epsilon} \end{aligned}$$

and

$$\begin{aligned} c_r(\ell_1,\ell_2,\ell_3) = c_\pi(\ell_1,\ell_2,\ell_3) &= \frac{1}{\ell_1+\ell_2} + \frac{1}{\ell_2+\ell_3} + \frac{1}{\ell_3-\ell_1} \\ &\geq \frac{1}{2s_1+\epsilon} + \frac{1}{\ell_2+\ell_3} + 0. \end{aligned}$$

Now we analyze the other side. Since (ℓ_1,ℓ_2,ℓ_3) satisfy (2.9) and $r_1 < \ell_1$, then also (r_1,ℓ_2,ℓ_3) satisfy (2.9)[2], so that

$$\begin{aligned} c_r(r_1,\ell_2,\ell_3) = c_\pi(r_1,\ell_2,\ell_3) &= \frac{1}{r_1+\ell_2} + \frac{1}{\ell_2+\ell_3} + \frac{1}{\ell_3-r_1} \\ &\leq \frac{1}{s_1} + \frac{1}{\ell_2+\ell_3} + \frac{1}{M-\epsilon}. \end{aligned}$$

For the other term we have

$$c_r(\ell_1,r_2,r_3) \leq c_\Delta(\ell_1,r_2,r_3) \leq c_\Delta(\ell_1,\ell_1,\ell_1) = \frac{\sqrt{3}}{\ell_1} \leq \frac{\sqrt{3}}{s_1-\epsilon},$$

where $c_\Delta(r_1,r_2,r_3) = c(r_1,0,r_2,\frac{2\pi}{3},r_3,\frac{4\pi}{3})$ denotes the cost when the angles are the ones of an equilateral triangle. The second inequality follows form the fact that we are keeping the angles fixed, but decreasing the size of the sides. By comparing the expressions and using Lemma 2.2.1 we get the desired inequality. □

Finally we come to the proof of Theorem 2.0.5.

Proof of Theorem 2.0.5. Let ε, M as in Lemma 2.2.1. Since ρ is fully supported and T is continuous, we have

$$T(x) \to s_2^- \text{ and } T^2(x) \to s_2^+ \text{ as } x \to 0,$$

and

$$T(x) \to s_1^+ \text{ and } T^2(x) \to +\infty \text{ as } x \to s_1^-.$$

This allows to choose triplets (r_1,r_2,r_3) and (ℓ_1,ℓ_2,ℓ_3) as in the hypothesis of Lemma 2.2.2 such that

$$(r_1,r_2,r_3) = (x,T(x),T^2(x)) \quad \text{and} \quad (\ell_1,\ell_2,\ell_3) = (y,T(y),T^2(y)).$$

Apply Lemma 2.2.2 to conclude that the support of the DDI map is not c_r-cyclically monotone. □

[2] It can be computed that $\varphi(r_1,r_2)$ it increasing in r_1.

We can exploit further Theorem 2.0.5 to produce a class of continuous counterexamples to the optimality of the DDI map.

First note that it is easy to construct ρ positive everywhere such that the condition

$$T(x)(T^2(x)-x)^3 - x(T^2(x)+T(x))^3 - T^2(x)(x+T(x))^3 \geq 0$$

holds for every $x \in (0, s_1)$, where T is the DDI map.

Given $s_2 > 0$, one can fix an arbitrary positive function ρ on the interval $[0, s_2]$ such that $\int_0^{s_2} \rho(x)dx = \frac{2}{3}$. Let s_1 such that $\int_0^{s_1} \rho(x)dx = \frac{1}{3}$. Observe that this defines the DDI map $T\colon [0, s_1] \to [s_1, s_2]$ given by

$$\int_0^x \rho(t)dt = \int_{T(x)}^{s_2} \rho(t)dt,$$

or equivalently by the Monge-Ampère equation

$$\begin{cases} T'(x) = -\dfrac{\rho(x)}{\rho(T(x))} \\ T(0) = s_2. \end{cases}$$

We can now transport ρ to the interval $[s_2, +\infty)$ via the map $\varphi(x, T(x))$, where φ was introduced in (2.9), *i.e.*, define ρ on the interval $[s_2, +\infty)$ by

$$\rho(\varphi(x, T(x)) = \frac{\rho(x)}{\frac{d}{dx}\varphi(x, T(x))}.$$

This is possible because $\varphi(0, s_2) = s_2$, $\lim\limits_{x \to s_1^-} \varphi(x, T(x)) = +\infty$ and $\varphi(x, T(x))$ is continuous.

Remark 5. In this construction, the transport function $\varphi(x, T(x))$ can be substituted with $\varphi(x, T(x)) + h(x)$ if h is a C^∞ function, $h(0) = 0$ and $h \geq 0$.

Suppose we construct the density ρ as above, using the transport function $\varphi(x, T(x)) + h(x)$. When is the obtained density continuous?

Since the density is arbitrary in the interval $[0, s_2]$, we can choose it to be continuous in that interval, and since the transport function is C^∞, the density will be continuous on $[s_2, +\infty]$. The problem is at the joining point s_2. On the one hand, from the Monge-Ampère equation we get

$$T'(0) = -\frac{\rho(0)}{\rho(s_2^-)}$$

and on the other hand we get

$$\left[\frac{d}{dx}\varphi(x, T(x)) + h'(x)\right]_{x=0} = \frac{\rho(0)}{\rho(s_2^+)},$$

thus the continuity is equivalent to

$$\left[\frac{d}{dx}\varphi(x,T(x))+h'(x)\right]_{x=0}=-T'(0).$$

By recalling that

$$\varphi(x,T(x))=\frac{5xT(x)+T(x)^2+(x+T(x))\sqrt{T(x)^2+12xT(x)-4x^2}}{2(T(x)-x)},$$

we get that

$$\left[\frac{d}{dx}\varphi(x,T(x))+h'(x)\right]_{x=0}=7+T'(0)+h'(0),$$

hence the equation

$$7+T'(0)+h'(0)=-T'(0)\implies -T'(0)=\frac{\rho(0)}{\rho(s_2^-)}=\frac{7}{2}+\frac{h'(0)}{2}$$

The condition $h\geq 0$ forces $h'(0)\geq 0$, and we deduce that, if $\rho(0)\geq \frac{7}{2}\rho(s_2^-)$, then there exists a continuous density that satisfies the hypotheses of Theorem 2.0.5, obtained for instance by choosing $h(x)=\lambda x$ for a suitable parameter λ. This density provides a continuous counterexample to the Seidl conjecture, which was so far missing in the literature.

One could, in principle, exploit the same idea in order to get differentiable counterexamples, and in general C^k or even C^∞ counterexamples, but the explicit computations for $\frac{d^k}{dx^k}\varphi(x,T(x))$ get very hard to treat. However, if ρ satisfies the strict inequality

$$\rho(0)>\frac{7}{2}\rho(s_2^-),$$

then a C^∞ counterexample can be implicitly constructed as follows. Let for simplicity $\psi(x):=\varphi(x)+h(x)$, and recall that we are searching for a C^∞ function h such that $h(0)=0$, $h\geq 0$. By differentiating n times the Monge-Ampère equation

$$\psi'(x)\rho(\psi(x))=\rho(x)\quad x\in[0,s_1]$$

we get

$$\sum_{k=0}^{n}\binom{n}{k}\psi^{(n-k+1)}(x)\frac{d^k}{dx^k}\rho(\psi(x))=\rho^{(n)}(x),$$

and (computing for $x=0$)

$$\sum_{k=0}^{n}\binom{n}{k}\psi^{(n-k+1)}(0)\left[\frac{d^k}{dx^k}\rho(\psi(x))\right]_{x=0}=\rho^{(n)}(0).$$

Suppose that we already defined $h'(0), \ldots, h^{(n)}(0)$. The only term containing $h^{(n+1)}(0)$ is obtained for $k = 0$ in the LHS, and reads $\psi^{(n+1)}(0)\rho(s_2)$. Hence we can isolate it and get

$$\psi^{(n+1)}(0)\rho(s_2) = \rho^{(n)}(0) - \sum_{k=1}^{n} \binom{n}{k} \psi^{(n-k+1)}(0) \left[\frac{d^k}{dx^k} \rho(\psi(x)) \right]_{x=0}.$$

Since ρ is positive everywhere, in particular $\rho(s_2) > 0$ and get a well-defined expression for $h^{(n+1)}(0)$ depending on $h'(0), \ldots, h^{(n)}(0)$ (alredy previously defined by induction). The base step is given by $h(0) = 0$.

By Borel's lemma there exists a smooth function $f : \mathbb{R} \to \mathbb{R}$ such that $f^{(k)} = h^{(k)}$ for all natural numbers k. Since the inequality $\rho(0) > \frac{7}{2}\rho(s_2^-)$ implies $h'(0) > 0$, there is a $\delta > 0$ and an interval $[0, \delta]$ such that $f(x) > 0$ for all $x \in [0, \delta]$. We now choose our h to coincide with f in the interval $[0, \frac{\delta}{2})$ and to be constant, equal to $f(\delta)$ on the interval $[\delta, \infty)$. On the interval $(\frac{\delta}{2}, \delta)$ we join these two parts smoothly, so that the function h is smooth on all of its domain $[0, \infty)$ — at 0 we mean by smoothness the existence of all derivatives from the right. This in turn defines a smooth density ρ on the interval $[s_2, +\infty)$ by transporting $\rho\lfloor_{[0,s_1]}$ with $\psi = \varphi + h$.

Chapter 3

Many-body limit of the multi-marginal OT functional

We consider a standard multi-marginal Optimal Transport (OT) problem defined by

$$C_N(\rho) = \inf \left\{ \int c_N(x_1, \dots, x_N) dP(x_1, \dots, x_N) \mid P \in \Pi_N(\rho) \right\} \tag{3.1}$$

where $\rho \in \mathcal{P}(\mathbb{R}^d)$.

As a cost function c_N we treat a two-particle interaction of the form

$$c_N(x_1, \dots, x_N) = \frac{2}{N(N-1)} \sum_{1 \le i < j \le N} \ell(|x_i - x_j|) \tag{3.2}$$

where $\ell \colon [0, +\infty] \to \mathbb{R}$ has the following properties:

(i) $\ell(r) \ge 0$;

(ii) ℓ is lower semi-continuous;

(iii) $\lim_{r \to +\infty} \ell(r) = 0$;

(iv) ℓ is locally integrable on $\mathbb{R}^d$, meaning $\int_{B(0,R)} \ell(|z|)dz < +\infty$ for every $R > 0$.

It is common in many applications (Density Functional Theory, crowd motion, statistics) to encounter minimum problems of the form

$$\inf_{\rho \in \mathcal{P}(\mathbb{R}^d)} \{ C_N(\rho) + \mathcal{F}(\rho) \},$$

where $\mathcal{F}$ is a suitable density functional. In this context it is important to understand the behaviour of this value and the structure of the minimizers for a large number N of particles/people, as this can be used to approximate the behaviour of large systems, often impossible to compute numerically in an exact way.

The first step in order to treat rigorously these instances is to understand the limit as $N \to \infty$ of the multi-marginal OT functional. In this setting, a natural tool is the notion of Γ-convergence with respect to the weak* topology of Radon measures on $\mathbb{R}^d$. In particular if $C_\infty := \Gamma\text{-}\lim_{n\to\infty} C_N$ exists and can be identified, it will possible to pass to the limit in minimum problems of the kind $\inf_\rho \{C_N(\rho) + \mathcal{F}(\rho)\}$. In particular, if $\mathcal{F}$ is weakly continuous, by applying a celebrated theorem of De Giorgi, we will obtain the convergence of the infima

$$\lim_{n\to\infty} \inf_\rho \{C_N(\rho) + \mathcal{F}(\rho)\} = \min_\rho \{C_\infty(\rho) + \mathcal{F}(\rho)\}$$

and the weak* convergence of minimizing sequences $\mathcal{P}(\mathbb{R}^d)$ to elements of $\arg\min \{C_\infty + \mathcal{F}\}$. Let us mention that in such a setting the mimimizers of the limit problem can be merely sub-probabilities due a possible loss of mass at infinity.

Having in mind the result of C. Cotar and M. Petrache [21], an ideal prospect for our work would be to get the next-order term of [21] as Γ-limit of a suitable renormalized sequence.

Structure of the chapter The material for this chapter comes from a work in collaboration with G. Bouchitté (Université de Toulon). I am grateful to the financial support of INdAM (Istituto Nazionale d'Alta Matematica), via the LIA LYSM project.

In Section 3.1 we introduce some notation and present some known results of convex analysis and measure theory.

In Section 3.2, we extend to general costs ℓ the relaxation and duality framework recently developed in the case of the Coulomb interaction energy (see [6]). We also prove the Γ-convergence of (C_N) and provide a characterization of the limit functional C_∞. We remark that the pointwise convergence of the multi-marginal OT functional was studied, in the case of a positive definite cost function, by B. Pass et al. in [36]. We give also a description of the main properties of C_∞ and of its Fenchel conjugate as a functional on $C_b(\mathbb{R}^d)$.

In Section 3.3 we come back to the minimum problem

$$\inf_\rho \{C_\infty(\rho) + \mathcal{F}(\rho)\}$$

in the special case of a linear and continuous functional $\mathcal{F}$. We derive optimality conditions and study the possibility of finding solutions of mass strictly less than 1 — which reveals a loss of mass at infinity.

Finally, in the appendix we will present some explicit computations of the relaxed functional $\overline{C_N}$ for a single Dirac delta (Theorem 3.A.2) and for a convex combination of two Dirac deltas (Theorem 3.A.3). As we shall see, even in these apparently simple cases the result is not trivial. Since in the literature very few results are known about the relaxed multi-marginal OT functional, we believe that every small step can be of interest.

3.1 Preliminary results

Tools and notation from convex analysis

Let X be a topological vector space, and $f\colon X \to [-\infty, +\infty]$. The *convex hull* of f, denoted by $\operatorname{cl} f$, is the largest lower semi-continuous convex function below f. It may be defined as the function whose epigraph is the closed convex hull of the epigraph of f in $X \times \overline{\mathbb{R}}$. Notice that, if f is lower semi-continuous and convex, $\operatorname{cl} f = f$.

The *Legendre-Fenchel conjugate* of f, denoted by f^*, is defined on X^* as

$$f^*(v) = \sup \{\langle v, x\rangle - f(x) \mid x \in X\}.$$

The following properties are well-known in the literature. We refer to [43, 7] for exhaustive treatments.

- f^* is convex and lower semi-continuous;
- if $f \le g$, then $f^* \ge g^*$;
- $f^{**} = \operatorname{cl} f$.

The *lower semi-continuous envelope* of f, denoted $\overline{f}$ is the greatest lower semi-continuous function below f, *i.e.*,

$$\overline{f}(x) = \sup \{g \le f \mid g \text{ is l.s.c.}\}.$$

Lemma 3.1.1 ([7, Proposition 1.31])**.** *Let $f\colon X \to [-\infty, \infty]$. Then*

$$\overline{f} = \inf_{x_n \to x} \liminf_{n\to\infty} f(x_n).$$

Lemma 3.1.2. *Let $f\colon X \to [-\infty, \infty]$. Then $(\overline{f})^{**} = f^{**}$.*

Proof. Since $\overline{f} \le f$, then $(\overline{f})^{**} \le f^{**}$. Given $x \in X$ and $\epsilon > 0$, let (x_n) be such that $\overline{f}(x) \ge \lim_{n\to\infty} f(x_n) - \epsilon$. Since every element $v \in X^*$ is a continuous linear functional,

$$\langle x, v\rangle - \overline{f}(x) \le \lim_{n\to\infty} \langle x_n, v\rangle - \overline{f}(x_n) + \epsilon \le f^*(v) + \epsilon.$$

By passing to the supremum over $x \in X$, since ϵ was arbitrary, we get

$$(\overline{f})^*(v) \le f^*(v),$$

which in turn implies $(\overline{f})^{**} \ge f^{**}$. □

Lemma 3.1.3. *Let $f\colon X \to [-\infty, \infty]$ be convex. Then $\overline{f} = f^{**}$.*

Proof. Since $f^{**} \le f$ and is l.s.c., then $f^{**} \le \overline{f}$. On the other hand, $\overline{f}$ is convex: given $x, y \in X$, let $f(x_n) \to \overline{f}(x)$ and $f(y_n) \to \overline{f}(y)$; if $t \in [0,1]$ we have

$$\begin{aligned}\overline{f}(tx + (1-t)y) &\le \liminf_{n\to\infty} f(tx_n + (1-t)y_n) \le \liminf_{n\to\infty} [tf(x_n) + (1-t)f(y_n)] \\ &= t\overline{f}(x) + (1-t)\overline{f}(y).\end{aligned}$$

Since $\overline{f}$ is convex l.s.c. below f, $\overline{f} \le f^{**}$. □

Proposition 3.1.4 ([7, Proposition 1.32]). *Let $f_n\colon X \to [-\infty, \infty]$. Then*

$$\Gamma\text{-}\lim_{n\to\infty} f_n = \Gamma\text{-}\lim_{n\to\infty} \overline{f_n}.$$

Tools and notation from measure theory

We will denote by $\langle \cdot, \cdot \rangle$ the duality between the continuous functions vanishing at infinity and the finite Borel measures on the Euclidean space:

$$\langle v, \mu \rangle = \int v d\mu.$$

We will denote by $\mathcal{M}$ the vector space of finite Borel measures, $\mathcal{P}$ the set of non-negative probability measures, by $\mathcal{P}_-$ the set of non-negative sub-probability measures.

If $v\colon \mathbb{R}^d \to \mathbb{R}$, we define $S_N v\colon (\mathbb{R}^d)^N \to \mathbb{R}$ as

$$S_N v(x_1, \dots, x_N) = \frac{1}{N} \sum_{j=1}^{N} v(x_j)$$

For a measure $\mu \in \mathcal{M}((\mathbb{R}^d)^N)$, we denote by $\mathrm{Sym}(\mu)$ its symmetrization, given by

$$\mathrm{Sym}(\mu)(E) = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \mu(\sigma(E)),$$

where $\sigma(E) = \big\{(x_1, \dots, x_N) \,|\, (x_{\sigma(1)}, \dots, x_{\sigma(N)}) \in E\big\}$ for a permutation $\sigma \in \mathfrak{S}_N$.

Given $\mu \in \mathcal{M}(\mathbb{R}^d)$ and $h \in \mathbb{R}^d$, let $\tau_h\mu$ be the translation of μ by the vector h, *i.e.*, $\tau_h\mu(E) = \mu(E - h)$ for every Borel set E.

Lemma 3.1.5. *Let $\mu \in \mathcal{M}(\mathbb{R}^d)$. Then $\tau_h\mu \rightharpoonup 0$ as $|h| \to \infty$.*

Proof. By splitting $\mu = \mu_+ - \mu_-$, we may assume μ to be a non-negative measure. Given $\epsilon > 0$ and $f \in C_0(\mathbb{R}^d)$, let $R > 0$ such that $|f(y)| < \epsilon$ as $|y| \geq R$ and $\mu(B(0,R)^c) \leq \epsilon$. Then, if $|h| \geq 2R$, we have

$$\begin{aligned}\left|\int f(x) d\tau_h \mu(x)\right| &= \left|\int f(x+h) d\mu(x)\right| \\ &\leq \int_{B(0,R)} |f(x+h)|\, d\mu(x) + \int_{B(0,R)^c} |f(x+h)|\, d\mu(x) \\ &\leq \epsilon\mu(B(0,R)) + \epsilon \sup |f| \leq \epsilon(\mu(\mathbb{R}^d) + \sup|f|). \qquad \square\end{aligned}$$

Moreover, we will make use of the following well-known result (see for instance [1, Section 5.1]).

Theorem 3.1.6. *Let $\mu_n, \mu \in \mathcal{M}(\mathbb{R}^d)$. Then the following are equivalent*

(i) $\mu_n \rightharpoonup \mu$;

(ii) $\liminf_{n\to\infty} \langle f, \mu_n\rangle \geq \langle f, \mu\rangle$ *for every f l.s.c. bounded from below;*

(iii) $\limsup_{n\to\infty} \langle f, \mu_n\rangle \leq \langle f, \mu\rangle$ *for every f u.s.c. bounded from above.*

3.2 Duality and Γ-convergence

Notice that the functional $C_N \colon \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is convex. Given $\rho_1, \rho_2 \in \mathcal{P}(\mathbb{R}^d)$, let $P_1 \in \Pi_N(\rho_1)$ and $P_2 \in \Pi_N(\rho_2)$ optimal in (3.1). Then $tP_1 + (1-t)P_2 \in \Pi_N(t\rho_1 + (1-t)\rho_2)$, and thus

$$\begin{aligned}C_N(t\rho_1 + (1-t)\rho_2) &\leq \langle c_N, tP_1 + (1-t)P_2\rangle \\ &= t\langle c_N, P_1\rangle + (1-t)\langle c_N, P_2\rangle \\ &= tC_N(\rho_1) + (1-t)C_N(\rho_2).\end{aligned}$$

In order to study the Γ-limit of C_N as $N \to \infty$, we consider the dual formulation given by

$$M_N(v) := C_N^*(v) = \sup_{\rho\in\mathcal{P}(\mathbb{R}^d)} \{\langle v, \rho\rangle - C_N(\rho)\}. \tag{3.3}$$

for $v \in C_0(\mathbb{R}^d)$ and

$$C_N^{**}(\rho) = M_N^*(\rho) = \sup_{v\in C_0(\mathbb{R}^d)} \{\langle v, \rho\rangle - M_N(v)\}.$$

By Lemma 3.1.3 and convexity of C_N, $\overline{C_N} = C_N^{**}$. Notice moreover that, if $\rho \notin \mathcal{P}_-(\mathbb{R}^d)$, then $\overline{C_N}(\rho) = +\infty$. This is expected, as $\mathcal{P}_-(\mathbb{R}^d)$ is the convex closure of $\mathcal{P}(\mathbb{R}^d)$.

An explicit formula for $\overline{C_N}(\rho)$ was given in [6, Theorem 2.3], and reads as follows:

$$\overline{C_N}(\rho) = \inf_{\substack{a_1,\dots,a_N\ge 0\\ \rho_1,\dots,\rho_N\in\mathcal{P}(\mathbb{R}^d)}} \left\{ \sum_{k=2}^N a_k \frac{k(k-1)}{N(N-1)} C_k(\rho_k) \mid \sum_{k=1}^N a_k \le 1, \sum_{k=1}^N \frac{k}{N} a_k \rho_k = \rho \right\}. \tag{3.4}$$

Observe in particular that, if $|\rho| \le \frac{1}{N}$, then one can choose $a_1 = 1$, $a_2 = \dots = a_N = 0$ to get $\overline{C_N}(\rho) = 0$. Notice also that, if $\rho \in \mathcal{P}(\mathbb{R}^d)$, the only choice in (3.4) is $a_1 = \dots = a_{N-1} = 0$, $a_N = 1$, which yields $\overline{C_N}(\rho) = C_N(\rho)$.

The interested reader may look to [6, 21] for more insights on $\overline{C_N}$ and the link with the grand-canonical formulation of Optimal Transport.

Lemma 3.2.1. *For every $v \in C_0(\mathbb{R}^d)$ one has*

$$M_N(v) = \sup\left\{ S_N v(x_1,\dots,x_N) - c_N(x_1,\dots,x_N) \mid x_1,\dots,x_N \in \mathbb{R}^d \right\}. \tag{3.5}$$

Proof. Given $x_1,\dots,x_N \in \mathbb{R}^d$, let

$$\rho = \frac{1}{N}\sum_{j=1}^N \delta_{x_j} \quad \text{and} \quad P = \frac{1}{N!}\sum_{\sigma\in\mathfrak{S}_N} \delta_{x_{\sigma(1)}} \otimes \dots \otimes \delta_{x_{\sigma(N)}}.$$

Observe that $\rho \in \mathcal{P}(\mathbb{R}^d)$, and $P \in \Pi_N(\rho)$. Hence

$$\begin{aligned} S_N v(x_1,\dots,x_N) - c_N(x_1,\dots,x_N) &= \int v d\rho - \int c_N dP \\ &\le \int v d\rho - C_N(\rho) \le M_N(v), \end{aligned}$$

which gives an inequality.

On the other hand, for any $\rho \in \mathcal{P}(\mathbb{R}^d)$, if $P \in \Pi_N(\rho)$ is optimal in (3.1) one has

$$\int v d\rho - C_N(\rho) = \int (S_N v - c_N) dP \le \sup(S_N v - c_N)$$

Passing to the supremum on the left-hand side one gets the converse inequality, and thus the thesis. $\square$

Lemma 3.2.2. *The sequence $(M_N)_{N\ge 2}$ is monotone decreasing and equi-Lipschitz (with Lipschitz constant equal to 1).*

Proof. Let us start by observing that

$$c_{N+1}(x_1,\dots,x_{N+1}) = \frac{1}{N+1}\sum_{k=1}^{N+1} c_N(x_1,\dots,\hat{x}_k,\dots,x_{N+1})$$

and

$$S_{N+1}v(x_1,\dots,x_{N+1}) = \frac{1}{N+1}\sum_{k=1}^{N+1} S_N v(x_1,\dots,\hat{x}_k,\dots,x_{N+1}).$$

Hence, for every $x_1,\dots,x_{N+1}\in\mathbb{R}^d$,

$$\begin{aligned}
&S_{N+1}v(x_1,\dots,x_{N+1}) - c_{N+1}(x_1,\dots,x_{N+1})\\
&= \frac{1}{N+1}\sum_{k=1}^{N+1}\left(S_N v(x_1,\dots,\hat{x}_k,\dots,x_{N+1}) - c_N(x_1,\dots,\hat{x}_k,\dots,x_{N+1})\right)\\
&\le \frac{1}{N+1}\sum_{k=1}^{N+1} M_N(v) = M_N(v)
\end{aligned}$$

Passing to the supremum on the left-hand side we get $M_{N+1}\le M_N$, as wanted.

In order to prove the second part of the statement, let $v_1, v_2 \in C_0(\mathbb{R}^d)$. Take $x_1,\dots,x_N\in\mathbb{R}^d$ optimal up to a threshold ϵ for $M_N(v_1)$ in (3.5). Then

$$M_N(v_1) - M_N(v_2) \le \frac{1}{N}\sum_{j=1}^{N}(v_1(x_j) - v_2(x_j)) + \epsilon \le \|v_1 - v_2\|_\infty + \epsilon.$$

By letting $\epsilon\to 0$, and then switching the roles of v_1 and v_2 we get the thesis. $\square$

Corollary 3.2.3. *The sequence $(\overline{C_N})_{N\ge 1}$ is monotone increasing.*

Proof. By Lemma 3.2.2 we have

$$\begin{aligned}
\overline{C_N}(\rho) &= \sup_{v\in C_0(\mathbb{R}^d)}\left\{\int v d\rho - M_N(v)\right\}\\
&\ge \sup_{v\in C_0(\mathbb{R}^d)}\left\{\int v d\rho - M_{N+1}(v)\right\} = \overline{C_{N+1}}(\rho).
\end{aligned}$$ $\square$

We get from Lemma 3.2.2 and Corollary 3.2.3 the existence of the pointwise limits

$$M_\infty(v) = \lim_{N\to\infty} M_N(v) \quad\text{and}\quad C_\infty(\rho) = \lim_{N\to\infty}\overline{C_N}(\rho).$$

Remark 6. The uniform Lipschitz continuity of $(M_N)_{N\ge 2}$ implies the same property for M_∞: given $v_1, v_2\in C_0(\mathbb{R}^d)$, for every $\epsilon>0$ let N such that $M_N(v_j) - M_\infty(v_j)\le\epsilon$ for $j=1,2$. Then

$$|M_\infty(v_1) - M_\infty(v_2)| \le |M_N(v_1) - M_N(v_2)| + 2\epsilon \le \|v_1 - v_2\| + 2\epsilon.$$

Letting $\epsilon\to 0$ we get that M_∞ is a Lipschitz continuous functional (with Lipschitz constant equal to 1).

Remark 7. It is easy to prove that $C_\infty = M_\infty^*$: due to Corollary 3.2.3 we have

$$\begin{aligned}
C_\infty(\rho) &= \sup_{N\geq 2} \overline{C_N}(\rho) \\
&= \sup_{N\geq 2} \sup_{v\in C_0(\mathbb{R}^d)} \langle v, \rho\rangle - M_N(v) \\
&= \sup_{v\in C_0(\mathbb{R}^d)} \sup_{N\geq 2} \langle v, \rho\rangle - M_N(v) \\
&= \sup_{v\in C_0(\mathbb{R}^d)} \langle v, \rho\rangle - M_\infty(v) = M_\infty^*(\rho).
\end{aligned}$$

Besides the pointwise convergence, in the following results we show that both the sequences (M_N) and $(\overline{C_N})$ Γ-converge to their pointwise limits.

Theorem 3.2.4. *The functionals M_N Γ-converge as $N \to \infty$ to the functional M_∞.*

Proof. When we have pointwise convergence of the functionals, the lim sup inequality is trivial: take $v_N = v$ as a recovery sequence to get

$$\limsup_{N\to\infty} M_N(v_N) = \limsup_{N\to\infty} M_N(v) = M_\infty(v).$$

Let $(v_N)_{N\geq 2}$ be any sequence in $C_0(\mathbb{R}^d)$ uniformly converging to v. Then by Lemma 3.2.2 we have $M_N(v_N) \geq M_\infty(v_N)$. By taking the lim inf on both sides and using the continuity of M_∞ (see Remark 6) we get

$$\liminf_{N\to\infty} M_N(v_N) \geq M_\infty(v). \qquad \square$$

Theorem 3.2.5. *The functionals $\overline{C_N}$ Γ-converge as $N \to \infty$ to the functional $C_\infty = M_\infty^*$.*

Proof. As before the pointwise convergence makes the lim sup inequality trivially satisfied.

Take $\rho_N \in \mathcal{P}_-(\mathbb{R}^d)$, $\rho_N \rightharpoonup \rho$. If $M_\infty^*(\rho) = +\infty$, for any $n \in \mathbb{N}$ let $v_n \in C_0(\mathbb{R}^d)$ such that $\langle v_n, \rho\rangle - M_\infty(v_n) \geq n$. Then

$$\overline{C_N}(\rho_N) = \sup_{v\in C_0(\mathbb{R}^d)} \langle v, \rho_N\rangle - M_N(v) \geq \langle v_n, \rho_N\rangle - M_N(v_n),$$

whence

$$\liminf_{N\to\infty} \overline{C_N}(\rho_N) \geq \liminf_{N\to\infty} \langle v_n, \rho_N\rangle - M_N(v_n) = \langle v_n, \rho\rangle - M_\infty(v_n) \geq n.$$

If on the contrary $M_\infty^*(\rho) < +\infty$, given $\epsilon > 0$, let $v_\epsilon \in C_0(\mathbb{R}^d)$ such that $M_\infty^*(\rho) - \epsilon \leq \langle v_\epsilon, \rho\rangle - M_\infty(v_\epsilon)$. Then

$$\overline{C_N}(\rho_N) = \sup_{v\in C_0(\mathbb{R}^d)} \langle v, \rho_N\rangle - M_N(v) \geq \langle v_\epsilon, \rho_N\rangle - M_N(v_\epsilon),$$

whence

$$\liminf_{N\to\infty} \overline{C_N}(\rho_N) \geq \liminf_{N\to\infty} \langle v_\epsilon, \rho_N\rangle - M_N(v_\epsilon) = \langle v_\epsilon, \rho\rangle - M_\infty(v_\epsilon) \geq M^*_\infty(\rho) - \epsilon$$

In both cases we get the lim inf inequality and thus the thesis. □

In the introduction we stated the minimum problem of interest using the functional C_N, but so far we treated the Γ-convergence of $\overline{C_N}$. This is justified by Proposition 3.1.4.

Some consequences of our first results, which we state and prove here, will be useful in the following.

Corollary 3.2.6. *For every $2 \leq k \leq N$ one has*

$$M_N(v) \geq \frac{k(k-1)}{N(N-1)} M_k\left(\frac{N-1}{k-1}v\right).$$

Proof. Let $x_1, \dots, x_k$ be optimal for $M_k\left(\frac{N-1}{k-1}v\right)$ in (3.5). Then we can send $x_{k+1}, \dots, x_N$ to ∞ to get

$$\begin{aligned} M_N(v) &\geq \frac{1}{N}\sum_{j=1}^{k} v(x_j) - \frac{2}{N(N-1)} \sum_{0\leq i<j\leq k} \ell(|x_i - x_j|) \\ &= \frac{k(k-1)}{N(N-1)}\left(\frac{N-1}{k(k-1)}\sum_{j=1}^{k} v(x_j) - \frac{2}{k(k-1)}\sum_{0\leq i<j\leq k} \ell(|x_i - x_j|)\right) \\ &= \frac{k(k-1)}{N(N-1)}\left(\frac{1}{k}\sum_{j=1}^{k}\frac{N-1}{k-1}v(x_j) - c_k(x_1, \dots, x_k)\right) \\ &= \frac{k(k-1)}{N(N-1)} M_k\left(\frac{N-1}{k-1}v\right). \end{aligned}$$

□

Corollary 3.2.7. *For every $v \in C_0(\mathbb{R}^d)$ and every $N \geq 2$*

$$\frac{\sup v}{N} \leq M_N(v) \leq \sup v.$$

Proof. First we prove the upper bound: since $c_N \geq 0$ we have

$$M_N(v) \leq \sup\left\{S_N v(x_1, \dots, x_N) \mid x_1, \dots, x_N \in \mathbb{R}^d\right\} = \sup v.$$

Given $\epsilon > 0$, fix x_1 be such that $v(x_1) \geq \sup v - \epsilon$. Next send all the other points $x_2, \dots, x_N$ to ∞ in such a way that all the distances $|x_i - x_j|$ with $1 \leq i < j \leq N$ go to $+\infty$. Since ℓ vanishes at infinity, we have that

$$M_N(v) \geq \frac{1}{N}(\sup v - \epsilon).$$

By letting $\epsilon \to 0$ we get the thesis. □

Study of the limit functionals

We want to refine the study of the limit functionals M_∞ and C_∞. First of all we prove that the domain of C_∞ is a subset of $\mathcal{P}_-(\mathbb{R}^d)$.

Lemma 3.2.8. *If $\rho \notin \mathcal{P}_-(\mathbb{R}^d)$ then $C_\infty(\rho) = +\infty$.*

Proof. If ρ is not a positive measure, let $v \in C_0(\mathbb{R}^d)$, $v \geq 0$ be such that $\langle v, \rho \rangle = -\lambda < 0$. For every $n \in \mathbb{N}$, let $v_n = -nv$, and notice that $v_n \in C_0(\mathbb{R}^d)$, $\langle v_n, \rho \rangle = n\lambda$. Hence

$$C_\infty(\rho) \geq \sup_{n\in\mathbb{N}} \{\langle v_n, \rho\rangle - M_\infty(v_n)\} = \sup_{n\in\mathbb{N}} n\lambda = +\infty.$$

Given $\rho \notin \mathcal{P}_-(\mathbb{R}^d)$, let $r > 0$ be such that $\rho(B(0, r)) > 1$. For every $n \in \mathbb{N}$ choose $v_n \in C_0(\mathbb{R}^d)$ radially decreasing such that $v_n \geq 0$ and $v_n(x) = n$ if $x \in B(0, r)$. Then we have $\sup v_n = n$, and recalling Corollary 3.2.7

$$C_\infty(\rho) \geq \sup_{n\in\mathbb{N}} \{\langle v_n, \rho\rangle - M_\infty(v_n)\} \geq \sup_{n\in\mathbb{N}} \{n\rho(B(0, r)) - \sup v_n\} = +\infty. \qquad \square$$

Moreover, the domain of C_∞ is a dense subset of $\mathcal{P}_-(\mathbb{R}^d)$. In order to prove it, we prepare a technical result.

Lemma 3.2.9. *For every ball $B(0, r)$, the uniform measure ρ_r concentrated on the ball is such that $C_\infty(\rho_r) < +\infty$.*

Proof. In order to bound $C_N(\rho_r)$, we consider the transport plan $P_{N,r} = \underbrace{\rho_r \otimes \cdots \otimes \rho_r}_{N \text{ times}}$, and observe that

$$\begin{aligned}
C_N(\rho_r) &\leq \int c(x_1, \dots, x_N)dP_{N,r} = \int c(x_1, x_2)d(\rho_r \otimes \rho_r)\\
&= \int \mathbb{1}_{B(0,r)}(x_1)\mathbb{1}_{B(0,r)}(x_2)\ell(|x_1 - x_2|)dx_1dx_2\\
&= \frac{1}{2}\int \mathbb{1}_{B(0,2r)}(w + z)\mathbb{1}_{B(0,2r)}(w - z)\ell(|z|)dzdw\\
&\leq \frac{1}{2}\int \mathbb{1}_{B(0,2r)}(w)\mathbb{1}_{B(0,2r)}(z)\ell(|z|)dzdw\\
&= 2^{d-1}r^d\omega_d \int_{B(0,2r)} \ell(|z|)dz = K(r, d),
\end{aligned}$$

where $K(r, d)$ is a finite constant, in view of the property (iv) of ℓ. Hence

$$C_\infty(\rho_r) = \lim_{N\to+\infty} C_N(\rho_r) \leq K(r, d) < +\infty. \qquad \square$$

Corollary 3.2.10. *The domain of C_∞ is a dense subset of $\mathcal{P}_-(\mathbb{R}^d)$.*

Proof. It suffices to show that every atomic probability can be approximated by elements of the domain. This amounts to prove that δ_x can be approximated for every $x \in \mathbb{R}^d$. We consider the probability ρ_n given by the normalized uniform measure restricted to the ball $B(x, 1/n)$ of radius $1/n$, *i.e.*,

$$\frac{d\rho_n}{dx} = \frac{n^d}{\omega_d} \mathbb{1}_{B(0,1/n)},$$

where ω_d is the measure of the unitary ball in $\mathbb{R}^d$. As it is well-known, $\rho_n \rightharpoonup \delta_x$ as $n \to +\infty$. Moreover, since C_∞ is invariant under translation, $C_\infty(\rho_n) < +\infty$ by Lemma 3.2.9. □

In view of Lemma 3.2.8, we may study the restriction of C_∞ to $\mathcal{P}_-(\mathbb{R}^d)$. We characterize in the next result the trivial case in which $C_\infty \equiv 0$ on $\mathcal{P}_-(\mathbb{R}^d)$.

Proposition 3.2.11. *The following are equivalent:*

(i) $C_\infty(\rho) = 0$ *for every* $\rho \in \mathcal{P}_-(\mathbb{R}^d)$*;*

(ii) $M_\infty(v) = \sup v$ *for every* $v \in C_0(\mathbb{R}^d)$*;*

(iii) $M_N(v) = \sup v$ *for every* $N \in \mathbb{N}$*,* $v \in C_0(\mathbb{R}^d)$*;*

(iv) $\ell(0) = 0$.

Proof. First we prove the equivalence between (i) and (ii). If (i) holds, then we have

$$M_\infty(v) = C_\infty^*(v) \sup_{\rho \in \mathcal{P}_-(\mathbb{R}^d)} \{\langle v, \rho \rangle - C_\infty(\rho)\} = \sup_{\rho \in \mathcal{P}_-(\mathbb{R}^d)} \langle v, \rho \rangle = \sup v.$$

On the other hand, if (ii) holds then

$$C_\infty(\rho) = M_\infty^*(\rho) = \sup_{v \in C_0(\mathbb{R}^d)} \{\langle v, \rho \rangle - \sup v\} \leq 0,$$

but C_∞ is non-negative.

The equivalence between (ii) and (iii) follows easily from Corollary 3.2.7 and Lemma 3.2.2.

Finally, we prove the equivalence between (iii) and (iv). Suppose that (iii) holds and choose $v \in C_0(\mathbb{R}^d)$ such that it has a unique maximum point x_0, *i.e.*, $v(x) < \sup v$ for every $x \neq x_0$. If $x_j \neq x_0$ for some j, then

$$S_N v(x_1, \dots, x_N) - c_N(x_1, \dots, x_N) < \sup v = M_N(v).$$

Hence the supremum is reached only for $x_1 = \dots = x_N = x_0$. Thus

$$M_N(v) = S_N v(x_0, \dots, x_0) - \ell(0) = \sup v - \ell(0),$$

proving that $\ell(0) = 0$.

If (iv) holds, given $v \in C_0(\mathbb{R}^d)$ let $x_\epsilon \in \mathbb{R}^d$ such that $v(x_\epsilon) \geq \sup v - \epsilon$. Choose $x_1 = \cdots = x_N = x_\epsilon$ in (3.5) to get

$$M_N(v) \geq S_N v(x_\epsilon, \ldots, x_\epsilon) - \ell(0) \geq \sup v - \epsilon.$$

This proves that $M_N(v) \geq \sup v$, which combined with Corollary 3.2.7 allows to conclude. □

Recall that M_∞ is continuous (actually 1-Lipschitz) and convex, since it is the pointwise limit of the convex functionals M_N. Hence, by duality,

$$M_\infty(v) = M_\infty^{**}(v) = C_\infty^*(v) = \sup_{\rho \in \mathcal{P}_-(\mathbb{R}^d)} \langle v, \rho \rangle - C_\infty(\rho). \tag{3.6}$$

We give in the following result some alternative formulations.

Proposition 3.2.12. *For every $v \in C_0(\mathbb{R}^d)$ one has*

$$\begin{aligned} M_\infty(v) = \sup_{\rho \in \mathcal{P}_-(\mathbb{R}^d)} \langle S_2 v - c_2, \rho \otimes \rho \rangle &= \sup_{\nu \in \mathcal{P}_-(\mathcal{P}(\mathbb{R}^d))} \int \langle S_2 v - c_2, Q \otimes Q \rangle \, d\nu(Q) \\ = \sup_{\rho \in \mathcal{P}(\mathbb{R}^d)} \langle S_2 v - c_2, \rho \otimes \rho \rangle &= \sup_{\nu \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))} \int \langle S_2 v - c_2, Q \otimes Q \rangle \, d\nu(Q). \end{aligned}$$

Proof. Clearly each term of the first line is bigger than or equal to the corresponding term of the second line.

Due to the pairwise-interaction structure of the cost c_N, for every $\rho \in \mathcal{P}_-(\mathbb{R}^d)$ and every $N \geq 2$ one has

$$\langle S_2 v - c_2, \rho \otimes \rho \rangle = |\rho|^{2-N} \langle S_N v - c_N, \underbrace{\rho \otimes \cdots \otimes \rho}_{N \text{ times}} \rangle \leq |\rho|^2 \sup(S_N v - c_N) \leq M_N(v);$$

by taking the supremum in ρ on the left-hand side and the infimum in N on the right-hand side we get

$$\sup_{\rho \in \mathcal{P}_-(\mathbb{R}^d)} \langle S_2 v - c_2, \rho \otimes \rho \rangle \leq M_\infty(v)$$

For every $\nu \in \mathcal{P}_-(\mathcal{P}(\mathbb{R}^d))$ one has

$$\begin{aligned} \int \langle S_2 v - c_2, Q \otimes Q \rangle \, d\nu(Q) &\leq \sup_{\rho \in \mathcal{P}(\mathbb{R}^d)} \langle S_2 v - c_2, \rho \otimes \rho \rangle \int d\nu(Q) \\ &\leq \sup_{\rho \in \mathcal{P}(\mathbb{R}^d)} \langle S_2 v - c_2, \rho \otimes \rho \rangle, \end{aligned}$$

yielding

$$\sup_{\nu \in \mathcal{P}_-(\mathcal{P}(\mathbb{R}^d))} \int \langle S_2 v - c_2, Q \otimes Q \rangle \, d\nu(Q) \leq \sup_{\rho \in \mathcal{P}(\mathbb{R}^d)} \langle S_2 v - c_2, \rho \otimes \rho \rangle.$$

Finally, for every $N \geq 2$ let $x_1^N, \dots, x_N^N \in \mathbb{R}^d$ be such that

$$M_N(v) \leq (S_N v - c_N)(x_1^N, \dots, x_N^N) + \frac{1}{N},$$

and define

$$\rho_N = \frac{1}{N} \sum_{j=1}^N \delta_{x_j^N} \quad \text{and} \quad P_N = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \delta_{x_{\sigma(1)}^N} \otimes \cdots \otimes \delta_{x_{\sigma(N)}^N}.$$

Observe that $\rho_N \in \mathcal{P}(\mathbb{R}^d)$ and $P_N \in \Pi_N(\rho_N)$. Denote by γ_N the 2-marginal projection of P_N, and notice that

$$(S_N v - c_N)(x_1^N, \dots, x_N^N) = \langle S_2 v - c_2, \gamma_N \rangle .$$

A classical result by Diaconis and Freedman [26, Theorem 13] gives for every N the existence of $\nu_N \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ such that $\left\| \gamma_N - \int (Q \otimes Q) d\nu_N(Q) \right\| \leq 2/N$, where $\|\cdot\|$ denotes the total variation. Hence we have

$$\begin{aligned}
M_N(v) &\leq \langle S_2 v - c_2, \gamma_N \rangle + \frac{1}{N} \\
&= \left\langle S_2 v - c_2, \gamma_N - \int (Q \otimes Q) d\nu_N(Q) \right\rangle \\
&\quad + \int \langle S_2 v - c_2, Q \otimes Q \rangle \, d\nu_N(Q) + \frac{1}{N} \\
&\leq \left\langle S_2 v - c_2, \gamma_N - \int (Q \otimes Q) d\nu_N(Q) \right\rangle \\
&\quad + \sup_{\nu \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))} \int \langle S_2 v - c_2, Q \otimes Q \rangle \, d\nu(Q) + \frac{1}{N}.
\end{aligned}$$

By Theorem 3.1.6, since $S_2 v - c_2$ is upper semi-continuous and bounded from above and $\gamma_N - \int (Q \otimes Q) d\nu_N(Q)$ converges to zero strongly, we have

$$\limsup_{N \to \infty} \left\langle S_2 v - c_2, \gamma_N - \int (Q \otimes Q) d\nu_N(Q) \right\rangle \leq 0.$$

Thus we get

$$M_\infty(v) = \lim_{N \to \infty} M_N(v) \leq \sup_{\nu \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))} \int \langle S_2 v - c_2, Q \otimes Q \rangle \, d\nu(Q),$$

completing the proof. □

Remark 8. The functionals M_N and M_∞ depend only on the positive part of its argument. Given $v \in C_0(\mathbb{R}^d)$, denote by v_+ its positive part, *i.e.*, $v_+ = \max\{v, 0\}$. For every choice of $(x_1, \dots, x_N) \in (\mathbb{R}^d)^N$, if for some k

we have $v(x_k) < 0$, let $x_k \to \infty$ while keeping the other x_j's fixed. This will increase the value of $v(x_k)$ and decrease the value of $c_N(x_1, \dots, x_N)$ by sending to zero the terms $\ell(|x_j - x_k|)$. Hence, the supremum in (3.5) is attained for points $(x_1, \dots, x_N)$ such that $v(x_j) \geq 0$ for every j, thus proving that $M_N(v) = M_N(v_+)$. By passing to the limit, we get the same property for M_∞.

Definition 4. The *direct energy* $D\colon \mathcal{M}(\mathbb{R}^d) \to [0, +\infty]$ is given by

$$D(\rho) = \begin{cases} \displaystyle\int \ell(|x-y|) d\rho(x) d\rho(y) & \text{if } \rho \in \mathcal{P}(\mathbb{R}^d) \\ +\infty & \text{otherwise.} \end{cases} \tag{3.7}$$

The name "energy" is inherited from a physical model where ρ represents a charge density, as this functional equals (up to constants) the potential energy due to the self-interaction of the density ρ.

Lemma 3.2.13. *For every $v \in C_0(\mathbb{R}^d)$ we have $M_\infty(v) = D^*(v)$.*

Proof. Recall that the definition of $D^*(v)$ is

$$D^*(v) = \sup_{\rho \in \mathcal{P}(\mathbb{R}^d)} \langle v, \rho \rangle - D(\rho)$$

and notice that $\langle v, \rho \rangle = \langle S_2 v, \rho \otimes \rho \rangle$ and $\int \ell(|x-y|) d\rho(x) d\rho(y) = \langle c_2, \rho \otimes \rho \rangle$ for every $\rho \in \mathcal{P}(\mathbb{R}^d)$. The conclusion follows from Proposition 3.2.12 $\square$

Corollary 3.2.14. *For every $\rho \in \mathcal{M}(\mathbb{R}^d)$ we have $C_\infty(\rho) = D^{**}(\rho)$.*

Proof. Combine Lemma 3.2.13 and Remark 7. $\square$

In the aforementioned work by B. Pass et al. [36], it was proven that $\lim_{N\to\infty} C_N(\rho) = D(\rho)$ in the case of a positive-definite cost function, *i.e.*, in the case when D is a convex functional. Corollary 3.2.14 is therefore an extension of their result, valid for every pairwise cost function.

Definition 5. For $1 \leq \alpha \leq 2$, let us denote by $D_\alpha\colon \mathcal{M}(\mathbb{R}^d) \to [0, +\infty]$ the α-homogeneous extension of D to sub-probabilities, *i.e.*,

$$D_\alpha(\rho) = \begin{cases} |\rho|^\alpha D\left(\dfrac{\rho}{|\rho|}\right) & \text{if } \rho \in \mathcal{P}_-(\mathbb{R}^d) \\ +\infty & \text{otherwise.} \end{cases}$$

Remark 9. Observe that $D_2(\rho) = \langle c_2, \rho \otimes \rho \rangle$ for every $\rho \in \mathcal{P}_-(\mathbb{R}^d)$.

Theorem 3.2.15. *For every $1 \leq \alpha \leq 2$ we have $C_\infty = D_\alpha^{**}$.*

Proof. We rely on Corollary 3.2.14 and prove that $D^{**} = D_\alpha^{**}$. Clearly we have $D_\alpha \leq D$ for every $\alpha \in [1,2]$, hence $D_\alpha^{**} \leq D^{**}$. Moreover, since $D_2 \leq D_\alpha$, it suffices to prove that $D^{**} \leq D_2^{**}$. We will proceed by proving that $\overline{D} \leq D_2$. In view of Lemma 3.1.2, we will get $D^{**} = (\overline{D})^{**} \leq D_2^{**}$, as wanted.

Recall that

$$\overline{D}(\rho) = \inf_{\rho_n \rightharpoonup \rho} \liminf_{n\to\infty} D(\rho_n).$$

Let $\rho \in \mathcal{P}_-(\mathbb{R}^d)$ with compact support. If $D_2(\rho) = +\infty$ then there is nothing to prove, so assume that $D_2(\rho) < +\infty$. Our goal is to construct a sequence $\rho_n \rightharpoonup \rho$ such that $\liminf D(\rho_n) \leq D_2(\rho)$. This will imply that

$$\overline{D}(\rho) \leq \liminf_{n\to\infty} D(\rho_n) \leq D_2(\rho).$$

Let us introduce the following technical result.

Lemma 3.2.16. *Let $\rho \in \mathcal{P}_-(\mathbb{R}^d)$ with compact support. Then*

$$\lim_{|h|\to\infty} \langle c_2, \rho \otimes \tau_h \rho \rangle = 0.$$

Proof. Given $\epsilon > 0$, let R be such that $\rho(B(0,R)^c) = 0$ and $\ell(r) < \epsilon$ for $r \geq R$. Then, if $|h| \geq 3R$ we have

$$\begin{aligned}\iint \ell(|x-y|) d\rho(x) d\tau_h\rho(y) &= \iint \ell(|x-y-h|) d\rho(x) d\rho(y) \\ &= \iint_{B(0,R)\times B(0,R)} \ell(|x-y-h|) d\rho(x) d\rho(y) \leq \epsilon. \quad \square\end{aligned}$$

Fix a unitary direction $u \in \mathbb{R}^d$. In view of Lemma 3.2.16, for every $n \geq 1$ let R_n be such that

$$\langle c_2, \rho \otimes \tau_{ru}\rho \rangle \leq \frac{1}{n^2}$$

as $|r| \geq R_n$. Let $h_{n,j} = R_n j u$ for $j = 0, \ldots, n$, so that $h_{n,i} - h_{n,j} = ru$ with $|r| \geq R_n$ for every $i \neq j$. We define

$$\rho_n = \rho + \frac{1 - |\rho|}{n} \sum_{j=1}^{n} \tau_{h_{n,j}} \rho.$$

Notice that $\rho_n \in \mathcal{P}(\mathbb{R}^d)$, and $\rho_n \rightharpoonup \rho$ by Lemma 3.1.5. By translation invariance of D_2 and the choice of the $h_{n,j}$'s we get

$$\begin{aligned}
D(\rho_n) &= D_2(\rho) + \frac{2(1-|\rho|)}{n} \sum_{\substack{i,j=0\\ i\neq j}}^{n} \left\langle c_2, \tau_{h_{n,i}}\rho \otimes \tau_{h_{n,j}}\rho \right\rangle \\
&\quad + \frac{(1-|\rho|)^2}{n^2} \sum_{j=1}^{n} \left\langle c_2, \tau_{h_{n,j}}\rho \otimes \tau_{h_{n,j}}\rho \right\rangle \\
&= D_2(\rho) + \frac{2(1-|\rho|)}{n} \sum_{\substack{i,j=0\\ i\neq j}}^{n} \left\langle c_2, \rho \otimes \tau_{(h_{n,i}-h_{n,j})}\rho \right\rangle + \frac{(1-|\rho|)^2}{n^2} \sum_{j=1}^{n} D_2(\rho) \\
&\le D_2(\rho) + \frac{2(1-|\rho|)}{n} + \frac{(1-|\rho|)^2}{n} D_2(\rho),
\end{aligned}$$

whence

$$\liminf_{n\to\infty} D(\rho_n) \le D_2(\rho).$$

If $\rho \in \mathcal{P}_-(\mathbb{R}^d)$ does not have compact support, consider for every $n \in \mathbb{N}$ the measure $\rho_n := \rho\lfloor_{B(0,n)}$. Notice that $\rho_n \rightharpoonup \rho$, since $\left|\int f(x)d(\rho(x)-\rho_n(x))\right| \le \sup|f|\,\rho(B(0,n)^c) \to 0$ for every $f \in C_0(\mathbb{R}^d)$ — actually, the same holds for every $f \in C_b(\mathbb{R}^d)$, *i.e.*, $\rho_n \to \rho$ tightly. Then

$$\overline{D}(\rho) \le \liminf_{n\to\infty} \overline{D}(\rho_n) \le \liminf_{n\to\infty} D_2(\rho_n) \le D_2(\rho).$$

This concludes the proof of Theorem 3.2.15. $\square$

Understanding in general the behaviour of D^{**}, even on probabilities, seems to be a difficult problem. This difficulty would be very much simplified, at least for sub-probabilities, if we could prove some homogeneity.

Let us fix $\rho \in \mathcal{P}(\mathbb{R}^d)$ and $\theta \in [0,1]$, and go back to the stratification formula (3.4), reported here for the sake of the reader:

$$\overline{C_N}(\theta\rho) = \inf_{\substack{a_1,\dots,a_N \ge 0\\ \rho_1,\dots,\rho_N \in \mathcal{P}(\mathbb{R}^d)}} \left\{ \sum_{k=2}^{N} a_k \frac{k(k-1)}{N(N-1)} C_k(\rho_k) \mid \sum_{k=1}^{N} a_k \le 1, \sum_{k=1}^{N} \frac{k}{N} a_k \rho_k = \theta\rho \right\}.$$

In an optimal choice $a_1, \dots, a_N$ there will be a minimum and a maximum index k such that $a_k > 0$, which we denote respectively by $\underline{k}(N, \theta\rho)$ and $\overline{k}(N, \theta\rho)$. Observe that

$$\sum_{k=1}^{N} \frac{k}{N} a_k \rho_k = \theta\rho \implies \sum_{k=1}^{N} k a_k = \theta N,$$

hence necessarily $\underline{k}(N, \theta\rho) \le \theta N \le \overline{k}(N, \theta\rho)$.

Conjecture 3.2.17. *The gap between* $\underline{k}(N, \theta\rho)$ *and* $\overline{k}(N, \theta\rho)$ *vanishes (with respect to* N*) as* $N \to \infty$, i.e.,

$$\limsup_{N\to\infty} \frac{\underline{k}(N, \theta\rho)}{N} = \liminf_{N\to\infty} \frac{\overline{k}(N, \theta\rho)}{N} = \theta.$$

This is weak version of the statement

$$\underline{k}(N, \theta\rho) = \lfloor \theta N \rfloor, \quad \overline{k}(N, \theta\rho) = \lceil \theta N \rceil,$$

which was initially conjectured by some people in the field but is probably not true, according to some recent works in preparation by S. Di Marino, M. Lewin and L. Nenna.[1]

An immediate application of Conjecture 3.2.17 is given in the following

Theorem 3.2.18. *Suppose that Conjecture 3.2.17 holds. Then the functional* C_∞ *is 2-homogeneous,* i.e.,

$$C_\infty(\theta\rho) = \theta^2 C_\infty(\rho) \quad \forall \rho \in \mathcal{P}(\mathbb{R}^d), \forall \theta \in [0, 1].$$

Proof. It suffices to consider $\theta \in (0, 1)$. One inequality is true independently from Conjecture 3.2.17, namely $C_\infty(\theta\rho) \le \theta^2 C_\infty(\rho)$. Indeed, from Corollary 3.2.6, by choosing a suitable sequence $(k_N)_{N\ge 2}$ such that $\lim \frac{k_N}{N} = \theta$, we get

$$M_\infty(\theta v) \ge \theta^2 M_\infty(v),$$

whence

$$C_\infty(\rho) = \sup_{v\in C_0(\mathbb{R}^d)} \langle v, \rho\rangle - M_\infty(v) \ge \frac{1}{\theta^2} \sup_{v\in C_0(\mathbb{R}^d)} \langle \theta v, \theta\rho\rangle - M_\infty(\theta v) = \frac{1}{\theta^2} C_\infty(\theta\rho).$$

For the converse, let $\rho_1, \dots, \rho_N$ and $a_1, \dots, a_N$ optimal in (3.4) for $\theta\rho$, with $a_1 = \dots = a_{k-1} = 0$. Then by convexity of C_k we have

$$\begin{aligned}\overline{C_N}(\theta\rho) &= \sum_{j=k}^{N} a_j \frac{j(j-1)}{N(N-1)} C_j(\rho_j) \\ &\ge \frac{k-1}{N-1} \sum_{j=k}^{N} a_j \frac{j}{N} C_k(\rho_j) \ge \frac{k-1}{N-1} \theta C_k(\rho).\end{aligned}$$

If there exists a sequence $(N_h)_{h\in\mathbb{N}}$ such that $\lim_{h\to\infty} \frac{\underline{k}(N_h, \theta\rho)}{N_h} = \theta$ (Conjecture 3.2.17), then we can conclude that $C_\infty(\theta\rho) \ge \theta^2 C_\infty(\rho)$, as wanted. □

[1] Personal communication.

3.3 Minimizers and ionization effect

Finally we come back to the problem stated in the introduction, namely a minimum problem of the form

$$\inf_{\rho\in\mathcal{P}(\mathbb{R}^d)} \{C_N(\rho)+\mathcal{F}(\rho)\} \tag{3.8}$$

for some functional $\mathcal{F}(\rho)$. We want to address the instances of this problem when $\mathcal{F}$ is linear and continuous, *i.e.*,

$$\mathcal{F}(\rho) = -\langle v,\rho\rangle$$

for some potential $v \in C_0$. As we already observed, (3.8) in this case makes sense only when $\sup v > 0$, otherwise the infimum is zero, and a minimizing sequence vanishes weakly by spreading mass to infinity.

Fix thus $v \in C_0(\mathbb{R}^d)$ such that $\sup v > 0$, and consider the function

$$\begin{aligned} g_v\colon (0,+\infty) &\to (0,\sup v] \\ \lambda &\mapsto \frac{M_\infty(\lambda v)}{\lambda}. \end{aligned} \tag{3.9}$$

Proposition 3.3.1. *Let g_v defined as in* (3.9)*. Then*

(i) g_v is increasing;

(ii) $\lim_{\lambda\to+\infty} g_v(\lambda) = \sup v$.

Proof. Since M_∞ is convex and $M_\infty(0)=0$,

$$M_\infty(\lambda v) = M_\infty\left(\frac{\lambda}{\lambda'}\lambda' v + \frac{\lambda'-\lambda}{\lambda'}0\right) \le \frac{\lambda}{\lambda'}M_\infty(\lambda' v) \quad \forall 0<\lambda\le\lambda',$$

thus proving (i).

(ii) Using Corollary 3.2.7 and Lemma 3.2.2, we have

$$M_\infty(\lambda v) \le M_N(\lambda v) \le \lambda \sup v,$$

whence

$$\limsup_{\lambda\to+\infty} \frac{M_\infty(\lambda v)}{\lambda} \le \sup v.$$

By Corollary 3.2.10, given $\epsilon > 0$ let $\rho_\epsilon \in \mathcal{P}(\mathbb{R}^d)$ such that:

- $\langle v,\rho_\epsilon\rangle \ge (1-\epsilon)\sup v$;
- $C_\infty(\rho_\epsilon) < +\infty$.

This can be achieved by taking a sequence $\{\rho_n\}$ in the domain of C_∞ such that $\rho_n \rightharpoonup \delta_x$, where $v(x) = \sup v$, and letting $\rho_\epsilon = \rho_n$ for a suitable n large enough. Then

$$M_\infty(\lambda v) \geq \lambda \langle v, \rho_\epsilon \rangle - C_\infty(\rho_\epsilon) \geq \lambda(1-\epsilon) \sup v - C_\infty(\rho_\epsilon),$$

whence

$$\liminf_{\lambda \to +\infty} \frac{M_\infty(\lambda v)}{\lambda} \geq (1-\epsilon) \sup v.$$

Since ϵ was arbitrary we get

$$\sup v \leq \liminf_{\lambda \to +\infty} \frac{M_\infty(\lambda v)}{\lambda} \leq \limsup_{\lambda \to +\infty} \frac{M_\infty(\lambda v)}{\lambda} \leq \sup v,$$

as wanted. □

Given $v \in C_0(\mathbb{R}^d)$ such that $\sup v > 0$, we denote $\mathcal{M}(v)$ the set of minimizers of the functional $\rho \mapsto C_\infty(\rho) - \langle v, \rho \rangle$. In other words, for every $\rho \in \mathcal{M}(v)$ we have $M_\infty(v) = \langle v, \rho \rangle - C_\infty(\rho)$. Here we see clearly the link between the minimum problem (3.8) and the Γ-limit of the Legendre-Fenchel conjugate M_∞.

It is of great interest to understand if the minimizers $\mathcal{M}(v)$ are probabilities or not. When the mass of a minimizing density ρ is less than one we have the so-called *ionization phenomenon*, taking the name from the case where ρ denotes the charge density of N electrons. For this reason we want to state, up to our knowledge, some conditions to have probability solutions to the problem (3.8).

Theorem 3.3.2. *Suppose that C_∞ is 2-homogeneous,* i.e., $C_\infty(t\rho) = t^2 C_\infty(\rho)$ $\forall t \in [0,1], \forall \rho \in \mathcal{P}(\mathbb{R}^d)$. *Then*

$$|\rho| \geq \min\left\{1, \frac{2M_\infty(v)}{\sup v}\right\} \quad \forall \rho \in \mathcal{M}(v).$$

Proof. Let $\rho \in \mathcal{M}(v)$. If ρ is a probability, there is nothing to prove, so assume that $\rho = t_0\mu_0$ for some $t_0 \in [0,1)$, $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$. In particular, t_0 must minimize the function $g(t) = \langle v, t\mu_0 \rangle - C_\infty(t\mu_0) = t\langle v, \mu_0 \rangle - t^2 C_\infty(\mu_0)$. Hence, by differentiating, $t_0 = \frac{\langle v, \mu_0 \rangle}{2C_\infty(\mu_0)}$. Recall however that $\rho \in \mathcal{M}(v)$, whence

$$M_\infty(v) = \langle v, \rho \rangle - C_\infty(\rho) = t_0 \langle v, \mu_0 \rangle - t_0^2 C_\infty(\mu_0) = \frac{t_0}{2} \langle v, \mu_0 \rangle \leq \frac{t_0}{2} \sup v,$$

as wanted, since $t_0 = |\rho|$. □

The Theorem 3.3.2 provides a sufficient condition in order to assure that the minimum problem has only probability solutions.

Corollary 3.3.3. *Suppose that C_∞ is 2-homogeneous, and let $v \in C_0(\mathbb{R}^d)$ such that $M_\infty(v) \geq \frac{\sup v}{2} > 0$. Then $\mathcal{M}(v) \subseteq \mathcal{P}(\mathbb{R}^d)$.*

Proof. It follows immediately from Theorem 3.3.2. □

When we make the potential grow in size, there is a threshold past which we always have probability solutions. This is reasonable: due to the minus sign in front of the potential term, we are putting a very strong confining potential, which does not allow the mass (or charge) to escape at infinity.

Proposition 3.3.4. *For every $v \in C_0$ such that $\sup v > 0$, there exists a threshold $\lambda^*(v) \geq 0$ such that*

$$\mathcal{M}(\lambda v) \subseteq \mathcal{P}(\mathbb{R}^d) \quad \forall \lambda \geq \lambda^*(v).$$

Proof. In view of Corollary 3.3.3, it suffices to show that $M_\infty(\lambda v) \geq \frac{\lambda}{2} \sup v$ for λ large enough. This follows from Proposition 3.3.1. □

3.A Explicit computations for $\overline{C_N}$

An explicit computation of the relaxed transport cost $\overline{C_N}(\rho)$ for a general sub-probability $\rho \in \mathcal{P}_-(\mathbb{R}^d)$ is often much involved, and in most case impossible to carry out. In this Section we present, to the best of our knowledge, some examples for atomic probabilities. We begin with the following result about a linear programming problem.

Lemma 3.A.1. *Let $0 \leq \theta \leq 1$ and $N \geq 2$. Then*

$$\max_{a_0,\dots,a_N \geq 0} \left\{ \sum k^2 a_k \mid \sum a_k = 1, \sum k a_k = \theta N \right\} = \theta N^2,$$
$$\min_{a_0,\dots,a_N \geq 0} \left\{ \sum k^2 a_k \mid \sum a_k = 1, \sum k a_k = \theta N \right\} = \theta^2 N^2 + \{\theta N\}(1 - \{\theta N\}),$$

where $\{\theta N\}$ denotes the fractional part of θN.

Proof. Let $a_0, \dots, a_N \geq 0$ be admissible parameters. Observe that for every k we have $\left|k - \frac{N}{2}\right| \leq \frac{N}{2}$, which yields

$$\sum_{k=0}^N k^2 a_k = \sum_{k=0}^N \left(k - \frac{N}{2}\right)^2 a_k + \theta N^2 - \frac{N^2}{4} \leq \frac{N^2}{4} + \theta N^2 - \frac{N^2}{4} = \theta N^2.$$

On the other hand, the choice $a_0 = 1 - \theta$, $a_N = \theta$, $a_k = 0$ for $k \neq 0, N$, fulfills the constraints and gives $\sum k^2 a_k = \theta N^2$.

Let $k_0 \leq \theta N \leq k_0 + 1$. We claim that an optimal solution for the minimum problem is given by $a_{k_0} = k_0 + 1 - \theta N$, $a_{k_0+1} = \theta N - k_0$, $a_k = 0$ if $k \neq k_0, k_0+1$. Indeed, let $b_0, \dots, b_N$ any admissible choice of parameters, observe that

$$a_{k_0} = \sum_{k=0}^N (k_0 + 1 - k) b_k, \quad a_{k_0+1} = \sum_{k=0}^N (k - k_0) b_k,$$

whence

$$
\begin{aligned}
k_0^2 a_{k_0} + (k_0+1)^2 a_{k_0+1} &= \sum_{k=0}^{N} (k_0^3 + k_0^2 - k_0^2 k + (k_0+1)^2 k - (k_0+1)^2 k_0) b_k \\
&= \sum_{k=0}^{N} (2k_0 k + k - k_0^2 - k_0) b_k \\
&= \sum_{k=0}^{N} k^2 b_k - \sum_{k=0}^{N} ((k_0-k)^2 + (k_0 - k)) b_k.
\end{aligned}
$$

Since each term of the last sum is positive (both if $k \le k_0$ and if $k \ge k_0$), we proved that the choice $a_{k_0} = k_0 + 1 - \theta N = 1 - \{\theta N\}$, $a_{k_0+1} = \theta N - k_0 = \{\theta N\}$ is optimal. Finally, notice that

$$
\begin{aligned}
k_0^2 a_{k_0} + (k_0+1)^2 a_{k_0+1} &= (\theta N - k_0)^2 a_{k_0} + (k_0 + 1 - \theta N)^2 a_{k_0+1} + \theta^2 N^2 \\
&= \theta^2 N^2 + a_{k_0+1}^2 a_{k_0} + a_{k_0}^2 a_{k_0+1} = \theta^2 N^2 + a_{k_0} a_{k_0+1}. \quad \square
\end{aligned}
$$

We want to compute the relaxed cost of a Dirac delta. Since every transport plan must be a Dirac delta on its own, concentrated on a point $X \in \mathbb{R}^d$ with equal coordinates, it is necessary to assume that $\ell(0)$ is finite in order to get a sensible result. From now on we assume thus that $\ell(0) < +\infty$.

Theorem 3.A.2. *Let $\theta \in [0,1]$. Then*

$$
\overline{C_N}(\theta\delta_0) = \begin{cases} 0 & \text{if } \theta N \le 1 \\ \left(\dfrac{\theta^2 N}{N-1} - \dfrac{\theta}{N-1} + \dfrac{\{\theta N\}(1 - \{\theta N\})}{N(N-1)} \right) \ell(0) & \text{otherwise.} \end{cases}
$$

Proof. As we already observed in section 3.2, if $\theta = |\theta\delta_0| \le \frac{1}{N}$, then $\overline{C_N}(\theta\delta_0) = 0$. Assume hence that $\theta N > 1$. Recalling that $C_k(\delta_0) = \ell(0)$ for every $k \ge 2$, by (3.4) we have

$$
\begin{aligned}
\overline{C_N}(\theta\delta_0) &= \frac{2\ell(0)}{N(N-1)} \inf_{a_1,\dots,a_N \ge 0} \left\{ \sum_{k=2}^{N} \frac{k(k-1)}{2} a_k \mid \sum_{k=1}^{N} a_k = 1, \sum_{k=1}^{N} \frac{k}{N} a_k = \theta \right\} \\
&= \frac{2\ell(0)}{N(N-1)} \inf_{a_0,\dots,a_N \ge 0} \left\{ \sum_{k=0}^{N} \frac{k(k-1)}{2} a_k \mid \sum_{k=0}^{N} a_k = 1, \sum_{k=0}^{N} \frac{k}{N} a_k = \theta \right\}.
\end{aligned}
$$

Indeed, every $a_1, \dots, a_N$ admissible in the first line produces a choice of parameters for the second line by letting $a_0 = 0$. Conversely, as we deduce from the proof of Lemma 3.A.1, since $\theta N > 1$, in an optimal solution for the second line we have $a_0 = 0$, and we get a choice of parameters for the first line. In view of Lemma 3.A.1, we conclude that

$$
\overline{C_N}(\theta\delta_0) = \frac{2\ell(0)}{N(N-1)} \left(\frac{\theta^2 N^2 + \{\theta N\}(1 - \{\theta N\})}{2} - \frac{\theta N}{2} \right),
$$

as wanted. $\square$

Recall that, by monotonicity, the Γ-limit of $\overline{C_N}$ is also the pointwise limit of C_N. Observe that, by Theorem 3.A.2, we have

$$C_\infty(\theta\delta_x) = \lim_{N\to\infty} \overline{C_N}(\theta\delta_0) = \theta^2\ell(0) = \theta^2 C_\infty(\delta_x),$$

since

$$\overline{C_N}(\delta_x) = C_N(\delta_x) = \ell(0),$$

as the set of transport plans consists only of the element $\rho \otimes \cdots \otimes \rho$. Thus, at least in the case when $\ell(0) < \infty$, we get 2-homogeneity of C_∞ for Dirac masses, in agreement with Theorem 3.2.18.

The next step is to compute the relaxed transport cost for a convex combination of two Dirac masses. As before, it is necessary to assume that $\ell(0)$ is finite.

Theorem 3.A.3. *Let $0 \le \theta \le 1$, and $\rho = \theta\delta_x + (1-\theta)\delta_y$. Then*

$$C_N(\rho) = [1 - \gamma(N,\theta)]\ell(0) + \gamma(N,\theta)\min\{\ell(|x-y|), \ell(0)\},$$

where

$$\gamma(N,\theta) = \frac{2\theta(1-\theta)N}{N-1} - \frac{2\{\theta N\}(1-\{\theta N\})}{N(N-1)}.$$

Proof. Every symmetric N-transport plan P_N with marginals ρ is given by

$$P_N = \sum_{k=0}^N a_k \operatorname{Sym}(\underbrace{\delta_x \otimes \cdots \otimes \delta_x}_{k \text{ times}} \otimes \underbrace{\delta_y \otimes \cdots \otimes \delta_y}_{N-k \text{ times}})$$

where the a_k's must satisfy $a_k \ge 0$, $\sum a_k = 1$ and the marginal condition. The latter may be written as follows:

$$\sum_{k=0}^N \frac{k}{N} a_k = \theta, \quad \sum_{k=0}^N \frac{N-k}{N} a_k = 1 - \theta. \tag{3.10}$$

Indeed, this is a general fact about the Sym operation: every marginal of $\operatorname{Sym}(\mu)$ is the arithmetic mean of the N marginals of μ.

In order to compute the cost of P_N, it is useful to observe that, due to the pairwise structure of the cost,

$$\langle c_N, P_N\rangle = \int_{\mathbb{R}^d\times\mathbb{R}^d} \ell(|u-v|)d\pi_2(P_N)(u,v), \tag{3.11}$$

where $\pi_2(P_N)$ denotes the 2-marginals projection of P_N.

Let us compute $\pi_2(P_N)$. Consider the $\binom{N}{k}$ different permutations of the string $(\underbrace{x,\dots,x}_{k \text{ times}}, \underbrace{y,\dots,y}_{N-k \text{ times}})$ given by the Sym operation: there are[2]

$$\begin{aligned}
&\binom{N-2}{k-2} \text{ of them starting with } (x,x),\\
&\binom{N-2}{k} \text{ of them starting with } (y,y),\\
&\binom{N-2}{k-1} \text{ of them starting with either } (x,y) \text{ or } (y,x).
\end{aligned}$$

The 2-marginals projection of P_N, which can be computed by integrating out all the variables except the first two, is given hence by

$$\begin{aligned}
\pi_2(P_N) = \sum_{k=0}^{N} \frac{a_k}{\binom{N}{k}} \Bigg[&\binom{N-2}{k-2} \delta_x \otimes \delta_x + \binom{N-2}{k} \delta_y \otimes \delta_y \\
&+ \binom{N-2}{k-1} \delta_x \otimes \delta_y + \binom{N-2}{k-1} \delta_x \otimes \delta_y \Bigg]
\end{aligned}$$

Let us define for simplicity $\alpha_k = \frac{\binom{N-2}{k-2}+\binom{N-2}{k}}{\binom{N}{k}}$, $\beta_k = 2\frac{\binom{N-2}{k-1}}{\binom{N}{k}}$, and observe that $\alpha_k + \beta_k = 1$ by the recursion formula for binomial coefficients. Recalling (3.11), the transport cost of P_N is given by

$$\langle c_N, P_N \rangle = \sum_{k=0}^{N} a_k \left[\alpha_k \ell(0) + \beta_k \ell(|x-y|) \right] = \ell(0) + \left[\ell(|x-y|) - \ell(0) \right] \sum_{k=0}^{N} a_k \beta_k$$

Observe that

$$\sum_{k=0}^{N} a_k \beta_k = \frac{2}{N(N-1)} \sum_{k=0}^{N} (-k^2 + kN) a_k = \frac{2\theta N}{N-1} - \frac{2}{N(N-1)} \sum_{k=0}^{N} k^2 a_k.$$

If $\ell(|x-y|) \geq \ell(0)$, by Lemma 3.A.1 the minimal value is

$$\sum_{k=0}^{N} a_k \beta_k = \frac{2\theta N}{N-1} - \frac{2}{N(N-1)} \theta N^2 = 0.$$

If $\ell(|x-y|) < \ell(0)$, again by Lemma 3.A.1 the minimal value is

$$\begin{aligned}
\sum_{k=0}^{N} a_k \beta_k &= \frac{2\theta N}{N-1} - \frac{2}{N(N-1)} \left(\theta^2 N^2 + \{\theta N\} (1 - \{\theta N\}) \right) \\
&= \frac{2\theta(1-\theta)N}{N-1} - \frac{2\{\theta N\}(1-\{\theta N\})}{N(N-1)} = \gamma(N, \theta).
\end{aligned}$$

□

[2] Here and in the following, we adopt the convention that $\binom{N}{k} = 0$ if $k \leq 0$ or $k > N$.

As a consequence we get

$$C_\infty(\theta\delta_x + (1-\theta)\delta_y) = (\theta^2 + (1-\theta^2))\ell(0) + 2\theta(1-\theta)\min\{\ell(|x-y|), \ell(0)\}.$$

Let $\rho = \theta\delta_x + (1-\theta)\delta_y$. It may be interesting to observe that, in the case when $\ell(0) > \ell(|x-y|)$ we get

$$C_\infty(\rho) = \int \ell(|x' - y'|) d\rho(x') d\rho(y'),$$

as in the Pass' result [36], while in the case $\ell(0) \leq \ell(|x-y|)$ we get a Monge-type integral

$$C_\infty(\rho) = \int \ell(|x' - T(x')|) d\rho(x'),$$

with $T = \mathrm{id}$.

Chapter 4

Smoothing of transport plans with fixed marginals

In this Chapter we focus on a multi-marginal Optimal Transport problem on the Euclidean space. Suppose we are given N Borel probability measures $\rho_1, \dots, \rho_N \in \mathcal{P}(\mathbb{R}^d)$, and a transport plan $\mu \in \Pi(\rho_1, \dots, \rho_n)$ — usually it will be an optimal transport plan associated to some cost function. As we already know from the previous chapters, even when $\rho_1, \dots, \rho_n$ share some regularity properties (*e.g.*, they are absolutely continuous w.r.t. the Lebesgue measure, or their densities are in some class of regular functions), typically the transport plan μ will not share the same regularity. In fact, optimal transport plans tends to concentrate on sets of zero Lebesgue measure, as is for instance the case when there is a Monge-type solution.

When considering quantum systems of particles, the Schrödinger equation can be naturally stated for wave-functions with Sobolev regularity, *i.e.*, a wave-function ψ lies in $H^1((\mathbb{R}^d)^N)$. If we consider, according to the Born interpretation, the measure $d\mu(X) = |\psi(X)|^2\, dX$ on $(\mathbb{R}^d)^N$ as the probability distribution of finding the particles in positions $X = (x_1, dotsc, x_N)$, then the marginals of μ are given by

$$\rho_j(x_j) = \int_{(\mathbb{R}^d)^{N-1}} |\psi(x_1, \dots, x_N)|^2 \,\mathrm{d}x_1 \cdots \mathrm{d}x_{j-1}\,\mathrm{d}x_{j+1} \cdots \mathrm{d}x_N.$$

It is not difficult to show (see for instance [34, Theorem 1.1]) that in this case $\sqrt{\rho_j} \in H^1(\mathbb{R}^d)$ for every $j = 1, \dots, N$. For this reason, we will concentrate on the case when the measures $\rho_1, \dots, \rho_N$ have a Sobolev-type regularity, as clarified in the following

Definition 6. If $p > 1$, we say that a probability measure $\mu \in \mathcal{P}(\mathbb{R}^m)$ is $W^{1,p}$-*regular* if μ is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^m$, and

$$\left(\frac{\mathrm{d}\mu}{\mathrm{d}\mathcal{L}^m}\right)^{1/p} \in W^{1,p}(\mathbb{R}^m).$$

In other words, μ is $W^{1,p}$-regular if there exists $f \in W^{1,p}(\mathbb{R}^m)$, $f \geq 0$, such that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mathcal{L}^m} = f^p.$$

Since in the following we will use this definition both for measures on $\mathbb{R}^d$ and on $(\mathbb{R}^d)^N$, we keep a generic dimension m for the Euclidean space. We will denote by $\mathcal{P}_p(\mathbb{R}^m)$ the space of $W^{1,p}$-regular probability measures. This definition arises naturally in the setting of Density Functional Theory as a generalization of the one given by Lieb in [34] for $p = 2$, but since the theory works for every Sobolev exponent $p > 1$, we prefer to keep it generic. Also, in the Hilbertian case $p = 2$, some results enjoy a simplified proof and sharper constants, as we will show in due time.

From here on, p will be a fixed real number greater than 1. The set $\mathcal{P}_p(\mathbb{R}^m)$ has a natural structure of metric space if endowed with the distance

$$d_p(\mu, \nu) = \left\| \left(\frac{\mathrm{d}\mu}{\mathrm{d}\mathcal{L}^m} \right)^{1/p} - \left(\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^m} \right)^{1/p} \right\|_{W^{1,p}},$$

which can be seen as a refined version of the Hellinger distance between two absolutely continuous probability measures, where the L^p norm of the p-th roots is replaced by the $W^{1,p}$ norm. We delay to Section 4.1 a more detailed study of the metric space $\mathcal{P}_p(\mathbb{R}^m)$.

As we already noticed, even when the marginals $\rho_1, \ldots, \rho_N$ are $W^{1,p}$-regular, the plan μ will in general be singular. On the other hand, for many applications, and in particular when dealing with Γ-convergence, it is useful to have regular transport plans which are "close" to a given optimal one (see for instance [4, 19, 32]). With this in mind, we want to address the following

Problem: *Given* $\rho_1, \ldots, \rho_N \in \mathcal{P}_p(\mathbb{R}^d)$, *and given* $\mu \in \Pi(\rho_1, \ldots, \rho_N)$, *find a family* $(\mu_\varepsilon)_{\varepsilon>0}$ *such that:*

(i) $\mu_\varepsilon \in \mathcal{P}_p((\mathbb{R}^d)^N)$ *for every* $\varepsilon > 0$*;*

(ii) $\mu_\varepsilon \in \Pi(\rho_1, \ldots, \rho_N)$*;*

(iii) $\mu_\varepsilon \to \mu$ *as* $\varepsilon \to 0$ *(for a suitable notion of convergence).*

In other words, we search for $W^{1,p}$-regular multi-marginal transport plans with marginals $\rho_1, \ldots, \rho_N$ which approximate a (non regular) transport plan μ. Since in general μ does not have any regularity property, the natural topology for (iii) is the tight convergence of probability measures, *i.e.*, weak convergence in duality with $C_b\left((\mathbb{R}^d)^N\right)$ (continuous and bounded functions).

Notice that, if μ is an optimal transport plan for some cost function c, and the cost function is upper semi-continuous and bounded from above, combining (iii) and the Portmanteau's Theorem we get

$$\lim_{\varepsilon \to 0} \int c(X) \, \mathrm{d}\mu^\varepsilon(X) = \int c(X) \, \mathrm{d}\mu(X),$$

whence we may say that μ^ε is "almost" optimal for small ε.

One could think that a very common technique for regularizing, namely the convolution with a smooth kernel, should be a good approach for dealing with this type of problems. However, the marginal constraint, which is crucial in all applications in optimal transport, is not stable under any convolution operation, which makes the problem not trivial.

The main result of this chapter is the following

Theorem 4.0.1. *Let $\mu \in \mathcal{P}((\mathbb{R}^d)^N)$ such that $\pi^k_\# \mu \in \mathcal{P}_p(\mathbb{R}^d)$ for every $k = 1, \dots, N$. Then for every $\varepsilon > 0$ there exists $\Theta^\varepsilon[\mu] \in \mathcal{P}_p(\mathbb{R}^d)$ such that the following hold.*

(i) $\pi^k_\# \Theta^\varepsilon[\mu] = \pi^k_\# \mu$ for every $k = 1, \dots, N$.

(ii) For every $\varepsilon > 0$,
$$W_2(\Theta^\varepsilon[\mu], \mu) \leq C(d)\sqrt{N\varepsilon},$$
where W_2 denotes the Wasserstein distance. If moreover $\mu \in \mathcal{P}_p((\mathbb{R}^d)^N)$, then $\Theta^\varepsilon[\mu] \to \mu$ in the d_p-metric as $\varepsilon \to 0$.

Observe that Theorem 4.0.1.(ii) implies that $\Theta^\varepsilon[\mu] \rightharpoonup \mu$ as $\varepsilon \to 0$, as wanted. We will call $\Theta^\varepsilon[\mu]$ given by Theorem 4.0.1 a "smoothing operator", viewed as a functional
$$\begin{aligned} \Theta \colon (0, +\infty) \times \mathcal{P}\left((\mathbb{R}^d)^N\right) &\longrightarrow \mathcal{P}_p\left((\mathbb{R}^d)^N\right) \\ (\varepsilon, \mu) &\longmapsto \Theta^\varepsilon[\mu] \end{aligned}$$

The construction of the operator Θ will depend on the choice of a function $\eta \in C^\infty(\mathbb{R}^d)$, which will play the role of a convolution kernel. Depending on the choice of η we will have additional properties, as stated in the following

Theorem 4.0.2. *Let η be supported on $B(0,1)$, and $\Omega \subseteq (\mathbb{R}^d)^N$ such that $\Omega + B(0,r) \subset (\mathbb{R}^d)^N \setminus \operatorname{supp} \mu$ for some $r > 0$. Then $\Theta^\varepsilon(\mu) = 0$ on Ω as soon as $\varepsilon < r/2$.*

Theorem 4.0.3. *Let η be a Gaussian kernel, and suppose that $\mu_n \rightharpoonup \mu$, and $\pi_\# \mu_n \to \pi_\# \mu$ in the d_p-topology. Then $\Theta^\varepsilon(\mu_n) \to \Theta^\varepsilon(\mu)$ in the d_p-topology for every $\varepsilon > 0$.*

In other words, when taking a Gaussian kernel, weak convergence is reinforced to Sobolev convergence as soon as the smoothing parameter ε is positive, and the marginals converge in the Sobolev sense.

This results will prove very useful for the applications in the final chapter, where we will choose the kernel η accordingly to our needs.

Finally, we want to point out that the definition of the smoothing operator, which we give in the case of Sobolev spaces due to physical interest, works in the same way for other classes of absolutely continuous measures, *e.g.*, measures with $C^{k,\alpha}$ density, with analogous regularity and continuity results.

Structure of the chpater The material of this chapter derives mainly from [3]. In Section 4.1 we will introduce some notation and some preliminary results about p-th powers and p-th roots of non-negative Sobolev functions. Moreover a short overview of the space $\mathcal{P}_p$ is given, with special attention to the map sending a measure to its marginals. Many of the proofs for this Section will be put in the Appendix, in order to focus better on the proof of Theorem 4.0.1.

In Section 4.2 we will present the proof of Theorem 4.0.1. In Section 4.3 we will prove Theorem 4.3.1 and Theorem 4.3.2. Finally, in the Appendix we will complete the missing proofs from Section 4.1 and Section 4.2.

4.1 Preliminary results and the space $\mathcal{P}_p$

Notation

If $f\colon (\mathbb{R}^d)^N \to \mathbb{R}$, and $1 \leq k \leq N$, we denote by

$$\int f(X)\,\mathrm{d}\hat{X}_k := \int f(x_1, \dots, x_N)\,\mathrm{d}x_1 \cdots \mathrm{d}\hat{x}_k \cdots \mathrm{d}x_N$$

the integral of f with respect to all the variables except x_k. This is a function of the variable x_k.

When $f \in W^{1,p}(\mathbb{R}^m)$, we will adopt the convention that

$$|\nabla f| := \left(\sum_{j=1}^{m} |\partial_{x_k} f|^p \right)^{1/p}, \tag{4.1}$$

i.e., when computing the norm of a gradient we take on $\mathbb{R}^m$ the p-th norm.

Roots and powers of Sobolev functions

When dealing with a smooth non-negative function u, we know that $\nabla(u^\alpha) = \alpha u^{\alpha-1}\nabla u$. This is also true for Sobolev functions if the RHS has the right summability. To make everything clear we state the following results, which will be useful later in order to have an expression for the weak derivatives of p-th powers and p-th roots of non-negative Sobolev functions. The proofs of Proposition 4.1.1 and Proposition 4.1.2 will be given in the Appendix.

Proposition 4.1.1. *Let $p > 1$. If $u \in W^{1,p}(\mathbb{R}^m)$, $u \geq 0$, then $u^p \in W^{1,1}(\mathbb{R}^m)$, and $\nabla u^p = pu^{p-1}\nabla u$.*

Viceversa, let $u \in W^{1,1}(\mathbb{R}^m)$, $u \geq 0$, such that

$$\int u^{1-p}\,|\nabla u|^p < \infty. \tag{4.2}$$

Then $u^{1/p} \in W^{1,p}(\mathbb{R}^m)$, and $\nabla u^{1/p} = \frac{1}{p} u^{\frac{1-p}{p}} \nabla u$.

The condition (4.2) in Proposition 4.1.1 is necessary, as the following example shows.

Example 1. In dimension $m = 1$, fix $p > 1$ and consider the $W^{1,1}$ function

$$f(x) = \begin{cases} \sin(x)^{p-1} & 0 \le x \le \pi \\ 0 & \text{otherwise,} \end{cases}$$

whose weak derivative is $f'(x) = \chi_{[0,\pi]}(p-1)\sin(x)^{p-2}\cos(x)$. However, $f^{1/p}$ does not belong to $W^{1,p}(\mathbb{R})$, since the weak derivative of $f^{1/p}$ should be $g(x) = \frac{p-1}{p}\chi_{[0,\pi]}\sin(x)^{-\frac{1}{p}}\cos(x)$, but

$$\int_0^\pi |g(x)|^p \, \mathrm{d}x = \frac{(p-1)^p}{p^p}\int_0^\pi \frac{|\cos(x)|^p}{\sin(x)}\,\mathrm{d}x$$

diverges at both 0 and π.

Proposition 4.1.2. *If $u_n \to u$ in $W^{1,p}(\mathbb{R}^m)$, $u_n, u \ge 0$, then $u_n^p \to u^p$ in $W^{1,1}(\mathbb{R}^m)$.*

Viceversa, let $u_n \to u$ in $W^{1,1}(\mathbb{R}^d)$, $u_n, u \ge 0$. Let $h_n, h \in L^1(\mathbb{R}^m)$ such that $u_n^{1-p}|\nabla u_n|^p \le h_n$, $u^{1-p}|\nabla u|^p \le h$, and

$$\lim_{n\to\infty}\int h_n = \int h. \tag{4.3}$$

Suppose also that for every subsequence $\{h_{n_k}\}$ there exists a further subsequence converging to h pointwise a.e. Then $u_n^{1/p} \to u^{1/p}$ in $W^{1,p}(\mathbb{R}^m)$.

The space $\mathcal{P}_p$ of regular measures

We aim to study the space $\left(\mathcal{P}_p\left((\mathbb{R}^d)^N\right), d_p\right)$ in relation with the map which sends a $W^{1,p}$-regular probability onto its marginals, namely

$$\begin{aligned} \pi\colon \mathcal{P}_p\left((\mathbb{R}^d)^N\right) &\longrightarrow \mathcal{P}(\mathbb{R}^d)^N \\ \mu &\longmapsto \left(\mu^1, \dots, \mu^N\right). \end{aligned} \tag{4.4}$$

We have the following

Proposition 4.1.3. *Let $p > 1$. Then*

(i) if $\mu \in \mathcal{P}_p((\mathbb{R}^d)^N)$, then $\mu^k \in \mathcal{P}_p(\mathbb{R}^d)$ for every $k = 1, \dots, N$;

(ii) the map $\pi\colon \mathcal{P}_p\left((\mathbb{R}^d)^N\right) \longrightarrow \mathcal{P}_p(\mathbb{R}^d)^N$ is continuous with respect to the distance d_p and the relative product topology on the codomain.

This will be proved in the Appendix. We remark that Proposition 4.1.3(ii) was alredy proved by Brezis in [34, Appendix] in the case $p = 2$. In what follows, if μ is $W^{1,p}$-regular, with a slight abuse of notation we will denote by $\mu(X)$ its density, whose p-th root belongs to $W^{1,p}\left((\mathbb{R}^d)^N\right)$. For $k = 1, \dots, N$ let

$$\mu^k(x_k) = \int \mu(X)\, \mathrm{d}\hat{X}_k, \quad \nabla\mu^k(x_k) = \int \nabla_{x_k}\mu(X)\, \mathrm{d}\hat{X}_k, \tag{4.5}$$

where $\nabla_{x_k}\mu$ is defined according to Proposition 4.1.1. It is easy to prove, approximating μ with smooth functions, that $\nabla\mu^k$ is the distributional gradient of μ^k, hence $\mu^k \in W^{1,1}(\mathbb{R}^d)$.

Remark 10. Notice that μ^k coincides with the (density of the) push-forward measure under the projection $\pi^k\colon (\mathbb{R}^d)^N \to \mathbb{R}^d$ on the k-th factor, which makes the notation consistent.

If $\mu \in \mathcal{P}_p(\mathbb{R}^m)$, it will be useful to deal with the Sobolev norm of $\mu^{1/p}$. However, since μ is a probability,

$$\left\|\mu^{1/p}\right\|_{W^{1,p}}^p = \int \mu(x)\,\mathrm{d}x + \int \left|\nabla\mu^{1/p}(x)\right|^p \mathrm{d}x = 1 + \int \left|\nabla\mu^{1/p}(x)\right|^p \mathrm{d}x,$$

hence all the information is contained in the second summand. Therefore we give the following

Definition 7. If $\mu \in \mathcal{P}_p$, the $W^{1,p}$-*energy* of μ is defined as

$$\mathcal{E}_p(\mu) = \int \left|\nabla\mu^{1/p}(x)\right|^p \mathrm{d}x. \tag{4.6}$$

In the special case $p = 2$, this quantity may be seen as the kinetic energy $\int |\nabla\psi|^2$ of a system described by a wave-function $\psi \in W^{1,2}(\mathbb{R}^m)$, which justifies the name. It is well-known (see for instance [34]) that the kinetic energy of a wave-function is bounded from below by (a constant times) the kinetic energy of its marginals. This is also true in our setting, as stated in the following

Lemma 4.1.4. *Let* $\mu \in \mathcal{P}_p\left((\mathbb{R}^d)^N\right)$. *Then*

$$\mathcal{E}_p(\mu) \geq \sum_{k=1}^N \mathcal{E}_p(\mu^k).$$

Moreover, if $\rho_1, \dots, \rho_N \in \mathcal{P}_p(\mathbb{R}^d)$,

$$\inf\left\{\mathcal{E}_p(\mu) \mid \mu \in \mathcal{P}_p(\mathbb{R}^m) \cap \Pi(\rho_1, \dots, \rho_N)\right\} = \sum_{k=1}^N \mathcal{E}_p(\rho_k).$$

Proof. See Appendix. □

Finally, the following proves a monotonicity property and a continuity property of the energy with respect to convolution, which will be useful later (since convolution will be one of the main tools for the proof of Theorem 4.0.1).

Lemma 4.1.5. *Let $\eta \in C^\infty(\mathbb{R}^m)$, $\eta \geq 0$ such that $\int \eta = 1$ and define $\eta^\varepsilon(x) = \frac{1}{\varepsilon^m}\eta\left(\frac{x}{\varepsilon}\right)$, for $\varepsilon > 0$. Then, for every $\mu \in \mathcal{P}_p(\mathbb{R}^m)$,*

$$\mathcal{E}_p(\mu * \eta^\varepsilon) \leq \mathcal{E}_p(\mu) \quad \text{and} \quad \lim_{\varepsilon \to 0} \mathcal{E}_p(\mu * \eta^\varepsilon) = \mathcal{E}_p(\mu).$$

Proof. See Appendix. □

4.2 Proof of Theorem 4.0.1

In this Section we deal with the proof of Theorem 4.0.1. To this end, we will define an operator

$$\begin{aligned} \Theta \colon \mathbb{R}^+ \times \mathcal{P}\left((\mathbb{R}^d)^N\right) &\longrightarrow \mathcal{P}\left((\mathbb{R}^d)^N\right) \\ (\varepsilon, \mu) \quad &\longmapsto \quad \Theta^\varepsilon[\mu] \end{aligned}$$

such that the following properties hold.

A. If $\mu^k \in \mathcal{P}_p(\mathbb{R}^d)$ for every $k = 1, \dots, N$, then

$$\Theta^\varepsilon[\mu] \in \mathcal{P}_p((\mathbb{R}^d)^N).$$

B. For every $\varepsilon > 0$, for every $k = 1, \dots, N$,

$$\Theta^\varepsilon[\mu]^k = \mu^k.$$

C. For every $\varepsilon > 0$,

$$W_2(\Theta^\varepsilon[\mu], \mu) \leq C(d)\sqrt{N\varepsilon};$$

if moreover $\mu \in \mathcal{P}_p((\mathbb{R}^d)^N)$, then

$$\lim_{\varepsilon \to 0} d_p(\Theta^\varepsilon[\mu], \mu) = 0.$$

Construction of Θ and proof of property B

Fix a radial function $\eta \in C^\infty(\mathbb{R}^d)$, $\int \eta = 1$, $\eta \geq 0$. We require that there exists $c(\eta) \in (0, +\infty)$ such that

$$\int \frac{|\nabla \eta(z)|^p}{\eta(z)^{p-1}}\,\mathrm{d}z \leq c(\eta) \quad \text{and} \quad \int |z|^2\,\eta(z)\,\mathrm{d}z \leq c(\eta)$$

Examples of functions with this properties are

$$\eta(z) = \begin{cases} e^{-\frac{1}{1-|z|^2}} & |z| < 1 \\ 0 & \text{otherwise} \end{cases}$$

or any Gaussian

$$\eta(z) = \frac{1}{(2\pi\sigma)^{d/2}} e^{-\frac{|z|^2}{2\sigma}}.$$

Given $\varepsilon > 0$, let $\eta^\varepsilon \colon \mathbb{R}^d \to (0, +\infty)$ be given by

$$\eta^\varepsilon(z) = \frac{1}{\varepsilon^d} \eta\left(\frac{z}{\varepsilon}\right).$$

The following property, immediate to obtain with a change of variables $\varepsilon u = z$, will be useful in the following:

$$\int \frac{|\nabla \eta^\varepsilon(z)|^p}{\eta^\varepsilon(z)^{p-1}} \, \mathrm{d}z = \frac{1}{\varepsilon^2} \int \frac{|\nabla \eta(u)|^p}{\eta(u)^{p-1}} \, \mathrm{d}u = \frac{c(\eta)}{\varepsilon^2}$$

For $\mu \in \mathcal{P}\left((\mathbb{R}^d)^N\right)$, we define the measure $\Lambda^\varepsilon[\mu]$ as the convolution of μ with the kernel $\eta^\varepsilon(x_1) \cdots \eta^\varepsilon(x_N)$, *i.e.*, if $\psi \colon (\mathbb{R}^d)^N \to \mathbb{R}$ is any continuous bounded function,

$$\int \psi(Y) \, \mathrm{d}\Lambda^\varepsilon[\mu](Y) := \iint \psi(Y) \prod_{j=1}^N \eta^\varepsilon(y_k - x_k) \, \mathrm{d}\mu(X) \, \mathrm{d}Y. \tag{4.7}$$

Notice that $\Lambda^\varepsilon[\mu]$ is absolutely continuous with respect to the Lebesgue measure, with density

$$\Lambda^\varepsilon[\mu](Y) = \int \prod_{j=1}^N \eta^\varepsilon(y_k - x_k) \, \mathrm{d}\mu(X).$$

Finally, if $\psi \colon (\mathbb{R}^d)^N \to \mathbb{R}$ is any continuous bounded function, we define $\Theta^\varepsilon[\mu]$ via the expression

$$\int \psi(X) \Theta^\varepsilon[\mu](X) := \iint \psi(X) \prod_{j=1}^N \frac{\eta^\varepsilon(y_k - x_k)}{(\mu^k * \eta^\varepsilon)(y_k)} \, \mathrm{d}\mu^k(x_k) \Lambda^\varepsilon[\mu](Y) \, \mathrm{d}Y, \tag{4.8}$$

where the denominator $(\mu^k * \eta^\varepsilon)(y_k)$ denotes the density of the measure $\mu^k * \eta^\varepsilon$ evaluated at y_k.

Remark 11. If $(\mu^k * \eta^\varepsilon)(y_k) = 0$, we have

$$0 = \int \eta^\varepsilon(y_k - x_k) \, \mathrm{d}\mu^k(x_k) = \int \Lambda^\varepsilon(Y) \, \mathrm{d}\hat{Y}_k,$$

so the numerator also vanishes everywhere. It is safe to define the integrand to be zero for such Y's. This convention will be assumed in the following.

Remark 12. This construction fits into the general framework for the composition of transport plans, as in [1, Section 5.3]. Indeed, the definition of $\Theta^\varepsilon[\mu]$ may be seen as follows: as a first step we regularize μ by convolution; secondly, we consider the 2-transport plans β^k for $k = 1, \ldots, N$ defined by

$$\int \phi(x,y)\,\mathrm{d}\beta^k(x,y) = \int \phi(x,y)\eta^\varepsilon(x-y)\,\mathrm{d}\mu^k(y)\,\mathrm{d}y$$

for any $\phi \in C_b(\mathbb{R}^d \times \mathbb{R}^d)$. Notice that β^k has marginals $\mu^k * \eta^\varepsilon$ and μ^k. Then $\Theta^\varepsilon[\mu]$ corresponds to the composition of $\Lambda^\varepsilon[\mu]$ with β^k on each corresponding k-th marginal.

Remark 13 (Property B). For every $\varepsilon > 0$ and for every $k = 1, \ldots, N$, we have $\Lambda^\varepsilon[\mu]^k = \mu^k * \eta^\varepsilon$. Indeed, using Fubini's Theorem,

$$\begin{aligned}
\Lambda^\varepsilon[\mu]^k(y_k) &= \iint \prod_{j=1}^N \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu(X)\,\mathrm{d}\hat{Y}_k \\
&= \int \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu(X) = \int \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu^k(x_k).
\end{aligned}$$

Moreover, we have that $\Theta^\varepsilon[\mu]^k = \mu^k$, which proves property B. Again by Fubini's theorem, and using the previous result,

$$\begin{aligned}
\int \phi(x_k)\,\mathrm{d}\Theta^\varepsilon[\mu](X) &= \iint \phi(x_k) \prod_{j=1}^N \frac{\eta^\varepsilon(y_k - x_k)}{(\mu^k * \eta^\varepsilon)(y_k)}\,\mathrm{d}\mu^k(x_k)\Lambda^\varepsilon[\mu](Y)\,\mathrm{d}Y \\
&= \int \phi(x_k)\frac{\eta^\varepsilon(y_k - x_k)}{(\mu^k * \eta^\varepsilon)(y_k)}\,\mathrm{d}\mu^k(x_k)\Lambda^\varepsilon[\mu](Y)\,\mathrm{d}Y \\
&= \int \phi(x_k)\eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu^k(x_k)\,\mathrm{d}y_k = \int \phi(x_k)\,\mathrm{d}\mu^k(x_k).
\end{aligned}$$

Proof of property A and energy estimates

In this Section we prove that Θ satisfies property A. Moreover, we will give upper bounds for the $W^{1,p}$-energy of $\Theta^\varepsilon[\mu]$. Let $\mu \in \mathcal{P}\left((\mathbb{R}^d)^N\right)$ such that $\mu^k \in \mathcal{P}_p(\mathbb{R}^d)$ for every $k = 1, \ldots, N$. Then $\Theta^\varepsilon[\mu]$ is absolutely continuous with respect to the Lebesgue measure, with density given by

$$\Theta^\varepsilon[\mu](X) = \int P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y,$$

where we denote by $P^\varepsilon[\mu]$ the integral kernel appearing in (4.8), namely

$$P^\varepsilon[\mu](X,Y) := \prod_{k=1}^N \frac{\eta^\varepsilon(y_k - x_k)}{(\mu^k * \eta^\varepsilon)(y_k)}\mu^k(x_k)\Lambda^\varepsilon[\mu](Y). \tag{4.9}$$

Let us denote by

$$\nabla_{x_k}\Theta^\varepsilon[\mu](X) := \frac{\nabla\mu^k(x_k)}{\mu^k(x_k)}\Theta^\varepsilon[\mu](X) - \int \frac{\nabla\eta^\varepsilon(y_k - x_k)}{\eta^\varepsilon(y_k - x_k)} P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y. \tag{4.10}$$

We claim that $\nabla_{x_k}\Theta^\varepsilon[\mu](X)$ is the weak gradient with respect to the k-th variable of $\Theta^\varepsilon[\mu](X)$ in $W^{1,1}((\mathbb{R}^d)^N)$. Indeed, if $\psi \in C_c^\infty((\mathbb{R}^d)^N)$, by the Fubini's Theorem we may perform first the integration in x_k to get

$$\begin{aligned}
-\int \nabla_{x_k}\psi(X)\Theta^\varepsilon[\mu](X)\,\mathrm{d}X &= -\iint \nabla_{x_k}\psi(X)P^\varepsilon[\mu](X,Y)\,\mathrm{d}X\,\mathrm{d}Y \\
&= \iint \psi(X)\frac{\nabla\mu^k(x_k)}{\mu^k(x_k)}P^\varepsilon[\mu](X,Y)\,\mathrm{d}X\,\mathrm{d}Y \\
&\quad - \iint \psi(X)\frac{\nabla\eta^\varepsilon(y_k - x_k)}{\eta^\varepsilon(y_k - x_k)}P^\varepsilon[\mu](X,Y)\,\mathrm{d}X\,\mathrm{d}Y \\
&= \int \psi(X)\frac{\nabla\mu^k(x_k)}{\mu^k(x_k)}\Theta^\varepsilon[\mu](X)\,\mathrm{d}X \\
&\quad - \int \psi(X)\int \frac{\nabla\eta^\varepsilon(y_k - x_k)}{\eta^\varepsilon(y_k - x_k)}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\,\mathrm{d}X.
\end{aligned}$$

To conclude that $\Theta^\varepsilon[\mu] \in \mathcal{P}_p\left((\mathbb{R}^d)^N\right)$, in view of Proposition 4.1.1, it suffices to show a suitable domination, which is given in the following

Lemma 4.2.1. *Let $\mu \in \mathcal{P}((\mathbb{R}^d)^N)$ such that $\mu^k \in \mathcal{P}_p(\mathbb{R}^d)$ for every $k = 1, \dots, N$. Then*

$$\begin{aligned}
&|\nabla_{x_k}\Theta^\varepsilon[\mu](X)|^p\,\Theta^\varepsilon[\mu](X)^{1-p} \\
&\leq 2^{p-1}\left(\frac{|\nabla\mu^k(x_k)|^p}{\mu^k(x_k)^p}\Theta^\varepsilon[\mu](X) + \int \frac{|\nabla\eta^\varepsilon(y_k - x_k)|^p}{\eta^\varepsilon(y_k - x_k)^p}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\right)
\end{aligned}$$

Proof. The triangular inequality for the p-th norm on $\mathbb{R}^d$ gives

$$|\nabla_{x_k}\Theta^\varepsilon[\mu](X)| \leq \frac{|\nabla\mu^k(x_k)|}{\mu^k(x_k)}\Theta^\varepsilon[\mu](X) + \int \frac{|\nabla\eta^\varepsilon(y_k - x_k)|}{\eta^\varepsilon(y_k - x_k)}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y.$$

Using the Hölder inequality with exponents p and $\frac{p}{p-1}$,

$$\begin{aligned}
&\int \frac{|\nabla\eta^\varepsilon(y_k - x_k)|}{\eta^\varepsilon(y_k - x_k)}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y \\
&\leq \left(\int \frac{|\nabla\eta^\varepsilon(y_k - x_k)|^p}{\eta^\varepsilon(y_k - x_k)^p}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\right)^{\frac{1}{p}}\Theta^\varepsilon[\mu](X)^{\frac{p-1}{p}},
\end{aligned}$$

and the thesis follows. □

Finally we get the proof of property A, together with the usual explicit formula for the weak gradient of $\Theta^\varepsilon[\mu]^{1/p}$.

Theorem 4.2.2 (Property A). *Let $\mu \in \mathcal{P}\left((\mathbb{R}^d)^N\right)$ such that $\mu^k \in \mathcal{P}_p(\mathbb{R}^d)$ for every $k = 1, \dots, N$. Then $\Theta^\varepsilon[\mu] \in \mathcal{P}_p\left((\mathbb{R}^d)^N\right)$, and*

$$\nabla_{x_k}\Theta^\varepsilon[\mu]^{1/p}(X) = \frac{1}{p}\Theta^\varepsilon[\mu](X)^{\frac{1-p}{p}}\nabla_{x_k}\Theta^\varepsilon[\mu](X).$$

Proof. Recalling Proposition 4.1.1, it suffices to check that condition (4.2) holds. Using Lemma 4.2.1 we have

$$\begin{aligned}
&\int |\nabla_{x_k}\Theta^\varepsilon[\mu](X)|^p\,\Theta^\varepsilon[\mu](X)^{1-p}\,\mathrm{d}X \\
&\le 2^{p-1}\left(\int \frac{|\nabla\mu^k(x_k)|^p}{\mu^k(x_k)^p}\Theta^\varepsilon[\mu](X)\,\mathrm{d}X + \iint \frac{|\nabla\eta^\varepsilon(y_k-x_k)|^p}{\eta^\varepsilon(y_k-x_k)^p}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\,\mathrm{d}X\right) \\
&= 2^{p-1}\left(\int \frac{|\nabla\mu^k(x_k)|^p}{\mu^k(x_k)^{p-1}}\,\mathrm{d}x_k + \int \frac{|\nabla\eta^\varepsilon(z)|^p}{\eta^\varepsilon(z)^{p-1}}\,\mathrm{d}z\right) \\
&\le 2^{p-1}p^p\left\|\nabla(\mu^k)^{1/p}\right\|_p^p + \frac{c(\eta)}{\varepsilon^2},
\end{aligned}$$

where $c(d,p)$ is a constant depending on the dimension d and the exponent p. □

From Theorem 4.2.2 we get also estimates on the $W^{1,p}$-energy of $\Theta^\varepsilon[\mu]$, as stated in the following

Theorem 4.2.3. *Let $\mu \in \mathcal{P}\left((\mathbb{R}^d)^N\right)$. Then there exists a constant $c(\eta,p)$ such that*

$$\mathcal{E}_p(\Theta^\varepsilon[\mu]) \le \sum_{k=1}^N\left(\mathcal{E}_p(\mu^k)^{1/p} + \frac{c(\eta,p)}{\varepsilon^{2/p}}\right)^p. \tag{4.11}$$

If in addition $\mu \in \mathcal{P}_p\left((\mathbb{R}^d)^N\right)$ and $p > 1$, then

$$\mathcal{E}_p(\Theta^\varepsilon[\mu]) \le \sum_{k=1}^N\left(\left\|\nabla_{x_k}\Lambda^\varepsilon[\mu]^{1/p}\right\|_p + c(p)\Delta^k(\varepsilon,p,\mu)\right)^p \tag{4.12}$$

where

$$\Delta^k(\varepsilon,p,\mu) = \begin{cases}\left[\left(\mathcal{E}_p(\mu^k)+\mathcal{E}_p(\mu^k*\eta^\varepsilon)\right)^{\frac{1}{p-1}} - 2^{\frac{1}{p-1}}\mathcal{E}_p(\mu^k*\eta^\varepsilon)^{\frac{1}{p-1}}\right]^{\frac{p-1}{p}} & 1<p<2 \\ \left(\mathcal{E}_p(\mu^k)-\mathcal{E}_p(\mu^k*\eta^\varepsilon)\right)^{\frac{1}{p}} & p\ge 2\end{cases}$$

and $c(p)$ is an explicit constant depending only on the exponent p.

Proof. Combining Theorem 4.2.2 and (4.10) we get

$$\begin{aligned}
&\int\left|\nabla_{x_k}\Theta^\varepsilon[\mu]^{1/p}(X)\right|^p\,\mathrm{d}X = \frac{1}{p^p}\int\Theta^\varepsilon[\mu](X)^{1-p}\,|\nabla_{x_k}\Theta^\varepsilon[\mu](X)|^p\,\mathrm{d}X \\
&= \frac{1}{p^p}\int\left|\frac{\nabla\mu^k(x_k)}{\mu^k(x_k)}\Theta^\varepsilon[\mu](X)^{1/p} + \Theta^\varepsilon[\mu](X)^{\frac{1-p}{p}}\int\frac{\nabla\eta^\varepsilon(y_k-x_k)}{\eta^\varepsilon(y_k-x_k)}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\right|^p\,\mathrm{d}X \\
&=: \frac{1}{p^p}\int|f(X)+g(X)|^p\,\mathrm{d}X.
\end{aligned}$$

Then we have

$$\int |f(X)|^p \,\mathrm{d}X = \int \frac{\left|\nabla\mu^k(x_k)\right|^p}{\mu^k(x_k)^p}\Theta^\varepsilon[\mu](X)\,\mathrm{d}X = \int \frac{\left|\nabla\mu^k(x_k)\right|^p}{\mu^k(x_k)^{p-1}}\,\mathrm{d}x_k = p^p\mathcal{E}_p(\mu^k).$$

By the Hölder inequality (as in the proof of Lemma 4.2.1), we have

$$\begin{aligned}|g(X)|^p &= \Theta^\varepsilon[\mu](X)^{1-p}\left|\int \frac{\nabla\eta^\varepsilon(y_k-x_k)}{\eta^\varepsilon(y_k-x_k)}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\right|^p \\ &\le \int \frac{|\nabla\eta^\varepsilon(y_k-x_k)|^p}{\eta^\varepsilon(y_k-x_k)^p}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y,\end{aligned}$$

thus, as in the proof of Theorem 4.2.2,

$$\int |g(X)|^p \,\mathrm{d}X \le \iint \frac{|\nabla\eta^\varepsilon(y_k-x_k)|^p}{\eta^\varepsilon(y_k-x_k)^p}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\,\mathrm{d}X \le \frac{c(\eta)}{\varepsilon^2}.$$

By the triangular inequality in $L^p((\mathbb{R}^d)^N)$ we conclude that

$$\mathcal{E}_p(\Theta^\varepsilon[\mu]) = \sum_{k=1}^N \int \left|\nabla_{x_k}\Theta^\varepsilon[\mu]^{1/p}(X)\right|^p \,\mathrm{d}X \le \sum_{k=1}^N \left(\mathcal{E}_p(\mu^k)^{1/p} + \frac{c(\eta,p)}{\varepsilon^{2/p}}\right)^p,$$

as wanted.

When the measure μ is regular, we may perform a change of variable in (4.10) to get

$$\begin{aligned}\nabla_{x_k}\Theta^\varepsilon[\mu](X) &= \int \left(\frac{\nabla\mu^k(x_k)}{\mu^k(x_k)} - \frac{\nabla(\mu^k*\eta^\varepsilon)(y_k)}{(\mu^k*\eta^\varepsilon)(y_k)}\right)P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y \\ &\quad + \int \nabla_{x_k}\Lambda^\varepsilon[\mu](Y)\prod_{j=1}^N \frac{\eta^\varepsilon(y_j-x_j)}{(\mu^j*\eta^\varepsilon)(y_j)}\mu^j(x_j)\,\mathrm{d}Y \\ &=: I(X) + II(X).\end{aligned}$$

By the Hölder inequality with exponents p and $\frac{p}{p-1}$,

$$\begin{aligned}|I(X)|^p &\le \Theta^\varepsilon[\mu](X)^{p-1}\int \left|\frac{\nabla\mu^k(x_k)}{\mu^k(x_k)} - \frac{\nabla(\mu^k*\eta^\varepsilon)(y_k)}{(\mu^k*\eta^\varepsilon)(y_k)}\right|^p P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y, \\ |II(X)|^p &\le \Theta^\varepsilon[\mu](X)^{p-1}\int \frac{|\nabla_{x_k}\Lambda^\varepsilon[\mu](Y)|^p}{\Lambda^\varepsilon[\mu](Y)^p}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y.\end{aligned}$$

When we integrate with respect to the X variable, the triangular inequality in L^p gives

$$
\begin{aligned}
&\left(\int \frac{|\nabla_{x_k}\Theta^\varepsilon[\mu](X)|^p}{\Theta^\varepsilon[\mu](X)^{p-1}}\,\mathrm{d}X\right)^{1/p} \\
&\le \left(\iint \left|\frac{\nabla\mu^k(x_k)}{\mu^k(x_k)} - \frac{\nabla(\mu^k*\eta^\varepsilon)(y_k)}{(\mu^k*\eta^\varepsilon)(y_k)}\right|^p P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\,\mathrm{d}X\right)^{1/p} \\
&\quad + \left(\iint \frac{|\nabla_{x_k}\Lambda^\varepsilon[\mu](Y)|^p}{\Lambda^\varepsilon[\mu](Y)^p} P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\,\mathrm{d}X\right)^{1/p} \\
&= \left(\iint \left|\frac{\nabla\mu^k(x)}{\mu^k(x)} - \frac{\nabla(\mu^k*\eta^\varepsilon)(y)}{(\mu^k*\eta^\varepsilon)(y)}\right|^p \eta^\varepsilon(y-x)\mu^k(x)\,\mathrm{d}x\,\mathrm{d}y\right)^{1/p} \\
&\quad + \left(\int \frac{|\nabla_{x_k}\Lambda^\varepsilon[\mu](Y)|^p}{\Lambda^\varepsilon[\mu](Y)^{p-1}}\,\mathrm{d}Y\right)^{1/p} \\
&=: D^{1/p} + p\left\|\nabla_{x_k}\Lambda^\varepsilon[\mu]^{1/p}\right\|_p
\end{aligned}
$$

Now we recall the following inequalities by Clarkson [12]: if $f, g \in L^p(\nu)$, then

$$
\left\|\frac{f-g}{2}\right\|^p \le \frac{1}{2}\|f\|^p + \frac{1}{2}\|g\|^p - \left\|\frac{f+g}{2}\right\|^p \qquad p \ge 2 \tag{4.13}
$$

$$
\left\|\frac{f-g}{2}\right\|^{\frac{p}{p-1}} \le \left(\frac{1}{2}\|f\|^p + \frac{1}{2}\|g\|^p\right)^{\frac{1}{p-1}} - \left\|\frac{f+g}{2}\right\|^{\frac{p}{p-1}} \qquad 1 < p < 2, \tag{4.14}
$$

where all the norms are $L^p(\nu)$ norms.

If we apply (4.13) on $\mathbb{R}^d \times \mathbb{R}^d$ with $f(x,y) = \frac{\nabla\mu^k(x)}{\mu^k(x)}$, $g(x,y) = \frac{\nabla(\mu^k*\eta^\varepsilon)(y)}{(\mu^k*\eta^\varepsilon)(y)}$ and $\frac{\mathrm{d}\nu}{\mathrm{d}\mathcal{L}^d}(x,y) = \eta^\varepsilon(y-x)\mu^k(x)$, we get for $p \ge 2$

$$
\begin{aligned}
D \le\ & 2^{p-1}\iint \left|\frac{\nabla\mu^k(x)}{\mu^k(x)}\right|^p \eta^\varepsilon(y-x)\mu^k(x)\,\mathrm{d}x\,\mathrm{d}y \\
&+ 2^{p-1}\iint \left|\frac{\nabla(\mu^k*\eta^\varepsilon)(y)}{(\mu^k*\eta^\varepsilon)(y)}\right|^p \eta^\varepsilon(y-x)\mu^k(x)\,\mathrm{d}x\,\mathrm{d}y \\
&- \iint \left|\frac{\nabla\mu^k(x)}{\mu^k(x)} + \frac{\nabla(\mu^k*\eta^\varepsilon)(y)}{(\mu^k*\eta^\varepsilon)(y)}\right|^p \eta^\varepsilon(y-x)\mu^k(x)\,\mathrm{d}x\,\mathrm{d}y \\
=:\ & 2^{p-1}p^p\left(\mathcal{E}_p(\mu^k) + \mathcal{E}_p(\mu^k*\eta^\varepsilon)\right) - S.
\end{aligned}
$$

Similarly, if $1 < p < 2$, using (4.14) we have

$$
D \le \left[2p^{\frac{p}{p-1}}\left(\mathcal{E}_p(\mu^k) + \mathcal{E}_p(\mu^k*\eta^\varepsilon)\right)^{\frac{1}{p-1}} - S^{\frac{1}{p-1}}\right]^{p-1}.
$$

Finally, we consider the convex function $\varphi(z) = |z|^p$ on $\mathbb{R}^d$, for which we have $\nabla\varphi(z) = p(|z_1|^{p-2} z_1, \ldots, |z_d|^{p-2} z_d)$. By convexity we have

$$
\varphi(w+z) \ge \varphi(2z) + \nabla\varphi(2z)\cdot(w-z),
$$

so letting $w = \frac{\nabla\mu^k(x)}{\mu^k(x)}$ and $z = \frac{\nabla(\mu^k * \eta^\varepsilon)(y)}{(\mu^k*\eta^\varepsilon)(y)}$ we get

$$\begin{aligned}
S \geq &\iint \left|\frac{2\nabla(\mu^k * \eta^\varepsilon)(y)}{(\mu^k * \eta^\varepsilon)(y)}\right|^p \eta^\varepsilon(y-x)\mu^k(x)\,\mathrm{d}x\,\mathrm{d}y \\
&+ 2^{p-1}p\sum_{j=1}^d \iint \left|\frac{\partial_j(\mu^k * \eta^\varepsilon)(y)}{(\mu^k * \eta^\varepsilon)(y)}\right|^{p-2} \frac{\partial_j(\mu^k * \eta^\varepsilon)(y)}{(\mu^k * \eta^\varepsilon)(y)}\frac{\partial_j\mu^k(x)}{\mu^k(x)}\eta^\varepsilon(y-x)\mu^k(x)\,\mathrm{d}x\,\mathrm{d}y \\
&- 2^{p-1}p\sum_{j=1}^d \iint \left|\frac{\partial_j(\mu^k * \eta^\varepsilon)(y)}{(\mu^k * \eta^\varepsilon)(y)}\right|^{p-2} \frac{\left|\partial_j(\mu^k * \eta^\varepsilon)(y)\right|^2}{(\mu^k * \eta^\varepsilon)(y)^2}\eta^\varepsilon(y-x)\mu^k(x)\,\mathrm{d}x\,\mathrm{d}y \\
= 2^p &\int \frac{\left|\nabla(\mu^k * \eta^\varepsilon)(y)\right|^p}{(\mu^k * \eta^\varepsilon)(y)^{p-1}}\,\mathrm{d}y = 2^p p^p \mathcal{E}_p(\mu^k * \eta^\varepsilon).
\end{aligned}$$

Hence, for $p \geq 2$,

$$D \leq 2^{p-1}p^p \left(\mathcal{E}_p(\mu^k) - \mathcal{E}_p(\mu^k * \eta^\varepsilon)\right),$$

while for $1 < p < 2$

$$D \leq 2^{p-1}p^p \left[\left(\mathcal{E}_p(\mu^k) + \mathcal{E}_p(\mu^k * \eta^\varepsilon)\right)^{\frac{1}{p-1}} - 2^{\frac{1}{p-1}}\mathcal{E}_p(\mu^k * \eta^\varepsilon)^{\frac{1}{p-1}}\right]^{p-1},$$

Putting all together and summing on k we get the thesis. □

In the particular case $p = 2$, the Hilbertian structure allows to simplify some computations and to get slightly sharper constants, as stated in the following result, which is proved in the Appendix.

Theorem 4.2.4. *Let $\mu \in \mathcal{P}\left((\mathbb{R}^d)^N\right)$ such that $\mu^k \in \mathcal{P}_2(\mathbb{R}^d)$ for every $k = 1, \ldots, N$. Then*

$$\mathcal{E}_2(\Theta^\varepsilon[\mu]) \leq \frac{Nc(\eta)}{\varepsilon^2} + \sum_{k=1}^N \mathcal{E}_2(\mu^k), \tag{4.15}$$

where $c(\eta)$ is a constant depending on the choice of η.

Remark 14. As one would expect, if the measure μ is not regular then the bound on the energy of $\Theta^\varepsilon[\mu]$ diverges as ε approaches zero, as in (4.11) and (4.15). On the contrary, if μ is $W^{1,p}$-regular then the bound on the energy of $\Theta^\varepsilon[\mu]$ in (4.12) converges to the energy of μ as $\varepsilon \to 0$. Indeed, on the one hand $\Delta^k(\varepsilon, p, \mu)$ converges to zero for every $k = 1, \ldots, N$ by Lemma 4.1.5. On the other hand, let $\lambda^\varepsilon(z_1, \ldots, z_N) = \eta^\varepsilon(z_1)\cdots\eta^\varepsilon(z_N)$, we have $\Lambda^\varepsilon[\mu] = \mu * \lambda^\varepsilon$, and hence

$$\left\|\nabla_{x_k}(\mu * \lambda^\varepsilon)^{1/p}\right\|_p \to \left\|\nabla_{x_k}\mu^{1/p}\right\|_p.$$

When we raise to the power p and sum over k we get $\mathcal{E}_p(\mu)$ in view of the usual condition (4.1).

Proof of property C

For the proof of property C we do not need to assume that the marginals of μ are regular. In order to simplify the notation, let as above $P^\varepsilon[\mu]$ be the measure over $(\mathbb{R}^d)^N \times (\mathbb{R}^d)^N$ given by

$$\iint \psi(X,Y)\,\mathrm{d}P^\varepsilon[\mu](X,Y) := \iint \psi(X,Y) \prod_{k=1}^N \frac{\eta^\varepsilon(y_k - x_k)}{(\mu^k * \eta^\varepsilon)(y_k)}\,\mathrm{d}\mu^k(x_k)\,\mathrm{d}\Lambda^\varepsilon[\mu](Y),$$

already introduced above, and let $Q^\varepsilon[\mu]$ be the measure over $(\mathbb{R}^d)^N \times (\mathbb{R}^d)^N$ given by

$$\iint \psi(X,Y)\,\mathrm{d}Q^\varepsilon[\mu](X,Y) := \iint \psi(X,Y) \prod_{k=1}^N \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu(X)\,\mathrm{d}Y$$

for any $\psi\colon (\mathbb{R}^d)^N \times (\mathbb{R}^d)^N \to \mathbb{R}$ bounded and countinuous.

Remark 15. Observe that $P^\varepsilon[\mu](X,Y)$ is the density of a transport plan between $\Theta^\varepsilon[\mu]$ and the measure $\Lambda^\varepsilon[\mu]$ introduced in (4.7). Analogously, $Q^\varepsilon[\mu]$ is a transport plan between $\Lambda^\varepsilon[\mu]$ and μ.

Theorem 4.2.5 (Property C, first part)**.** *For every $\varepsilon > 0$,*

$$W_2(\Theta^\varepsilon[\mu],\mu) \le \varepsilon C(\eta)\sqrt{N},$$

where $C(\eta)$ is a constant depending on the choice of η.

Proof. The idea is to estimate $W_2(\Theta^\varepsilon[\mu],\Lambda^\varepsilon[\mu])$ and $W_2(\Lambda^\varepsilon[\mu],\mu)$, and then use the triangular inequality for the Wasserstein distance. On the one hand,

$$\begin{aligned}
W_2^2(\Lambda^\varepsilon[\mu],\mu) &\le \iint |X-Y|^2\,\mathrm{d}Q^\varepsilon[\mu](X,Y) = \sum_{k=1}^N \iint |x_k - y_k|^2\,\mathrm{d}Q^\varepsilon[\mu](X,Y)\\
&= \sum_{k=1}^N \iint |x_k - y_k|^2\,\eta_\varepsilon(y_k - x_k)\,\mathrm{d}y_k\,\mathrm{d}\mu^k(x_k)\\
&= \sum_{k=1}^N \iint |z_k|^2\,\eta_\varepsilon(z_k)\,\mathrm{d}z_k\,\mathrm{d}\mu^k(x_k)\\
&\le \varepsilon^2 \sum_{k=1}^N \int_{\mathbb{R}^d} |z|^2\,\eta(z)\,\mathrm{d}z = N\varepsilon^2 c(\eta).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
W_2^2(\Theta^\varepsilon[\mu],\Lambda^\varepsilon[\mu]) &\le \iint |X-Y|^2\,\mathrm{d}P^\varepsilon[\mu](X,Y)\\
&= \sum_{k=1}^N \iint |x_k - y_k|^2\,\mathrm{d}P^\varepsilon[\mu](X,Y)\\
&= \sum_{k=1}^N \iint |x_k - y_k|^2\,\eta_\varepsilon(y_k - x_k)\,\mathrm{d}y_k\,\mathrm{d}\mu^k(x_k) \le N\varepsilon^2 c(\eta),
\end{aligned}$$

as before. In conclusion,

$$W_2(\Theta^\varepsilon[\mu],\mu) \le 2\varepsilon\sqrt{Nc(\eta)}. \qquad \square$$

Corollary 4.2.6. *As $\varepsilon \to 0$, $\Theta^\varepsilon[\mu]$ converges weakly to μ in duality with $C_b((\mathbb{R}^d)^N)$.*

Proof. As it is well-known, the Wasserstein convergence implies the weak convergence — see e.g. [1, Proposition 7.1.5]. $\square$

We conclude this section with the second part of property C. So far we proved that $\Theta^\varepsilon[\mu]$ converges to μ in the Wasserstein and weak sense, which is the natural notion of convergence as far as μ is no more regular than a measure. However if μ is regular, since $\Theta^\varepsilon[\mu] \in \mathcal{P}_p$ for every $\varepsilon > 0$ it is natural to ask whether $\Theta^\varepsilon[\mu] \to \mu$ in the d_p-topology. The answer is positive, as stated in the following

Theorem 4.2.7 (Property C, second part)**.** *Let $\mu \in \mathcal{P}_p\left((\mathbb{R}^d)^N\right)$. Then*

$$\lim_{\varepsilon\to 0} d_p(\Theta^\varepsilon[\mu],\mu) = 0.$$

Proof. Combining the fact that the family $\Theta^\varepsilon[\mu]^{1/p}$ is bounded in $W^{1,p}$ due to Theorem 4.2.3 and the result of Corollary 4.2.6 we get that $\Theta^\varepsilon[\mu]^{1/p} \to \mu^{1/p}$ weakly in $W^{1,p}\left((\mathbb{R}^d)^N\right)$ as $\varepsilon \to 0$. Since $W^{1,p}$ is uniformly convex, we need only to check that

$$\lim_{\varepsilon\to 0} \left\|\Theta^\varepsilon[\mu]^{\frac{1}{p}}\right\|_{W^{1,p}} = \left\|\mu^{\frac{1}{p}}\right\|_{W^{1,p}}.$$

The L^p-norms are identically equal to 1, so we need to prove the limit for the norms of the gradients. The weak convergence of $\nabla\Theta^\varepsilon[\mu]^{\frac{1}{p}}$ to $\nabla\mu^{\frac{1}{p}}$ implies that

$$\liminf_{\varepsilon\to 0} \left\|\nabla\Theta^\varepsilon[\mu]^{\frac{1}{p}}\right\|_{L^p} \ge \left\|\nabla\mu^{\frac{1}{p}}\right\|_{L^p}.$$

The other inequality follows from Remark 14. $\square$

4.3 Proof of Theorems 4.3.1 and 4.3.2

We recall and prove Theorem 4.3.1.

Theorem 4.3.1. *Let η be supported on $B(0,1)$, and $\Omega \subseteq (\mathbb{R}^d)^N$ such that $\Omega + B(0,r) \subset (\mathbb{R}^d)^N \setminus \operatorname{supp}\mu$ for some $r > 0$. Then $\Theta^\varepsilon(\mu) = 0$ on Ω as soon as $\varepsilon < r/2$.*

Proof. Let $X \in \Omega$ and $\varepsilon < r/2$. We have

$$\begin{aligned}\Theta^\varepsilon[\mu](X) &= \int P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y \\ &= \int_{|X-Y|<\frac{r}{2}} P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y + \int_{|X-Y|\ge\frac{r}{2}} P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y,\end{aligned}$$

where we recall that

$$P^\varepsilon[\mu](X,Y) = \prod_{k=1}^{N} \frac{\eta^\varepsilon(y_k - x_k)\mu^k(x_k)}{(\mu^k * \eta^\varepsilon)(y_k)} \Lambda^\varepsilon[\mu](Y).$$

We claim that both integrals are equal to zero. If $|X - Y| < \frac{r}{2}$, then $Y + B(0, r/2) \subset (\mathbb{R}^d)^N \setminus \operatorname{supp}\mu$. Hence, $\Lambda^\varepsilon[\mu] = 0$ for $\varepsilon < \frac{r}{2\sqrt{N}}$, since $\Lambda^\varepsilon[\mu] = \mu * \lambda^\varepsilon$ and $\operatorname{supp}\lambda^\varepsilon \subseteq B(0, \varepsilon\sqrt{N})$.

On the other hand, if $|X - Y| \geq \frac{r}{2}$, then there exists k such that $|x_k - y_k| \geq \frac{r}{2\sqrt{N}}$. Once again, if $\varepsilon < \frac{r}{2\sqrt{N}}$, then $\eta^\varepsilon(y_k - x_k) = 0$, and so $P^\varepsilon[\mu](X,Y) = 0$. □

Finally we prove Theorem 4.3.2.

Theorem 4.3.2. *Let η be a Gaussian kernel, and suppose that $\mu_n \rightharpoonup \mu$, and $\pi_{\#}\mu_n \to \pi_{\#}\mu$ in the d_p-topology. Then $\Theta^\varepsilon(\mu_n) \to \Theta^\varepsilon(\mu)$ in the d_p-topology for every $\varepsilon > 0$.*

We fix $\varepsilon > 0$ and we take a Gaussian mollifier

$$\eta^\varepsilon = \frac{1}{(2\pi\varepsilon)^{d/2}} e^{-\frac{|z|^2}{2\varepsilon}}.$$

The main idea is to use the Dominated Convergence Theorem, but in order to do so we must first prove some fine upper-bound on the integral kernel $P^\varepsilon[\mu]$. With a slight abuse of notation, since $\Lambda^\varepsilon[\mu]$ and $\mu^k * \eta^\varepsilon$ are absolutely continuous with respect to the Lebesgue measure, we will use the same symbol for the measure and its density.

Lemma 4.3.3. *Let $\mu \in \mathcal{P}\left((\mathbb{R}^d)^N\right)$. Then:*

(i)

$$\Lambda^\varepsilon[\mu](Y) \leq (2\pi\varepsilon)^{-\frac{(N-1)d}{2N}} \prod_{k=1}^{N} (\mu^k * \eta^\varepsilon)(y_k)^{\frac{1}{N}}.$$

(ii) Let $R > 0$, $\gamma \in [0,1]$ be such that $\mu^k(B(0,R)) \geq \gamma$. Then

$$(\mu^k * \eta^\varepsilon)(y_k) \geq \frac{\gamma}{(2\pi\varepsilon)^{d/2}} \exp\left(-\frac{(|y_k| + R)^2}{2\varepsilon}\right).$$

Proof. (i) We apply a general version of the Hölder's inequality with exponents $p_1 = \cdots = p_N = N$, and use the fact that $\eta^\varepsilon(z) \le \eta^\varepsilon(0) = (2\pi\varepsilon)^{-d/2}$, to get

$$\begin{aligned}\Lambda^\varepsilon[\mu](Y) &= \int \prod_{k=1}^N \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu(X) \le \prod_{k=1}^N \left(\int \eta^\varepsilon(y_k - x_k)^N\,\mathrm{d}\mu(X)\right)^{\frac{1}{N}} \\ &\le (2\pi\varepsilon)^{-\frac{(N-1)d}{2N}} \prod_{k=1}^N \left(\int \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu(X)\right)^{\frac{1}{N}} \\ &= (2\pi\varepsilon)^{-\frac{(N-1)d}{2N}} \prod_{k=1}^N (\mu^k * \eta^\varepsilon)(y_k)^{\frac{1}{N}}.\end{aligned}$$

as wanted.

(ii) We start by observing that

$$(\mu^k * \eta^\varepsilon)(y_k) = \int \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu^k(x_k) \ge \int_{B(0,R)} \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu^k(x_k).$$

When x_k belongs to the ball $B(0,R)$, the minimum value of $\eta^\varepsilon(y_k - x_k)$ is attained at $x_k = -R\frac{y_k}{|y_k|}$, or at any boundary point if $y_k = 0$. Thus, in this region,

$$\eta^\varepsilon(y_k - x_k) \ge \frac{1}{(2\pi\varepsilon)^{d/2}} \exp\left(-\frac{(|y_k| + R)^2}{2\varepsilon}\right)$$

and the thesis follows. $\square$

Lemma 4.3.4. *Let $\rho_n, \rho \in \mathcal{P}_p(\mathbb{R}^d)$ such that $\rho_n \to \rho$ in the d_p-topology. Then, for every $\gamma > 0$ there exists $R > 0$ such that $\rho_n(B(0,R)) \ge 1 - \gamma$ and $\rho(B(0,R)) \ge 1 - \gamma$.*

Proof. It suffices to show that $\rho_n \rightharpoonup \rho$, which implies that the family $\{\rho_n\}_{n\in\mathbb{N}}$ is tight; however, by Proposition 4.1.2 we have the stronger property $\rho_n \to \rho$ in $W^{1,1}(\mathbb{R}^d)$. $\square$

Proposition 4.3.5. *Suppose that $\mu_n \rightharpoonup \mu$, with $\mu_n^k \to \mu^k$ in $\mathcal{P}_p(\mathbb{R}^d)$ and $\mu_n^k \to \mu^k$ pointwise a.e. on $\mathbb{R}^d$ for every $k = 1, \dots, N$. Then $\Theta^\varepsilon[\mu_n] \to \Theta^\varepsilon[\mu]$ pointwise a.e. on $(\mathbb{R}^d)^N$.*

Assume in addiction that $\nabla\mu_n^k \to \nabla\mu^k$ pointwise a.e. on $\mathbb{R}^d$. Then $\nabla\Theta^\varepsilon[\mu_n] \to \nabla\Theta^\varepsilon[\mu]$ pointwise a.e. on $(\mathbb{R}^d)^N$.

Proof. Let $P^\varepsilon[\mu](X,Y)$ be the integral kernel defining $\Theta^\varepsilon[\mu]$, namely

$$P^\varepsilon[\mu](X,Y) = \prod_{k=1}^N \frac{\eta^\varepsilon(y_k - x_k)}{(\mu^k * \eta^\varepsilon)(y_k)} \mu^k(x_k)\Lambda^\varepsilon[\mu](Y).$$

We claim that $P^\varepsilon[\mu_n]$ converges pointwise a.e. to $P^\varepsilon[\mu]$. For every $Y \in (\mathbb{R}^d)^N$ and every $k \in \{1, \ldots, N\}$ we have

$$\begin{aligned}\left|(\mu_n^k * \eta^\varepsilon)(y_k) - (\mu^k * \eta^\varepsilon)(y_k)\right| &\le \int \eta^\varepsilon(y_k - x_k)\left|\mu_n^k(x_k) - \mu^k(x_k)\right| \mathrm{d}x_k \\ &\le \frac{1}{(2\pi\varepsilon)^{\frac{d}{2}}}\left\|\mu_n^k - \mu^k\right\|_1 \to 0\end{aligned}$$

by Proposition 4.1.2. Moreover

$$\begin{aligned}&|\Lambda^\varepsilon[\mu_n](Y) - \Lambda^\varepsilon[\mu](Y)| \\ &\le \left|\int \prod_{k=1}^N \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu_n(X) - \int \prod_{k=1}^N \eta^\varepsilon(y_k - x_k)\,\mathrm{d}\mu(X)\right|\end{aligned}$$

goes to zero for every Y because $\prod \eta^\varepsilon(y_k - x_k)$ is a fixed countinuous bounded function, and $\mu_n \rightharpoonup \mu$. Finally fix $X \in (\mathbb{R}^d)^N$ in the set of full measure such that $\mu_n^k(x_k) \to \mu^k(x_k)$ for every $k = 1, \ldots, N$.

We need only to find a domination for $P^\varepsilon[\mu_n]$. For every $k = 1, \ldots, N$ let R_k given by Lemma 4.3.4 for $\gamma = \frac{1}{2}$, and let $R = \max_k R_k$. Using Lemma 4.3.3 (i) and (ii) one has

$$\begin{aligned}P^\varepsilon[\mu_n](X,Y) &\le (2\pi\varepsilon)^{-\frac{(N-1)d}{2N}} \prod_{k=1}^N \frac{\eta^\varepsilon(y_k - x_k)\mu_n^k(x_k)}{(\mu_n^k * \eta^\varepsilon)(y_k)^{N-1/N}} \\ &\le 2^N \prod_{k=1}^N \eta^\varepsilon(y_k - x_k)\mu_n^k(x_k)\exp\left(\frac{(N-1)(|y_k| + R)^2}{2N\varepsilon}\right) \\ &= 2^N e^{\frac{(N-1)R^2}{2\varepsilon}} \prod_{k=1}^N \mu_n^k(x_k) e^{\frac{-|x_k|^2}{2\varepsilon}} e^{\frac{-|y_k|^2 + (2N|x_k| + 2(N-1)R)|y_k|}{2N\varepsilon}}.\end{aligned}$$

When X and ε are fixed, the latter is an integrable function of the variable $Y = (y_1, \ldots, y_N)$, and we conclude the first part of the proof thanks to the Dominate Convergence Theorem.

Recalling (4.10) we have

$$\nabla_{x_k}\Theta^\varepsilon[\mu_n](X) = \frac{\nabla\mu_n^k(x_k)}{\mu_n^k(x_k)}\Theta^\varepsilon[\mu_n](X) - \int \frac{\nabla\eta^\varepsilon(y_k - x_k)}{\eta^\varepsilon(y_k - x_k)} P^\varepsilon[\mu_n](X,Y)\,\mathrm{d}Y$$

and

$$\nabla_{x_k}\Theta^\varepsilon[\mu](X) = \frac{\nabla\mu^k(x_k)}{\mu^k(x_k)}\Theta^\varepsilon[\mu](X) - \int \frac{\nabla\eta^\varepsilon(y_k - x_k)}{\eta^\varepsilon(y_k - x_k)} P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y$$

Using the first part and the additional assumption on the pointwise convergence of the gradients, we immediately see that

$$\frac{\nabla\mu_n^k(x_k)}{\mu_n^k(x_k)}\Theta^\varepsilon[\mu_n](X) \longrightarrow \frac{\nabla\mu^k(x_k)}{\mu^k(x_k)}\Theta^\varepsilon[\mu](X),$$

pointwise a.e. on $\mathbb{R}^d \times \mathbb{R}^d$.

As for the second term, like before the integrands converge pointwise a.e., and the domination is obtained using Lemma 4.3.3 (i) and (ii). □

From Proposition 4.3.5, using some dominations already seen for the proof of Property A, we obtain the following corollary.

Corollary 4.3.6. *Suppose that $\mu_n \rightharpoonup \mu$, with $\mu_n^k \to \mu^k$ in $\mathcal{P}_p(\mathbb{R}^d)$ and $\mu_n^k \to \mu^k$ pointwise a.e. on $\mathbb{R}^d$ for every $j = 1, \dots, N$. Then $\Theta^\varepsilon[\mu_n]^{1/p} \to \Theta^\varepsilon[\mu]^{1/p}$ in $L^p((\mathbb{R}^d)^N)$.*

Assume in addiction that $\nabla \mu_n^k \to \nabla \mu^k$ pointwise a.e. on $\mathbb{R}^d$. Then $\Theta^\varepsilon[\mu_n]^{1/p} \to \Theta^\varepsilon[\mu]^{1/p}$ in $W^{1,p}((\mathbb{R}^d)^N)$.

Proof. By Proposition 4.3.5 we already have pointwise a.e. convergence of the functions. Using

$$|a^\gamma - b^\gamma| \leq |a-b|^\gamma \quad \forall \gamma \in (0,1], \forall a, b > 0, \tag{4.16}$$

we get

$$\begin{aligned}\left|\Theta^\varepsilon[\mu_n](X)^{1/p} - \Theta^\varepsilon[\mu](X)^{1/p}\right|^p &\leq |\Theta^\varepsilon[\mu_n](X) - \Theta^\varepsilon[\mu](X)| \\ &\leq \Theta^\varepsilon[\mu_n](X) + \Theta^\varepsilon[\mu](X).\end{aligned}$$

The latter converges pointwise to $2\Theta^\varepsilon[\mu](X)$, and

$$\int \Theta^\varepsilon[\mu_n](X)\,\mathrm{d}X + \int \Theta^\varepsilon[\mu](X)\,\mathrm{d}X = 2,$$

which allows to conclude the first part of the proof thanks to the Dominated Convergence Theorem.

Using the expression given by Theorem 4.2.2 and (4.16) we have

$$\begin{aligned}&\int \left|\nabla_{x_k}\Theta^\varepsilon[\mu_n]^{1/p}(X) - \nabla_{x_k}\Theta^\varepsilon[\mu]^{1/p}(X)\right|^p \mathrm{d}X \\ &\leq \frac{1}{p^p}\int \left|\Theta^\varepsilon[\mu_n](X)^{\frac{1-p}{p}}\nabla_{x_k}\Theta^\varepsilon[\mu_n](X) - \Theta^\varepsilon[\mu](X)^{\frac{1-p}{p}}\nabla_{x_k}\Theta^\varepsilon[\mu](X)\right|^p \mathrm{d}X.\end{aligned}$$

By Proposition 4.3.5 we have pointwise convergence to zero of the integrand. In order to control the gradients we recall Lemma 4.2.1 and get

$$\begin{aligned}&\left|\Theta^\varepsilon[\mu_n](X)^{\frac{1-p}{p}}\nabla_{x_k}\Theta^\varepsilon[\mu_n](X) - \Theta^\varepsilon[\mu](X)^{\frac{1-p}{p}}\nabla_{x_k}\Theta^\varepsilon[\mu](X)\right|^p \\ &\leq 2^{p-1}\left(\Theta^\varepsilon[\mu_n](X)^{1-p}\left|\nabla_{x_k}\Theta^\varepsilon[\mu_n](X)\right|^p + \Theta^\varepsilon[\mu](X)^{1-p}\left|\nabla_{x_k}\Theta^\varepsilon[\mu](X)\right|^p\right) \\ &\leq 4^{p-1}\left(\frac{\left|\nabla\mu_n^k(x_k)\right|^p}{\mu_n^k(x_k)^p}\Theta^\varepsilon[\mu_n](X) + \int\frac{|\nabla\eta^\varepsilon(y_k-x_k)|^p}{\eta^\varepsilon(y_k-x_k)^p}P^\varepsilon[\mu_n](X,Y)\,\mathrm{d}Y\right) \\ &\quad + 4^{p-1}\left(\frac{\left|\nabla\mu^k(x_k)\right|^p}{\mu^k(x_k)^p}\Theta^\varepsilon[\mu](X) + \int\frac{|\nabla\eta^\varepsilon(y_k-x_k)|^p}{\eta^\varepsilon(y_k-x_k)^p}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\right) \\ &=: 4^{p-1}g_n(X) + 4^{p-1}g(X)\end{aligned}$$

By hypothesis we have that that $g_n \to g$ pointwise a.e. as in the proof of Proposition 4.3.5. Moreover, as already seen above,

$$\int g_n(X) = p^p \int \left| \nabla \left(\mu_n^k \right)^{\frac{1}{p}} (x_k) \right|^p \mathrm{d}x_k + \int \frac{|\nabla \eta^\varepsilon(z)|^p}{\eta^\varepsilon(z)^{p-1}} \mathrm{d}z$$

and

$$\int g(X) = p^p \int \left| \nabla \left(\mu^k \right)^{\frac{1}{p}} (x_k) \right|^p \mathrm{d}x_k + \int \frac{|\nabla \eta^\varepsilon(z)|^p}{\eta^\varepsilon(z)^{p-1}} \mathrm{d}z,$$

which allows to conclude thanks to the Dominated Convergence Theorem. □

As a final result we obtain the proof of Theorem 4.3.2

Proof of Theorem 4.3.2. By contradiction, suppose that there exist $\delta > 0$ and a subsequence of (μ_n) (denoted again (μ_n) for simplicity) such that

$$d_p \left(\Theta^\varepsilon[\mu_n], \Theta^\varepsilon[\mu] \right) \geq \delta. \tag{4.17}$$

Extract a further subsequence $(\mu_{n_j})_j$ such that $\mu_{n_j}^k \to \mu^k$ in $\mathcal{P}_p(\mathbb{R}^d)$, and in addition $\mu_{n_j}^k \to \mu^k$ and $\nabla \mu_{n_j}^k \to \nabla \mu^k$ pointwise a.e. on $\mathbb{R}^d$ for every $k = 1, \ldots, N$. Due to Corollary 4.3.6 we should have $\Theta^\varepsilon[\mu_{n_j}]^{1/p} \to \Theta^\varepsilon[\mu]^{1/p}$ in $W^{1,p}((\mathbb{R}^d)^N)$, contradicting (4.17). □

4.4 Appendix

Here we complete the missing proofs from Sections 4.1-4.2. First we recall the results of paragraph 4.1

Proposition 4.1.1. *Let $p > 1$. If $u \in W^{1,p}(\mathbb{R}^m)$, $u \geq 0$, then $u^p \in W^{1,1}(\mathbb{R}^m)$, and $\nabla u^p = p u^{p-1} \nabla u$.*

Viceversa, let $u \in W^{1,1}(\mathbb{R}^m)$, $u \geq 0$, such that

$$\int u^{1-p} |\nabla u|^p < \infty. \tag{4.2}$$

Then $u^{1/p} \in W^{1,p}(\mathbb{R}^m)$, and $\nabla u^{1/p} = \frac{1}{p} u^{\frac{1-p}{p}} \nabla u$.

Proof. If $u \in W^{1,p}(\mathbb{R}^m)$ clearly $u^p \in L^1(\mathbb{R}^m)$, and viceversa if $u \in W^{1,1}(\mathbb{R}^m)$ then $u^{\frac{1}{p}} \in L^p(\mathbb{R}^m)$. Let $u_n \in C^\infty(\mathbb{R}^m) \cap W^{1,p}(\mathbb{R}^m)$ such that $u_n \to u$ in $W^{1,p}(\mathbb{R}^m)$. Then by the Hölder inequality with exponents p and $\frac{p}{p-1}$

$$\begin{aligned} \int \left| u_n^{p-1} \nabla u_n - u^{p-1} \nabla u \right| &\leq \int u_n^{p-1} \left| \nabla u_n - \nabla u \right| + \int |\nabla u| \left| u_n^{p-1} - u^{p-1} \right| \\ &= \|u_n\|_p^{p-1} \|\nabla u - \nabla u_n\|_p \\ &\quad + \|\nabla u\|_p \left\| \left| u_n^{p-1} - u^{p-1} \right|^{\frac{1}{p-1}}_p \right\|^{p-1}. \end{aligned}$$

Recall that

$$|a^q - b^q| \le |a-b|\,|a+b|^{q-1} \quad \forall q \in [1,\infty), \forall a,b > 0. \tag{4.18}$$

If $p \ge 2$ we use (4.18) and the Hölder inequality to get

$$\left\| \left| u_n^{p-1} - u^{p-1} \right|^{\frac{1}{p-1}} \right\|_p^{p-1} \le \|u_n - u\|_p \, \|u_n + u\|_p^{p-2};$$

if $1 < p < 2$, let $\gamma = p - 1 \in (0,1)$ and use (4.16) to get

$$\left\| \left| u_n^{p-1} - u^{p-1} \right|^{\frac{1}{p-1}} \right\|_p^{p-1} \le \|u_n - u\|_p^{p-1}.$$

This completes the proof of the first part. Suppose on the contrary that $u \in W^{1,1}(\mathbb{R}^m)$, $u \ge 0$, and that the condition (4.2) holds. Fix $\phi \in C_c^\infty(\mathbb{R}^m)$ and $\varepsilon > 0$. We want to prove that

$$\int (u+\varepsilon)^{\frac{1}{p}} \nabla\phi = -\frac{1}{p} \int \phi (u+\varepsilon)^{\frac{1-p}{p}} \nabla u. \tag{4.19}$$

To this end, let $u_n \to u$ in $W^{1,1}(\mathbb{R}^m)$, where $u_n \in C^\infty$, $u_n \ge 0$; up to a subsequence we may suppose also $u_n \to u$ and $\nabla u_n \to \nabla u$ pointwise almost everywhere. Putting u_n in place of u in (4.19) we have pointwise convergence of both the integrands, and we conclude via the Dominated Convergence Theorem using the dominations

$$\left| \phi (u_n+\varepsilon)^{\frac{1-p}{p}} \nabla u_n \right| \le \varepsilon^{\frac{1-p}{p}} |\phi|\,|\nabla u_n|, \quad \left| \phi (u+\varepsilon)^{\frac{1-p}{p}} \nabla u \right| \le \varepsilon^{\frac{1-p}{p}} |\phi|\,|\nabla u|.$$

Finally, letting $\varepsilon \to 0$ in (4.19), we have once again pointwise convergence of the integrands, and we conclude by the classical Lebesgue's dominated covergence Theorem thanks to the hypothesis and the domination

$$\left| \phi (u+\varepsilon)^{\frac{1-p}{p}} \nabla u \right|^p \le |\phi|^p \, u^{1-p} \, |\nabla u|^p. \qquad \square$$

Proposition 4.1.2. *If $u_n \to u$ in $W^{1,p}(\mathbb{R}^m)$, $u_n, u \ge 0$, then $u_n^p \to u^p$ in $W^{1,1}(\mathbb{R}^m)$.*

Viceversa, let $u_n \to u$ in $W^{1,1}(\mathbb{R}^d)$, $u_n, u \ge 0$. Let $h_n, h \in L^1(\mathbb{R}^m)$ such that $u_n^{1-p} |\nabla u_n|^p \le h_n$, $u^{1-p} |\nabla u|^p \le h$, and

$$\lim_{n\to\infty} \int h_n = \int h. \tag{4.3}$$

Suppose also that for every subsequence $\{h_{n_k}\}$ there exists a further subsequence converging to h pointwise a.e. Then $u_n^{1/p} \to u^{1/p}$ in $W^{1,p}(\mathbb{R}^m)$.

Proof of Proposition 4.1.2. If $p = 1$ there is nothing to prove, so assume $p > 1$, and take $u_n \to u$ in $W^{1,p}(\mathbb{R}^m)$. Using (4.18) and the Hölder inequality with exponents p and $\frac{p}{p-1}$,

$$\int |u_n^p - u^p| \le \|u_n - u\|_p \|u_n + u\|_p^{p-1}.$$

Since $u_n \to u$ in $W^{1,p}(\mathbb{R}^m)$ and hence in particular u_n is bounded in $L^p(\mathbb{R}^m)$, we get that $u_n^p \to u^p$ (strongly) in $L^1(\mathbb{R}^m)$.

Moreover, $\nabla u_n^p = p u_n^{p-1} \nabla u_n$ and $\nabla u^p = p u^{p-1} \nabla u$ by Proposition 4.1.1, hence by the Hölder inequality

$$\begin{aligned}\int |\nabla u_n^p - \nabla u^p| &\le p \int u_n^{p-1} |\nabla u_n - \nabla u| + p \int |\nabla u| \left|u_n^{p-1} - u^{p-1}\right| \\ &\le p \|u_n\|_p^{p-1} \|\nabla u_n - \nabla u\|_p + p \int |\nabla u| \left|u_n^{p-1} - u^{p-1}\right|,\end{aligned}$$

which converges to zero as in the proof of Proposition 4.1.1.

To prove the converse, suppose by contradiction that there is a subsequence (denoted again u_n) such that

$$\left\| u_n^{1/p}, u^{1/p} \right\|_{W^{1,p}} \ge \delta > 0. \tag{4.20}$$

By hypothesis, up to a further subsequence we may assume that $u_{n_j} \to u$, $\nabla u_{n_j} \to \nabla u$ and $h_{n_j} \to h$ pointwise almost everywhere. Then we have by (4.16), with $\gamma = \frac{1}{p}$,

$$\int \left| u_{n_j}^{1/p} - u^{1/p} \right|^p \le \int |u_{n_j} - u| = \|u_{n_j} - u\|_1,$$

and

$$\left\| \nabla u_{n_j}^{1/p} - \nabla u^{1/p} \right\|_p = \frac{1}{p^p} \int \left| u_{n_j}^{\frac{1-p}{p}} \nabla u_{n_j} - u^{\frac{1-p}{p}} \nabla u \right|^p.$$

Here the integrand converges to zero pointwise, and using the domination

$$\left| u_{n_j}^{\frac{1-p}{p}} \nabla u_{n_j} - u^{\frac{1-p}{p}} \nabla u \right|^p \le 2^{p-1} \left(u_{n_j}^{1-p} |\nabla u_{n_j}|^p + u^{1-p} |\nabla u|^p \right) \le 2^{p-1} (h_{n_j} + h)$$

and the condition (4.3) we conclude thanks to the Dominated Convergence Theorem that $u_{n_j}^{1/p} \to u^{1/p}$ in $W^{1,p}(\mathbb{R}^m)$, contradicting (4.20). □

We are ready for the proof of Proposition 4.1.3, which we recall for the sake of the reader.

Proposition 4.1.3. *Let $p > 1$. Then*

(i) if $\mu \in \mathcal{P}_p((\mathbb{R}^d)^N)$, then $\mu^k \in \mathcal{P}_p(\mathbb{R}^d)$ for every $k = 1, \ldots, N$;

(ii) the map $\pi\colon \mathcal{P}_p\left((\mathbb{R}^d)^N\right) \longrightarrow \mathcal{P}_p(\mathbb{R}^d)^N$ is continuous with respect to the distance d_p and the relative product topology on the codomain.

Proof. For the first part, by Proposition 4.1.1 it suffices to show that, if $\mu \in \mathcal{P}_p((\mathbb{R}^d)^N)$, then $\int \mu^k(x)^{1-p} \left|\nabla \mu^k(x)\right|^p \mathrm{d}x$ is finite. Using the Hölder inequality with exponents $\frac{p}{p-1}$ and p, we get

$$\begin{aligned}
\left|\nabla \mu^k(x_k)\right| &\leq p \int \mu(X)^{\frac{p-1}{p}} \left|\nabla_{x_k} \mu^{\frac{1}{p}}(X)\right| \mathrm{d}\hat{X}_k \\
&\leq p \left(\int \mu(X)\, \mathrm{d}\hat{X}_k\right)^{\frac{p-1}{p}} \left(\int \left|\nabla_{x_k} \mu^{\frac{1}{p}}(X)\right|^p \mathrm{d}\hat{X}_k\right)^{\frac{1}{p}} \\
&= p\mu^k(x_k)^{\frac{p-1}{p}} \left(\int \left|\nabla_{x_k} \mu^{\frac{1}{p}}(X)\right|^p \mathrm{d}\hat{X}_k\right)^{\frac{1}{p}}.
\end{aligned}$$

Hence

$$\int \mu^k(x_k)^{1-p} \left|\nabla \mu^k(x_k)\right|^p \mathrm{d}x_k \leq p^p \int \left|\nabla_{x_k} \mu^{1/p}(X)\right|^p \mathrm{d}X$$

which is finite by hypothesis.

Now we come to the second part. Let $\mu_n \to \mu$ in $\mathcal{P}_p\left((\mathbb{R}^d)^N\right)$ and fix $k \in \{1, \ldots, k\}$. First of all we prove that $\mu_n^k \to \mu^k$ in $W^{1,1}(\mathbb{R}^d)$. Using (4.18) and the Hölder inequality,

$$\begin{aligned}
\int \left|\mu_n^k(x_k) - \mu^k(x_k)\right| \mathrm{d}x_k &= \int \left|\int \mu(X) - \mu_n(X)\, \mathrm{d}\hat{X}_k\right| \mathrm{d}x_k \\
&\leq \int \left|\mu_n(X) - \mu(X)\right| \mathrm{d}X
\end{aligned}$$

and

$$\begin{aligned}
\int \left|\nabla \mu_n^k(x_k) - \nabla \mu^k(x_k)\right| \mathrm{d}x_k &= \int \left|\int \nabla_{x_k} \mu_n(X) - \nabla_{x_k} \mu(X)\, \mathrm{d}\hat{X}_k\right| \mathrm{d}x_k \\
&\leq \int \left|\nabla_{x_k} \mu_n(X) - \nabla_{x_k} \mu(X)\right| \mathrm{d}X.
\end{aligned}$$

We deduce that $\mu_n^k \to \mu^k$ in $W^{1,1}(\mathbb{R}^d)$, since $\mu_n \to \mu$ in $W^{1,p}((\mathbb{R}^d)^N)$ by Proposition 4.1.1. Now we want to apply Proposition 4.1.2, with

$$h_n(x_k) = p^p \int \left|\nabla_{x_k} \mu_n^{1/p}(X)\right|^p \mathrm{d}\hat{X}_k, \quad h(x_k) = p^p \int \left|\nabla_{x_k} \mu^{1/p}(X)\right|^p \mathrm{d}\hat{X}_k$$

in order to conclude that $(\mu_n^k)^{1/p} \to (\mu_n^k)^{1/p}$ in $W^{1,p}(\mathbb{R}^d)$.

By the first part we have $(\mu_n^k)^{1-p} \left|\nabla \mu_n^k\right|^p \leq h_n$ and $(\mu^k)^{1-p} \left|\nabla \mu^k\right|^p \leq h$. Condition (4.3) is ensured by

$$\begin{aligned}
\lim_{n\to\infty} \int h_n(x_k)\, \mathrm{d}x_k &= p^p \lim_{n\to\infty} \int \left|\nabla_{x_k} \mu_n^{1/p}(X)\right|^p \mathrm{d}X \\
&= p^p \lim_{n\to\infty} \left\|\nabla_{x_k} \mu_n^{1/p}\right\|_p^p \\
&= p^p \left\|\nabla_{x_k} \mu^{1/p}\right\|_p^p \\
&= \int h(x_k)\, \mathrm{d}x_k.
\end{aligned}$$

We now follow a construction similar to the one of the Riesz-Fischer theorem, and already used for the analogous result by Brezis in [34, Appendix]. Recall that, by Proposition 4.1.2, $\mu_n \to \mu$ in $W^{1,1}\left((\mathbb{R}^d)^N\right)$. For every subsequence of $(h_n)_n$(denoted again $(h_n)_n$ for simplicity), extract a further subsequence $(h_{n_j})_j$ such that:

(i) $\nabla \mu_{n_j}^{1/p} \to \nabla \mu^{1/p}$ pointwise a.e.;

(ii) $\left\|\nabla \mu_{n_j}^{1/p} - \nabla \mu^{1/p}\right\|_{L^p}^p \leq 2^{-j}$.

Let

$$F(X) = \left|\nabla \mu^{1/p}(X)\right|^p + \sum_{j=1}^{\infty} \left|\nabla \mu_{n_j}^{1/p}(X) - \nabla \mu^{1/p}(X)\right|^p.$$

Since $F \in L^1\left((\mathbb{R}^d)^N\right)$ and clearly

$$\left|\nabla \mu_{n_j}^{1/p}(X)\right|^p \leq 2^{p-1} F(X), \quad \left|\nabla \mu^{1/p}(X)\right|^p \leq F(X)$$

we have that $h_{n_j} \to h$ pointwise a.e. by dominated convergence. □

Next we prove the estimates on the energy given in paragraph 4.1.

Lemma 4.1.4. *Let $\mu \in \mathcal{P}_p\left((\mathbb{R}^d)^N\right)$. Then*

$$\mathcal{E}_p(\mu) \geq \sum_{k=1}^{N} \mathcal{E}_p(\mu^k).$$

Moreover, if $\rho_1, \ldots, \rho_N \in \mathcal{P}_p(\mathbb{R}^d)$,

$$\inf\left\{\mathcal{E}_p(\mu) \mid \mu \in \mathcal{P}_p(\mathbb{R}^m) \cap \Pi(\rho_1, \ldots, \rho_N)\right\} = \sum_{k=1}^{N} \mathcal{E}_p(\rho_k).$$

Proof. Fix $\mu \in \mathcal{P}_p\left((\mathbb{R}^d)^N\right)$. By Proposition 4.1.3 we have

$$\left|\nabla(\mu^k)^{1/p}(x_k)\right|^p = \frac{1}{p^p}\mu^k(x_k)^{1-p}\left|\nabla\mu^k(x_k)\right|^p \leq \int \left|\nabla_{x_k}\mu^{1/p}(X)\right|^p \mathrm{d}\hat{X}_k.$$

Summing on k and recalling the condition (4.1) we get the thesis. As for the second statement, due to the first one clearly we have

$$\inf\left\{\mathcal{E}_p(\mu) \mid \mu \in \Pi(\rho_1, \dots, \rho_N)\right\} \geq \sum_{k=1}^N \mathcal{E}_p(\rho_k).$$

Let however $\mu(X) := \rho_1(x_1)\cdots\rho_N(x_N)$; then μ is such that $\mu \in \mathcal{P}_p\left((\mathbb{R}^d)^N\right)$ and

$$\nabla_{x_k}\mu^{1/p} = \nabla\rho_k^{1/p}\prod_{\substack{j=1\\ j\neq k}}^N \rho_j(x_j)^{1/p};$$

hence

$$\int \left|\nabla_{x_k}\mu^{\frac{1}{p}}(X)\right|^p \mathrm{d}X = \int \left|\nabla\rho_k^{1/p}(x_k)\right|^p \mathrm{d}x_k = \mathcal{E}_p(\rho_k).$$

Finally summing on k and taking into account the usual condition (4.1),

$$\mathcal{E}_p(\mu) = \sum_{k=1}^N \mathcal{E}_p(\rho_k). \qquad \square$$

Lemma 4.1.5. *Let $\eta \in C^\infty(\mathbb{R}^m)$, $\eta \geq 0$ such that $\int \eta = 1$ and define $\eta^\varepsilon(x) = \frac{1}{\varepsilon^m}\eta\left(\frac{x}{\varepsilon}\right)$, for $\varepsilon > 0$. Then, for every $\mu \in \mathcal{P}_p(\mathbb{R}^m)$,*

$$\mathcal{E}_p(\mu * \eta^\varepsilon) \leq \mathcal{E}_p(\mu) \quad \textit{and} \quad \lim_{\varepsilon\to 0}\mathcal{E}_p(\mu * \eta^\varepsilon) = \mathcal{E}_p(\mu).$$

Proof. By the Hölder inequality with exponents p and $\frac{p}{p-1}$ we have

$$\begin{aligned}
|\nabla(\mu * \eta^\varepsilon)(x)| &= |((\nabla\mu) * \eta^\varepsilon)(x)| \\
&\leq \int |\nabla\mu(y)|\,\eta^\varepsilon(x-y)\,\mathrm{d}y \\
&\leq \left(\int \mu(y)^{1-p}\,|\nabla\mu(y)|^p\,\eta^\varepsilon(x-y)\,\mathrm{d}y\right)^{\frac{1}{p}}(\mu * \eta^\varepsilon)(x)^{\frac{p-1}{p}}.
\end{aligned}$$

Since $\mu * \eta^\varepsilon \in C^\infty(\mathbb{R}^m)$ we have

$$\begin{aligned}
\left|\nabla(\mu * \eta^\varepsilon)^{1/p}(x)\right| &= \frac{1}{p}(\mu * \eta^\varepsilon)(x)^{\frac{1-p}{p}}\,|\nabla(\mu * \eta^\varepsilon)(x)| \\
&\leq \frac{1}{p}\left(\int \mu(y)^{1-p}\,|\nabla\mu(y)|^p\,\eta^\varepsilon(x-y)\,\mathrm{d}y\right)^{1/p}, \qquad (4.21)
\end{aligned}$$

whence

$$\begin{aligned}\mathcal{E}_p(\mu * \eta^\varepsilon) &= \int \left|\nabla(\mu * \eta^\varepsilon)^{1/p}(x)\right|^p \mathrm{d}x \\ &\le \frac{1}{p^p}\int \mu(y)^{1-p}\,|\nabla\mu(y)|^p\,\eta^\varepsilon(x-y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int \left|\nabla\mu^{1/p}(y)\right|^p \mathrm{d}y = \mathcal{E}_p(\mu).\end{aligned}$$

In order to prove the second part, it suffices to show that $(\mu * \eta^\varepsilon)^{1/p}$ converges strongly to $\mu^{1/p}$ in $W^{1,p}$ to get that

$$\lim_{\varepsilon\to 0}\mathcal{E}_p(\mu * \eta^\varepsilon) = \lim_{\varepsilon\to 0}\left\|(\mu * \eta^\varepsilon)^{1/p}\right\|^p_{W^{1,p}} - 1 = \left\|\mu^{1/p}\right\|_{W^{1,p}} - 1 = \mathcal{E}_p(\mu).$$

Since $(\mu^{1-p}\,|\nabla\mu|^p) * \eta^\varepsilon \longrightarrow \mu^{1-p}\,|\nabla\mu|^p$ pointwise a.e., the inequality (4.21) gives a domination which allows to conclude thanks to Proposition 4.1.2. $\square$

Finally, we present a slight modification of Theorem 4.2.3 for the case $p = 2$, which allows to get sharper constants, as stated in Theorem 4.2.4.

Theorem 4.2.4. *Let $\mu \in \mathcal{P}\left((\mathbb{R}^d)^N\right)$ such that $\mu^k \in \mathcal{P}_2(\mathbb{R}^d)$ for every $k = 1, \dots, N$. Then*

$$\mathcal{E}_2(\Theta^\varepsilon[\mu]) \le \frac{Nc(\eta)}{\varepsilon^2} + \sum_{k=1}^N \mathcal{E}_2(\mu^k), \tag{4.15}$$

where $c(\eta)$ is a constant depending on the choice of η.

Proof. We use the formula for the gradient given by Theorem 4.2.2, and exploit the Hilbertian structure of $W^{1,2}((\mathbb{R}^d)^N)$ to get

$$\begin{aligned}\mathcal{E}_2(\Theta^\varepsilon[\mu]) \;&=\; \int \left|\nabla\sqrt{\Theta^\varepsilon[\mu]}(X)\right|^2 \mathrm{d}X \\ &= \frac{1}{4}\int \frac{|\nabla\Theta^\varepsilon[\mu](X)|^2}{\Theta^\varepsilon[\mu](X)}\,\mathrm{d}X \\ &= \frac{1}{4}\sum_{k=1}^N \int \frac{\left|\nabla\mu^k(x_k)\right|^2}{\mu^k(x_k)^2}\Theta^\varepsilon[\mu](X)\,\mathrm{d}X \\ &\quad - \sum_{k=1}^N \iint \frac{\nabla\mu^k(x_k)\cdot\nabla\eta^\varepsilon(y_k - x_k)}{\mu^k(x_k)\eta^\varepsilon(y_k - x_k)} P^\varepsilon[\mu](X,Y)\,\mathrm{d}X\,\mathrm{d}Y \\ &\quad + \sum_{k=1}^N \int \frac{1}{\Theta^\varepsilon[\mu](X)}\left|\int \frac{\nabla\eta^\varepsilon(y_k - x_k)}{\eta^\varepsilon(y_k - x_k)} P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\right|^2 \mathrm{d}X \\ &=: I - II + III.\end{aligned}$$

We treat the three terms in order. First we have

$$I = \frac{1}{4}\sum_{k=1}^N \int \frac{\left|\nabla\mu^k(x_k)\right|^2}{\mu^k(x_k)}\,\mathrm{d}x_k = \sum_{k=1}^N \int \left|\nabla\sqrt{\mu^k}(x_k)\right|^2 \mathrm{d}x_k = \sum_{k=1}^N \mathcal{E}_2(\mu^k).$$

The second term vanishes. Indeed, using Fubini's theorem and a change of variables,

$$
\begin{aligned}
II &= \iint \frac{\nabla\mu^k(x_k)\cdot\nabla\eta^\varepsilon(y_k-x_k)}{\mu^k(x_k)\eta^\varepsilon(y_k-x_k)}\frac{\eta^\varepsilon(y_k-x_k)}{(\mu^k*\eta^\varepsilon)(y_k)}\mu^k(x_k)\Lambda^\varepsilon[\mu](Y)\,\mathrm{d}x_k\,\mathrm{d}Y \\
&= \iint \nabla\mu^k(x_k)\cdot\nabla\eta^\varepsilon(y_k-x_k)\,\mathrm{d}x_k\,\mathrm{d}y_k \\
&= \left(\iint \nabla\mu^k(x_k)\,\mathrm{d}x_k\right)\cdot\left(\int \nabla\eta^\varepsilon(z)\,\mathrm{d}z\right),
\end{aligned}
$$

and the second term is zero, as it can be seen, for instance, integrating in spherical coordinates — recall that we chose a radial kernel η.

Finally, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\left|\int \frac{\nabla\eta^\varepsilon(y_k-x_k)}{\eta^\varepsilon(y_k-x_k)}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\right|^2 \\
&\le \int \frac{|\nabla\eta^\varepsilon(y_k-x_k)|^2}{\eta^\varepsilon(y_k-x_k)^2}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\int P^\varepsilon[\mu](X,Y')\,\mathrm{d}Y' \\
&= \Theta^\varepsilon[\mu](X)\int \frac{|\nabla\eta^\varepsilon(y_k-x_k)|^2}{\eta^\varepsilon(y_k-x_k)^2}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y.
\end{aligned}
$$

Hence

$$
\begin{aligned}
III &\le \sum_{k=1}^N \iint \frac{|\nabla\eta^\varepsilon(y_k-x_k)|^2}{\eta^\varepsilon(y_k-x_k)^2}P^\varepsilon[\mu](X,Y)\,\mathrm{d}Y\,\mathrm{d}X \\
&= \sum_{k=1}^N \iint \frac{|\nabla\eta^\varepsilon(y_k-x_k)|^2}{\eta^\varepsilon(y_k-x_k)}\mu^k(x_k)\Lambda^\varepsilon[\mu](Y)\,\mathrm{d}Y\,\mathrm{d}x_k \\
&= \sum_{k=1}^N \iint \frac{|\nabla\eta^\varepsilon(y_k-x_k)|^2}{\eta^\varepsilon(y_k-x_k)}\mu^k(x_k)\,\mathrm{d}y_k\,\mathrm{d}x_k = N\int \frac{|\nabla\eta^\varepsilon(z)|^2}{\eta^\varepsilon(z)}\,\mathrm{d}z \le \frac{Nc(\eta)}{\varepsilon^2}.
\end{aligned}
$$

where $c(\eta)$ is a constant depending on the choice of η. $\square$

Chapter 5

Applications to Density Functional Theory

Introduction

Consider a system of N particles (*e.g.* electrons) interacting through Coulomb force both with each other and with M other fixed particles (*e.g.* nuclei) at positions $(r_1, \ldots, r_M)$ with charges $(-Z_1, \ldots, -Z_M)$. According to the Schrödinger equation, the ground state energy of the system is given by

$$\min_{\psi \text{ wave-function}} \left\{ \hbar^2 T(\psi) + V_{ee}(\psi) + V_{ne}(\psi) \right\} \tag{5.1}$$

where

$$\begin{aligned}
T(\psi) &= \frac{1}{2} \int |\nabla \psi(X)|^2 \, dX && \text{is the kynetic energy,} \\
V_{int}(\psi) &= \sum_{1 \le i < j \le N} \int \frac{|\psi(X)|^2}{|x_i - x_j|} dX && \text{is the internal interaction energy,} \\
V_{ext}(\psi) &= -\sum_{k=1}^{M} \sum_{i=1}^{N} \int \frac{Z_k |\psi(X)|^2}{|x_i - r_k|} dX && \text{is the external interaction energy}
\end{aligned}$$

and $\hbar$ is the Planck constant.

We must now specify which is the set of admissible wave-functions. If the particles are electrons (or, more in general, fermions), then they follow the Fermi-Dirac statistics and the set of wave-functions is given by

$$\mathcal{A} = \left\{ \psi \in H^1((\mathbb{R}^3 \times \{\uparrow, \downarrow\})^N; \mathbb{C}) \mid \int |\psi(X)|^2 \, dX = 1, \psi \text{ is antisymmetric} \right\},$$

where we say that a wave-function is *antisymmetric* if

$$\psi(x_1, \alpha_1, \ldots, x_N, \alpha_N) = \operatorname{sign}(\sigma) \psi(x_{\sigma(1)}, \alpha_{\sigma(1)}, \ldots, x_{\sigma(N)}, \alpha_{\sigma(N)})$$

for every permutation $\sigma \in \mathfrak{S}_N$. On the contrary, if the particles are bosons, then they follow the Bose-Einstein statistics and the set of wave-functions is given by

$$\mathcal{S} = \left\{ \psi \in H^1((\mathbb{R}^3 \times \{\uparrow, \downarrow\})^N; \mathbb{C}) \mid \int |\psi(X)|^2 \, dX = 1, \psi \text{ is symmetric} \right\},$$

where we say that a wave-function is *symmetric* if

$$\psi(x_1, \alpha_1, \ldots, x_N, \alpha_N) = \psi(x_{\sigma(1)}, \alpha_{\sigma(1)}, \ldots, x_{\sigma(N)}, \alpha_{\sigma(N)})$$

In the above we adopted the common notation

$$\int f(X) dX := \sum_{(\alpha_1, \ldots, \alpha_N) \in \{\uparrow, \downarrow\}^N} \int_{(\mathbb{R}^3)^N} f(x_1, \alpha_1, \ldots, x_N, \alpha_N) dx_1 \cdots dx_N$$

for every $f \colon (\mathbb{R}^3 \times \{\uparrow, \downarrow\})^N \to \mathbb{C}$.

According to the usual Born interpretation, $|\psi(x_1, \alpha_1, \ldots, x_N, \alpha_N)|^2$ is the probability distribution of finding the N particles in positions $(x_1, \ldots, x_N)$ with spin values $(\alpha_1, \ldots, \alpha_N)$ and, in agreement with the indistinguishability principle, it is invariant with respect to permutations of the N variables (x_i, α_i).

Computing the ground values above amounts to solving a Schrödinger equation in $\mathbb{R}^{3N}$ and the numerical cost scales exponentially with N. The Density Functional Theory (DFT from now on) is an alternative introduced in the late sixties by Hohenberg, Kohn and Sham. However the desire to describe the system in term of a different variable is much older and we may consider the Thomas-Fermi model as a precursor of this theory. Nowadays, DFT is recognized as the absolute best tool for computing the ground state of complex systems, in virtue of its excellent compromise between computational efficiency and accuracy. See [27] for a classical monography, and the works of P. Gori-Giorgi, M. Seidl, A. Gerolin, S. Di Marino for recent contributions and the current state-of-art.

The idea is to associate to every wave function ψ a probability density on $\mathbb{R}^3$ defined as follows [1]:

$$\rho[\psi](x) := \sum_{(\alpha_1, \ldots, \alpha_N) \in \{\uparrow, \downarrow\}^N} \int_{(\mathbb{R}^3)^{N-1}} |\psi(x, \alpha_1, x_2, \alpha_2, \ldots, x_N, \alpha_N)|^2 \, dx_2 \cdots dx_N.$$

It can be shown (see for instance [34, Theorems 1.1 and 1.2]) that the set of feasible densities is given by

$$\begin{aligned} \mathcal{R} &:= \left\{ \rho \in \mathcal{P}(\mathbb{R}^3) \mid \rho = \rho[\psi] \text{ for some } \psi \in \mathcal{A} \right\} \\ &= \left\{ \rho \in \mathcal{P}(\mathbb{R}^3) \mid \rho = \rho[\psi] \text{ for some } \psi \in \mathcal{S} \right\} \\ &= \left\{ \rho \colon \mathbb{R}^3 \to [0, +\infty] \mid \int \rho = 1, \ \sqrt{\rho} \in H^1(\mathbb{R}^d) \right\} \end{aligned}$$

[1] In the usual definition in quantum chemistry, the integral in the definition of ρ_ψ is also multiplied by a factor N, but for our presentation from the mathematical viewpoint we prefer to deal with probability measures.

The key remark which constitutes the starting point of DFT is the following: the minimum problem (5.1) can be factorized as

$$\min_{\rho\in\mathcal{R}} \min_{\psi\mapsto\rho} \left\{\hbar^2 T(\psi) + V_{int}(\psi) + V_{ext}(\psi)\right\},$$

where $\psi \mapsto \rho$ denotes the fact that $\rho = \rho[\psi]$.

Observe that, if $\psi \mapsto \rho$, then by symmetry

$$V_{ext}(\psi) = -\sum_{i=1}^{N}\sum_{k=1}^{M} \int \frac{Z_k \left|\psi(X)\right|^2}{|x_i - r_k|} dX = -N \sum_{k=1}^{M} \int_{\mathbb{R}^3} \frac{Z_k \rho(x)}{|x - r_k|} dx = v_{ext}(\rho).$$

Hence, it is natural to define the Levy-Lieb functional

$$F_\hbar(\rho) = \min_{\psi\mapsto\rho} \left\{\hbar^2 T(\psi) + V_{int}(\psi)\right\},$$

so that we reduce (5.1) to the minimum problem

$$\min_{\rho\in\mathcal{R}} \left\{F_\hbar(\rho) + v_{ext}(\rho)\right\}.$$

The Levy-Lieb functional $F_\hbar$ is also called "universal Levy-Lieb functional", since it does not depend on the specific external potential — it only depends on the density ρ, the number of particles N and the parameter $\hbar$. We should be careful, however, in the definition of the Levy-Lieb functional, to distinguish between the fermionic and the bosonic case. For the sake of clarity, let us define

$$F_\hbar^{\mathcal{A}} := \min\left\{\hbar^2 T(\psi) + V_{int}(\psi) \mid \psi \in \mathcal{A}, \psi \mapsto \rho\right\}$$

when we consider fermionic particles, and by

$$F_\hbar^{\mathcal{S}} := \min\left\{\hbar^2 T(\psi) + V_{int}(\psi) \mid \psi \in \mathcal{S}, \psi \mapsto \rho\right\}$$

when we consider bosonic particles.

In this chapter we are going to deepen the study of the Levy-Lieb functional by treating two main topics of DFT:

(1) **Continuity of the Levy-Lieb functional**

(2) **Semiclassical limit of the Levy-Lieb functional**

In order to give a more general mathematical framework, in the following we will denote by d the spatial dimension of the system of particles, without necessarily reducing the setting to the physical case $d = 3$.

5.1 Continuity of the Levy-Lieb functional

The main question of this Section is the following: is the Levy-Lieb functional continuous with respect to the density ρ?

First of all we need to specify the topology for the space $\mathcal{R}$ of the densities. Recall that

$$\mathcal{R} = \left\{ \rho\colon \mathbb{R}^3 \to [0, +\infty] \mid \int \rho = 1, \sqrt{\rho} \in H^1(\mathbb{R}^3) \right\}.$$

A natural topology is given by considering $\mathcal{R}$ as a metric space endowed with the distance

$$\delta(\rho_1, \rho_2) = \|\rho_1 - \rho_2\|_{H^1(\mathbb{R}^3)} .$$

This turns out to be also a physically meaningful topology, since we have the following

Theorem 5.1.1 ([34, Theorem 1.3])**.** *Let $\psi_n, \psi \in \mathcal{A}$ or $\mathcal{S}$, and suppose that $\psi_n \to \psi \in H^1((\mathbb{R}^d \times \{\uparrow, \downarrow\})^N; \mathbb{C})$. Then $\sqrt{\rho[\psi_n]} \to \sqrt{\rho[\psi]} \in H^1(\mathbb{R}^d)$.*

This amounts to say that the map $\psi \mapsto \rho[\psi]$ is continuous, but it is quite clear that this map is not invertible — in fact, different wave-functions may well share the same single particle density. However, suppose that $(\sqrt{\rho_k})_{k\geq 1}$ converge to $\sqrt{\rho}$ in H^1, and take ψ such that $\rho = \rho[\psi]$. Can we find $(\psi_k)_{k\geq 1}$ such that $\rho_k = \rho[\psi_k]$ and $\psi_k \to \psi$ in H^1? In other words, is the map $\psi \mapsto \rho[\psi]$ open? This problem, to our knowledge first stated in [34, Question 2], is still open.

This question is strongly related to the continuity of the Levy-Lieb functional, since we have the following result.

Theorem 5.1.2. *If the map $\psi \mapsto \rho[\psi]$ is open, the Levy-Lieb functional is continuous.*

Proof. The idea is to use the De Giorgi's main theorem on Γ-convergence (see [22, Theorem 7.4]). Given a sequence (ρ_k) such that $\rho_k \to \rho \in \mathcal{R}$, consider the functionals

$$\mathcal{F}_k(\psi) = \hbar^2 T(\psi) + V_{int}(\psi) + \chi_{\{\psi\in\mathcal{A} \mid \psi\mapsto\rho_k\}}(\psi)$$

and

$$\mathcal{F}(\psi) = \hbar^2 T(\psi) + V_{int}(\psi) + \chi_{\{\psi\in\mathcal{A} \mid \psi\mapsto\rho\}}(\psi),$$

where χ is the characteristic function of a set, taking values 0 on the set and $+\infty$ outside. By definition we have that $F_\hbar^{\mathcal{A}}(\rho_k) = \min\left\{\mathcal{F}_k(\psi) \mid \psi \in H^1\right\}$ and $F_\hbar^{\mathcal{A}}(\rho_k) = \min\left\{\mathcal{F}(\psi) \mid \psi \in H^1\right\}$. Thus, we need to prove that $\mathcal{F}_k$ Γ-converges to $\mathcal{F}$.[2]

[2]The condition $\rho_k \to \rho$ implies tightness of the set $S = \{\rho\} \cup \{\rho_k\}_{k\in\mathbb{N}}$, and taking the counterimage of a tight set by the map sending a measure to its marginal preserves tightness. Thus the set $\{\psi \mid \rho[\psi] \in S\}$ is sequentially compact.

$\liminf$-**inequality** Take any sequence $\psi_k \to \psi$. If $\rho \neq \rho[\psi]$, then by Theorem 5.1.1 we have that $\rho[\psi_k] \neq \rho_k$ definitively, so that the lim inf-inequality is trivially satisfied. Hence we can assume that $\rho[\psi] = \rho$. Since the functionals T and V_{int} are continuous[3], we get

$$\mathcal{F}(\psi) = \lim_{k\to\infty} \hbar^2 T(\psi_k) + V_{int}(\psi_k) \leq \liminf_{k\to\infty} \mathcal{F}_k(\psi_k).$$

$\limsup$-**inequality** If $\rho[\psi] \neq \rho$ there is nothing to prove, so assume that $\rho = \rho[\psi]$. Take as a recovery sequence the one given by the openness of the map $\psi \mapsto \rho[\psi]$, *i.e.*, take (ψ_k) such that $\rho[\psi_k] = \rho_k$ and $\psi_k \to \psi \in H^1$. Then we have

$$\mathcal{F}(\psi) = \hbar^2 T(\psi) + V_{int}(\psi) = \lim_{k\to\infty} \hbar^2 T(\psi_k) + V_{int}(\psi_k) = \lim_{k\to\infty} \mathcal{F}_k(\psi_k). \qquad \square$$

We will provide a partial positive answer when the spin variables are not taken into account. In particular, we prove the following results.

Theorem 5.1.3 ([5, Theorem 1.1])**.** *Let $\psi \in H^1((\mathbb{R}^d)^N;\mathbb{R})$ symmetric and non-negative. Given $(\rho_n)_{n\geq 1}$ such that $\sqrt{\rho_n} \to \sqrt{\rho[\psi]}$ in $H^1(\mathbb{R}^d)$, there exist $(\psi_n)_{n\geq 1}$ symmetric and non-negative such that $\rho_n = \rho[\psi_n]$ and $\psi_n \to \psi$ in $H^1((\mathbb{R}^d)^N;\mathbb{R})$.*

Theorem 5.1.4 ([5, Theorem 1.2])**.** *Let $\psi \in H^1((\mathbb{R}^d)^N;\mathbb{R})$ symmetric. Given $(\rho_n)_{n\geq 1}$ such that $\sqrt{\rho_n} \to \sqrt{\rho[\psi]}$ in $H^1(\mathbb{R}^d)$, there exist $(\psi_n)_{n\geq 1}$ symmetric and complex-valued such that $\rho_n = \rho[\psi_n]$ and $\psi_n \to \psi$ in $H^1((\mathbb{R}^d)^N;\mathbb{C})$.*

Notice that the first result is already of physical interest, since in many cases the ground state of a system of N-particles is non-negative.

The main tools will be the smoothing of transport plans as introduced and studied in [3, 4] and an application of the weighted Sobolev spaces. In Section 5.1 we will start by constructing explicitly an L^2 approximation of $|\psi|$ which respects the marginal constraint. Then in Section 5.1 we regularize it in order to obtain a Sobolev regular sequence which converges in H^1 to $|\psi|$ and still maintains the marginal constraint. The main tool will be the smoothing operator defined in Chapter 4. This will complete the proof of Theorem 5.1.3. In the final Section 5.2, making use of a suitable weighted Sobolev space, we show how to deal with the sign of the wave-function, finally proving the main result in its completeness.

Construction of L^2 wavefunctions

In this section we start the construction by proving the following

[3]The fact that V_{int} is continuous on the set of wave functions follows from the observation that $f(z) = \frac{1}{|z|}$ is such that $f \in L^\infty(\mathbb{R}^d)+L^{d/2}(\mathbb{R}^d)$, and the Hölder and Sobolev inequalities.

Theorem 5.1.5. *Let $\rho_n, \rho \in L^1(\mathbb{R}^d)$ such that $\sqrt{\rho_n} \to \sqrt{\rho}$ in $L^2(\mathbb{R}^d)$, $\int \rho = \int \rho_n = 1$ and let $\varphi \in L^2((\mathbb{R}^d)^N)$ symmetric, $\varphi \geq 0$, such that $\rho = \rho[\varphi]$. Then there exists a sequence $(\varphi_n) \subseteq L^2((\mathbb{R}^d)^N)$ such that φ_n is symmetric, $\rho_n = \rho[\varphi_n]$ and $\varphi_n \to \varphi$ in $L^2((\mathbb{R}^d)^N)$.*

For fixed $n \in \mathbb{N}$, let $\sigma_n^0 = \rho$ and $\varphi_n^0 = \varphi$ and define inductively for $k \geq 0$

$$\begin{aligned}
E_n^k &= \left\{ x \in \mathbb{R}^d \mid \sigma_n^k(x) > \rho_n(x) \right\}, \\
S_n^k(X) &= \frac{1}{N} \sum_{j=1}^{N} \frac{\sigma_n^k(x_j) - \rho_n(x_j)}{\sigma_n^k(x_j)} \chi_{E_n^k}(x_j), \\
\varphi_n^{k+1}(X) &= \varphi_n^k(X) \sqrt{1 - S_n^k(X)}, \\
\sigma_n^{k+1}(x) &= \int \varphi_n^{k+1}(x, x_2, \ldots, x_N)^2 dx_2 \cdots dx_N.
\end{aligned}$$

Notice that, for every k, n, the function φ_n^k is symmetric. The sequence $(\varphi_n^k)_{k \geq 0}$ is monotone decreasing, as proved in the following

Lemma 5.1.6. *(i) $0 \leq \varphi_n^{k+1} \leq \varphi_n^k \leq \varphi$;*

(ii) $0 \leq \sigma_n^{k+1} \leq \sigma_n^k \leq \rho$;

(iii) $E_n^{k+1} \subseteq E_n^k \subseteq E_n^0$.

Proof. Since $0 \leq S_n^k(X) \leq 1$, the factor $\sqrt{1 - S_n^k(X)}$ is less or equal than 1, and the inequalities in (i) are obvious; (ii) and (iii) follow. ☐

In order to estimate some L^2 norms which will appear later, the following lemma will also prove useful.

Lemma 5.1.7. *If $k \geq 0$ and $E \subseteq E_n^k$, then*

$$\int_E (\sigma_n^k(x) - \rho_n(x)) dx \leq \left(\frac{N-1}{N} \right)^k \int_E (\rho(x) - \rho_n(x)) dx.$$

Proof. By induction on k. For $k = 0$ the inequality is in fact an equality.

Suppose now the thesis is true for k, and fix $E \subseteq E_n^{k+1}$. Using the fact that $E_n^{k+1} \subseteq E_n^k$ one has

$$\begin{aligned}
\int_E (\sigma_n^{k+1}(x) - \rho_n(x)) dx &= \int_{E \times (\mathbb{R}^d)^{N-1}} \varphi_n^k(X)^2 (1 - S_n^k(X)) dX - \int_E \rho_n(x) dx \\
&= \int_E (\sigma_n^k(x) - \rho_n(x)) dx - \int_{E \times (\mathbb{R}^d)^{N-1}} \varphi_n^k(X)^2 S_n^k(X) dX.
\end{aligned}$$

Notice that

$$\begin{aligned}
&\int_{E\times(\mathbb{R}^d)^{N-1}} \varphi_n^k(X)^2 S_n^k(X)dX \\
&\qquad = \frac{1}{N}\sum_{j=1}^{N}\int_{E\times(\mathbb{R}^d)^{N-1}} \varphi_n^k(X)^2\frac{\sigma_n^k(x_j)-\rho_n(x_j)}{\sigma_n^k(x_j)}\chi_{E_n^k}(x_j)dX \\
&\qquad \geq \frac{1}{N}\int_{E\times(\mathbb{R}^d)^{N-1}} \varphi_n^k(X)^2\frac{\sigma_n^k(x_1)-\rho_n(x_1)}{\sigma_n^k(x_1)}\chi_{E_n^k}(x_1)dX \\
&\qquad = \frac{1}{N}\int_E (\sigma_n^k(x)-\rho_n(x))dx,
\end{aligned}$$

because the (first) marginal of φ_n^k is σ_n^k, and $E \subseteq E_n^k$. Hence, using the inductive hypothesis,

$$\begin{aligned}
\int_E (\sigma_n^{k+1}(x)-\rho_n(x))dx &\leq \left(1-\frac{1}{N}\right)\int_E (\sigma_n^k(x)-\rho_n(x))dx \\
&\leq \left(1-\frac{1}{N}\right)^{k+1}\int_E (\rho(x)-\rho_n(x))dx.
\end{aligned}$$

as wanted. □

The following proposition specifies that the sequence $(\varphi_n^k)_{k\geq 0}$ is not too far away from the target function φ with respect to the L^2 topology.

Proposition 5.1.8. *For every $k \geq 0$,*

$$\|\varphi\|^2_{L^2((\mathbb{R}^d)^N)} - \left\|\varphi_n^k\right\|^2_{L^2((\mathbb{R}^d)^N)} \leq 2N\left\|\sqrt{\rho}-\sqrt{\rho_n}\right\|_{L^2(\mathbb{R}^d)}.$$

Proof. We denote for simplicity the L^2-norm as $\|\cdot\|$ both on $(\mathbb{R}^d)^N$ and on $\mathbb{R}^d$, since there cannot be any confusion. By the definition of the φ_n^j's we may compute for every $j \geq 0$

$$\left\|\varphi_n^{j+1}\right\|^2 = \int \varphi_n^j(X)^2(1-S_n^j(X))dX = \left\|\varphi_n^j\right\|^2 - \int_{E_n^j}(\sigma_n^j(x)-\rho_n(x))dx.$$

Hence, using Lemma 5.1.7,

$$\begin{aligned}
\|\varphi\|^2 - \left\|\varphi_n^k\right\|^2 &= \sum_{j=0}^{k-1}\left(\left\|\varphi_n^j\right\|^2 - \left\|\varphi_n^{j+1}\right\|^2\right) = \sum_{j=0}^{k-1}\int_{E_n^j}(\sigma_n^j-\rho_n) \\
&\leq \sum_{j=0}^{k-1}\left(\frac{N-1}{N}\right)^j\int_{E_n^j}(\rho-\rho_n) = \sum_{j=0}^{k-1}\left(\frac{N-1}{N}\right)^j\int_{E_n^j}|\rho-\rho_n| \\
&\leq \sum_{j=0}^{k-1}\left(\frac{N-1}{N}\right)^j\int|\rho-\rho_n| \leq N\int|\rho-\rho_n|,
\end{aligned}$$

Now the Hölder inequality and the elementary estimate $(\sqrt{a}+\sqrt{b})^2 \leq 2(a+b)$ lead to

$$\int |\rho-\rho_n| \leq \left(\int |\sqrt{\rho}+\sqrt{\rho_n}|^2\right)^{1/2}\left(\int |\sqrt{\rho}-\sqrt{\rho_n}|^2\right)^{1/2}$$
$$= \left(2\int(\rho+\rho_n)\right)^{1/2}\|\sqrt{\rho}-\sqrt{\rho_n}\| = 2\,\|\sqrt{\rho}-\sqrt{\rho_n}\|\,.\square$$

We are ready to define the functions φ_n. Let

$$\varphi_n^\infty(X) = \lim_{k\to\infty}\varphi_n^k$$
$$\sigma_n^\infty(x) = \int \varphi_n^\infty(x,x_2,\dots,x_N)^2 dx_2\cdots dx_N.$$

They are well defined due to Lemma 5.1.6, and φ_n^∞ is symmetric, since it is the pointwise limit of symmetric functions; let moreover

$$q_n = \int (\rho_n(x)-\sigma_n^\infty(x))\,dx$$
$$\alpha_n(X) = \frac{1}{q_n^{N-1}}\prod_{j=1}^N(\rho_n(x_j)-\sigma_n^\infty(x_j))$$
$$\varphi_n(X) = \sqrt{\varphi_n^\infty(X)^2+\alpha_n(X)}.$$

where the second term is set to zero if $q_n = 0$. Observe that the function φ_n is symmetric, because α_n is symmetric by construction. The definition is well-posed since α_n is non-negative, as proved in the following

Lemma 5.1.9. $\rho_n(x)-\sigma_n^\infty(x) \geq 0.$

Proof. Using that $\frac{\sigma_n^k-\rho_n}{\sigma_n^k}\chi_{E_n^k} \geq 0$, one has

$$\rho_n(x)-\sigma_n^{k+1}(x) = \rho_n(x) - \int_{(\mathbb{R}^d)^{N-1}}\varphi_n^{k+1}(X)^2dx_2\cdots dx_N$$
$$= \rho_n(x)-\sigma_n^k(x) + \int_{\mathbb{R}^{(N-1)d}}\varphi_n^k(X)^2S_n^k(X)dx_2\cdots dx_N$$
$$\geq (\rho_n(x)-\sigma_n^k(x))\left(1-\frac{1}{N}\chi_{E_n^k}(x)\right)$$

If $x \in E_n^k$ then

$$(\rho_n(x)-\sigma_n^k(x))\left(1-\frac{1}{N}\chi_{E_n^k}(x)\right) = \frac{N-1}{N}(\rho_n(x)-\sigma_n^k(x));$$

on the other hand, if $x \in (E_n^k)^c$, then $\rho_n(x) - \sigma_n^k(x) \geq 0$, and hence

$$(\rho_n(x) - \sigma_n^k(x))\left(1 - \frac{1}{N}\chi_{E_n^k}(x)\right) = \rho_n(x) - \sigma_n^k(x) \geq \frac{N-1}{N}(\rho_n(x) - \sigma_n^k(x)).$$

So that for every $x \in \mathbb{R}^d$,

$$\rho_n(x) - \sigma_n^{k+1}(x) \geq \frac{N-1}{N}(\rho_n(x) - \sigma_n^k(x)),$$

and letting $k \to \infty$,

$$\rho_n(x) - \sigma_n^\infty(x) \geq \frac{N-1}{N}(\rho_n(x) - \sigma_n^\infty(x)) \implies \rho_n(x) - \sigma_n^\infty(x) \geq 0. \qquad \square$$

Finally, $\varphi_n \to \varphi$ in L^2 as n goes to ∞, as is proved in the following

Proposition 5.1.10.

$$\|\varphi_n - \varphi\|_{L^2}^2 \leq 2(2N+1)\left\|\sqrt{\rho} - \sqrt{\rho_n}\right\|_{L^2}$$

Proof. By the monotonicity described in Lemma 5.1.6, $\varphi \geq \varphi_n^\infty$ and then

$$\begin{aligned}|\varphi_n - \varphi|^2 &= \varphi^2 + (\varphi_n^\infty)^2 + \alpha_n - 2\varphi\sqrt{(\varphi_n^\infty)^2 + \alpha_n} \\ &\leq \varphi^2 + (\varphi_n^\infty)^2 + \alpha_n - 2\varphi\varphi_n^\infty \leq \varphi^2 - (\varphi_n^\infty)^2 + \alpha_n\end{aligned}$$

Integrating over $\mathbb{R}^d$ leads to

$$\|\varphi_n - \varphi\|_{L^2}^2 \leq \|\varphi\|_{L^2}^2 - \|\varphi_n^\infty\|_{L^2}^2 + \|\alpha_n\|_{L^1}.$$

Letting $k \to \infty$ in Proposition 5.1.8 and using the monotone convergence theorem, one has

$$\|\varphi\|_{L^2}^2 - \|\varphi_n^\infty\|_{L^2}^2 \leq 2N\left\|\sqrt{\rho} - \sqrt{\rho_n}\right\|_{L^2}.$$

On the other hand, recalling the final step of the proof of Proposition 5.1.8 and using again the monotone convergence theorem,

$$\begin{aligned}\|\alpha_n\|_{L^1} &= \int (\rho_n(x) - \sigma_n^\infty(x))\,dx \\ &\leq \int |\rho_n(x) - \rho(x)|\,dx + \int (\rho(x) - \sigma_n^\infty(x))\,dx \\ &\leq 2\left\|\sqrt{\rho} - \sqrt{\rho_n}\right\|_{L^2} + \int \varphi(X)^2 dX - \int \varphi_n^\infty(X)^2 dX \\ &= 2\left\|\sqrt{\rho} - \sqrt{\rho_n}\right\|_{L^2} + \|\varphi\|_{L^2}^2 - \|\varphi_n^\infty\|_{L^2}^2 \\ &\leq 2(N+1)\left\|\sqrt{\rho} - \sqrt{\rho_n}\right\|_{L^2}. \qquad \square\end{aligned}$$

This concludes the proof of Theorem 5.1.5.

Sobolev regularity and convergence

Let ρ_n, ρ, φ_n and φ as in Section 5.1 and assume, additionally, that $\varphi \in H^1$ and $\sqrt{\rho_n} \to \sqrt{\rho}$ in $H^1(\mathbb{R}^d)$. The sequence (φ_n) constructed in Section 5.1 is such that $\varphi_n \in L^2((\mathbb{R}^d)^N)$ with $\varphi_n \to \varphi$ in $L^2((\mathbb{R}^d)^N)$. We will now improve the regularity and the convergence of (φ_n) using the results of Chapter 4.

Given $\varphi \in L^2((\mathbb{R}^d)^N)$, we write for simplicity φ^ε for the square root of the density of the measure $\Theta^\varepsilon(\mu_\varphi)$, where $\mathrm{d}\mu_\varphi(X) = |\varphi(X)|^2 \,\mathrm{d}X$ and Θ is the smoothing operator of Theorem 4.0.1. We take in this case a Gaussian kernel, since we need Theorem 4.3.2. By the results of Chapter 4, $\Theta^\varepsilon(\varphi)$ is indeed absolutely continuous w.r.t. the Lebesgue measure, and $\varphi^\varepsilon \in H^1((\mathbb{R}^d)^N)$. Moreover, if φ is a symmetric function, *i.e.*, $\varphi(x_1, \dots, x_N) = \varphi(x_{\sigma(1)}, \dots, x_{\sigma(N)})$ for every permutation $\sigma \in \mathfrak{S}_N$, the construction of Chapter 4 produces a symmetric function φ^ε.

We are now able to prove the following

Theorem 5.1.11. *Let $\rho_n, \rho \in \mathcal{R}$ such that $\sqrt{\rho_n} \to \sqrt{\rho}$ in $H^1(\mathbb{R}^d)$, and let $\varphi \in H^1((\mathbb{R}^d)^N)$ symmetric and non-negative be such that $\rho[\varphi] = \rho$. Then there exist $u_n \in H^1((\mathbb{R}^d)^N)$ such that $u_n \to \varphi$ in $H^1((\mathbb{R}^d)^N)$ and $\rho[u_n] = \rho_n$.*

Proof. In this proof we denote by $\|\cdot\|$ the H^1-norm on $(\mathbb{R}^d)^N$. Let φ_n, φ defined in Section 5.1: the idea is to take a suitable diagonal sequence $u_n := \varphi_n^{\varepsilon(n)}$. Let $N_0 = 1$, and for $k \geq 1$ choose $N_k \in \mathbb{N}$ such that

(i) $N_k > N_{k-1}$;

(ii) $\left\|\varphi_n^{2^{-k}} - \varphi^{2^{-k}}\right\| \leq 2^{-k}$ for every $n \geq N_k$.

The sequence $(N_k)_{k\geq 0}$ is well defined due to Theorem 4.3.2.(iii) and increasing. Given $n \geq 1$, let k be such that $N_k \leq n < N_{k+1}$, and set $\varepsilon(n) = 2^{-k}$. When $N_k \leq n < N_{k+1}$, by construction we have

$$\left\|\varphi_n^{\varepsilon(n)} - \varphi\right\| \leq \left\|\varphi_n^{2^{-k}} - \varphi^{2^{-k}}\right\| + \left\|\varphi^{2^{-k}} - \varphi\right\| \leq 2^{-k} + \left\|\varphi^{2^{-k}} - \varphi\right\|.$$

As $n \to \infty$, also $k \to \infty$ and the right-hand side goes to zero due to Theorem 4.2.7. $\square$

To avoid any confusion, in the following we will denote by $(\varphi_n)_{n\in\mathbb{N}}$ a sequence such that $\varphi_n \to \varphi$ in $H^1((\mathbb{R}^d)^N)$ and $\rho[\varphi_n] = \rho_n$.

Remark 16. If the original wave-function was symmetric and non-negative, then by Theorem 5.1.11 we already get the desired approximating wave-functions which are also symmetric and non-negative, thus proving Theorem 5.1.3.

5.2 Approximation with signs

Let ρ_n, ρ, φ_n and φ like in the previous sections, and assume now that $\varphi = |\psi|$, where $\psi \in H^1((\mathbb{R}^d)^N;\mathbb{R})$. In this section, starting from φ_n, we will construct $\psi_n \in H^1((\mathbb{R}^d)^N;\mathbb{C})$ such that $\psi_n \mapsto \rho_n$ and $\psi_n \overset{H^1}{\to} \psi$. Some weighted Sobolev spaces will be the main tool of the construction.

To every measurable $\lambda : (\mathbb{R}^d)^N \to \mathbb{R}^n$ (scalar or vectorial) we may associate some spaces related to the measure $|\lambda|^2(X)dX$. In particular this will be used for λ equal to the wave function ψ or equal to the gradient of the wave function $\nabla\psi$. The most natural space is

$$L^2(|\lambda|^2 dX;\mathbb{C}) := \left\{ f\colon (\mathbb{R}^d)^N \to \mathbb{C} \mid \int |f(X)|^2|\lambda(X)|^2 dX < +\infty \right\}.$$

When $\lambda \in H^1((\mathbb{R}^d)^N;\mathbb{R})$ we may define also the Sobolev spaces relative to the measure $|\lambda|^2 dX = \lambda^2 dX$. First we need a definition of the gradient:

Definition 8. If $\lambda \in H^1((\mathbb{R}^d)^N;\mathbb{R})$ and $f \in L^2(\lambda^2 dX)$, the gradient $\nabla_\lambda f$ is defined by the identity

$$\int \nabla_\lambda f \varphi \lambda^2 dX = -\int f \nabla\varphi \lambda^2 dX - 2\int f\varphi \frac{\nabla\lambda}{\lambda}\lambda^2 dX \quad \forall \varphi \in \mathcal{C}_c^\infty. \tag{5.2}$$

It is then natural to define the Sobolev space

$$H^1(\lambda^2 dX) := \left\{ f \in L^2(\lambda^2 dX)\colon \int |\nabla_\lambda f(X)|^2 \lambda^2 dX < +\infty \right\}.$$

Remark 17. If f is a $\mathcal{C}^1$ function then $\nabla_\lambda f = \nabla f$, the usual gradient. Indeed, if $\varphi \in \mathcal{C}_c^\infty$, then

$$\begin{aligned}
\int \nabla_\lambda f \varphi \lambda^2 dX &= -\int f \nabla\varphi \lambda^2 dX - 2\int f\varphi \frac{\nabla\lambda}{\lambda}\lambda^2 dX \\
&= \int \nabla f \varphi \lambda^2 dX + 2\int f\varphi \frac{\nabla\lambda}{\lambda}\lambda^2 dX - 2\int f\varphi \frac{\nabla\lambda}{\lambda}\lambda^2 dX \\
&= \int \nabla f \varphi \lambda^2 dX.
\end{aligned}$$

Remark 18. If $f, g \in H^1(\lambda^2 dX)$, and $fg, f\nabla_\lambda g, g\nabla_\lambda f \in L^2(\lambda^2 dX)$, then $fg \in H^1(\lambda^2 dX)$ and $\nabla_\lambda(fg) = f\nabla_\lambda g + g\nabla_\lambda f$. This is Corollary 2.6 in [11].

The next construction relates these ideas to the objects we know from the previous sections. Let $\psi \in H^1((\mathbb{R}^d)^N;\mathbb{R})$ so that also $|\psi| \in H^1((\mathbb{R}^d)^N)$; then there exists a measurable function $e : (\mathbb{R}^d)^N \to \{-1, 1\}$ such that $\psi = e|\psi|$. The function e coincides almost everywhere with $\psi/|\psi|$ in the set where $\psi \neq 0$. From now on, let $\lambda = |\psi|$.

Lemma 5.2.1. *It holds that* $e \in H^1(|\psi|^2 dX)$ *and* $\nabla_\lambda e = 0$ $|\psi|^2 dX - a.e.$ *Moreover, since* $|e| \leq 1$, $e \in L^2(|\nabla\psi|^2 dX)$.

Proof. Since $|e| = 1$ $|\psi|^2 dX - a.e.$, $e \in L^2(|\psi|^2 dX)$. Let $\varphi \in \mathcal{C}_c^\infty$,

$$\int \nabla_\lambda e \varphi |\psi|^2 dX = -\int e \nabla\varphi |\psi|^2 dX - 2\int e\varphi|\psi|\frac{\psi}{|\psi|}\nabla\psi dX = 0,$$

since

$$\begin{aligned}\int e\nabla\varphi|\psi|^2 dX &= \int \nabla\varphi\psi|\psi| dX = -\int \varphi\nabla\psi|\psi| dX - \int \varphi\psi\frac{\psi}{|\psi|}\nabla\psi dX \\ &= -2\int \varphi|\psi|\nabla\psi dX.\end{aligned}$$

□

We are interested in smooth approximations in these Sobolev spaces, a well-studied question in the literature. The following is a consequence of [11, Theorem 2.7]; we invite the reader to see also the references in that paper for a more complete picture.

Theorem 5.2.2. *There exists a sequence* $\{e_n\} \in \mathcal{C}^\infty \cap H^1(|\psi|^2 dX)$ *such that*

(i) $|e_n| \leq 1$,

(ii) $e_n \to e$ *in* $H^1(|\psi|^2 dX)$,

(iii) $e_n \to e$ *in* $L^2(|\nabla\psi|^2 dX)$.

Proof. Choose a sequence of smooth cut-off functions $(c_n)_{n\geq 1}$ such that:

- $0 \leq c_n \leq 1$;
- $c_n \equiv 1$ on $B(0, n-1)$, $c_n \equiv 0$ on $B(0,n)^c$;
- $\mathrm{Lip}(c_n) \leq 2$.

First we consider $e \cdot c_n$ and we prove that it satisfies the properties (i)-(iii) above. The property (i) is obvious. Also the L^2-convergence is easy:

$$\int |ec_n - e|\,\lambda^2 dX \leq \int_{B(0,n)^c} \lambda^2 dX, \quad \int |ec_n - e|\,|\nabla\lambda|^2\,dX \leq \int_{B(0,n)^c} |\nabla\lambda|^2\,dX$$

converge to zero since $\lambda, \nabla\lambda \in L^2(\mathbb{R}^{Nd})$. Combining Remark 17 and Remark 18 we have $\nabla_\lambda(ec_n) = c_n\nabla_\lambda e + e\nabla_\lambda c_n = e\nabla_\lambda c_n = e\nabla c_n$, and we must prove that it converge to 0 in $L^2(\lambda^2 dX)$. Indeed we have

$$\int |e\nabla c_n|^2\,\lambda^2 dX \leq 4\int_{B(0,n-1)^c} \lambda^2 dX.$$

Now the second step is to regularize by convolution with a standard mollifier of compact support J_ε defined by $J_\varepsilon(X) = 1/\varepsilon^{Nd} J(X/\varepsilon)$, where J is non-negative and supported in the unit ball, with $\int J = 1$. It is shown in [11, Theorem 2.7] that $J_\varepsilon * (ec_n)$ converge to ec_n for fixed n as $\varepsilon \to 0$. Thus it suffices to take ε_n small enough so that $\|J_{\varepsilon_n} * (ec_n) - ec_n\|_{H^1(\lambda^2 dX)}$ converges to zero to conclude. □

Remark 19. If the function e is symmetric, it is possible to make e_n to be symmetric as well. It suffices to choose c_n (the cut-off functions) to be symmetric. Then the process of convolution maintains symmetry if the kernel is symmetric.

In order to have a good behaviour of the approximating sequence $(e_n)_{n\geq 1}$ for the estimates that will be needed in the proof of Theorem 5.2.5, we must also control the Lipschitz constant of e_n. This may be done as a consequence of the following

Lemma 5.2.3. *Given sequences of non-negative real numbers (M_n) and (a_k) such that $a_k \to 0$, there exists a choice (n_k) of indexes such that*

(i) $n_k \nearrow +\infty$;

(ii) $M_{n_k} a_k \to 0$.

Proof. Given n, let $K(n)$ such that $M_n a_k < 2^{-n}$ for all $k \geq K(n)$, and choose also $K(n+1) > K(n)$. Now we define the sequence (n_k) as follows:

$$n_k = \begin{cases} 1 & \text{if } k < K(1) \\ n & \text{if } K(n) \leq k < K(n+1). \end{cases}$$

By construction we have $M_{n_k} a_k < 2^{-n}$ for all $k \geq K(n)$, thus proving (ii). On the other hand, given $L \in \mathbb{N}$, if $k \geq K(L)$ we have $n_k \geq L$, which proves (i). □

Corollary 5.2.4. *Given (a_n) such that $a_n \to 0$, the sequence in Theorem 5.2.2 may be chosen such that* $\mathrm{Lip}(e_n)a_n \to 0$.

Proof. Apply Lemma 5.2.3 with $M_n = \mathrm{Lip}(e_n)$ to select a suitable sequence (e_{n_k}) with the desired property. □

Definition 9. Let $\omega \in \mathcal{C}^1([-1,1], S^1_+)$ be defined by

$$s \mapsto e^{i(1-s)\frac{\pi}{2}}.$$

The function ω is such that $|\omega| = 1$, $\omega(-1) = -1$ and $\omega(1) = 1$ so that $\omega(e(x)) = e(x)$ a.e. in the set $\psi \neq 0$. Moreover, observe that $|\omega'| = \frac{\pi}{2}$ and $|\omega(s) - \omega(t)| \leq \frac{\pi}{2}|s-t|$ for all $s, t \in [-1,1]$.

Theorem 5.2.5. *Let ω be the function defined above. Let $e_n \in \mathcal{C}^\infty_c$ with values in $[-1,1]$ be such that $e_n \to e$ in $H^1(|\psi|^2 dX)$ and $L^2(|\nabla\psi|^2 dX)$. Then $\psi_n := \omega(e_n)\varphi_n \to \psi$ in H^1.*

Proof. First the L^2 convergence which is easier.

$$\begin{aligned}
\|\psi_n - \psi\|_{L^2} &= \|\omega(e_n)\varphi_n - e|\psi|\|_{L^2} \\
&\le \|\omega(e_n)\varphi_n - \omega(e_n)|\psi|\|_{L^2} + \|\omega(e_n)|\psi| - e|\psi|\|_{L^2} \\
&= \|\varphi_n - |\psi|\|_{L^2} + \|\omega(e_n) - e\|_{L^2(|\psi|^2dX)} \\
&\le \|\varphi_n - |\psi|\|_{L^2} + \|\omega(e_n) - \omega(e)\|_{L^2(|\psi|^2dX)} + \|\omega(e) - e\|_{L^2(|\psi|^2dX)} \\
&\le \|\varphi_n - |\psi|\|_{L^2} + \frac{\pi}{2}\|e_n - e\|_{L^2(|\psi|^2dX)}.
\end{aligned}$$

The last term converges to 0 by Theorem 5.2.2 above.

For the L^2 convergence of gradients, let us first compute

$$\nabla\psi_n = \omega'(e_n)\nabla e_n\varphi_n + \omega(e_n)\nabla\varphi_n,$$

$$\nabla\psi = \nabla(e|\psi|) = e\nabla|\psi|,$$

and in the second computation we used that $\nabla e = 0$ a.e. where $|\psi| \neq 0$.

$$\begin{aligned}
\|\nabla\psi_n - \nabla\psi\|_{L^2} &= \|\omega'(e_n)\nabla e_n\varphi_n + \omega(e_n)\nabla\varphi_n - e\nabla|\psi|\| \\
&\le \|\omega'(e_n)\nabla e_n\varphi_n - \omega'(e_n)\nabla e_n|\psi|\| + \|\omega'(e_n)\nabla e_n|\psi|\| \\
&\quad + \|\omega(e_n)\nabla\varphi_n - e\nabla|\psi|\|.
\end{aligned}$$

The three terms on the right-hand-side above may be studied separately, the most difficult one being the first. We have

$$\begin{aligned}
\left\|\omega'(e_n)\nabla e_n\varphi_n - \omega'(e_n)\nabla e_n|\psi|\right\|^2 &\le \frac{\pi}{2}\int |\nabla e_n|^2 \left|\varphi_n - |\psi|\right|^2 dX \\
&\le \frac{\pi}{2}\operatorname{Lip}(e_n)^2 \int \left|\varphi_n - |\psi|\right|^2 dX.
\end{aligned}$$

The last term of the inequality converges to 0 if we choose $a_n = \|\varphi_n - |\psi|\|_{L^2}$ in Corollary 5.2.4.

The second term

$$\left\|\omega'(e_n)\nabla e_n\,|\psi|\right\|^2 \le \frac{\pi}{2}\int |\nabla e_n|^2\,|\psi|^2\,dX$$

and this goes to 0 by Theorem 5.2.2 ii) and Lemma 5.2.1. Finally we control the third term by breaking it down again.

$$\begin{aligned}
\|\omega(e_n)\nabla\varphi_n - e\nabla\,|\psi|\|_{L^2} &\le \|\omega(e_n)\nabla\varphi_n - \omega(e_n)\nabla\,|\psi|\|_{L^2} \\
&\quad + \|\omega(e_n)\nabla\,|\psi| - e\nabla\,|\psi|\|_{L^2} \\
&\le \|\nabla\varphi_n - \nabla\,|\psi|\|_{L^2} + \|\omega(e_n)\nabla\,|\psi| - \omega(e)\nabla\,|\psi|\|_{L^2} \\
&\le \|\nabla\varphi_n - \nabla\,|\psi|\|_{L^2} + \frac{\pi}{2}\,\|e_n - e\|_{L^2(|\nabla\psi|^2dX)}.
\end{aligned}$$

The last term converges to 0 by the convergence of φ_n to $|\psi|$ in $H^1((\mathbb{R}^d)^N)$ and by Theorem 5.2.2 iii). $\square$

In conclusion, notice that the approximating sequence built in this way maintains the symmetry property. Indeed, φ_n is symmetric for every n, and so is the sign function e. By Remark 19 we may choose e_n to be symmetric. Finally, if e_n is symmetric, so is $\omega(e_n)$, and hence $\omega(e_n)\varphi_n$ is symmetric, finally proving Theorem 5.1.4.

5.3 The semiclassical limit of the Hohenberg-Kohn functional

In this Section we address the following problem: what is the behaviour of $F_\hbar^{\mathcal{A}}$ and $F_\hbar^{\mathcal{S}}$ as $\hbar \to 0$?

This is a very relevant topic in physics, and is called the "semiclassical limit". It corresponds to the case of "strongly correlated electrons", *i.e.*, to a system in which the internal potential energy V_{int} due to the interaction between particles dominates the kinetic energy T.

Let us start with a crucial remark.

Remark 20. If we put $\hbar = 0$ in the definition of $F_\hbar^{\mathcal{S}}$ we get

$$F_0^{\mathcal{S}} := \inf \left\{ V_{int}(\psi) \mid \psi \in \mathcal{S}, \psi \mapsto \rho \right\}.$$

Recall the Kantorovich formulation of the multi-marginal optimal transport problem for a cost function $c(x_1, \ldots, x_N)$ with all marginals equal to ρ:

$$C(\rho) = \min \left\{ \int c(X) d\gamma(X) \mid \gamma \in \Pi(\rho, \ldots, \rho) \right\}.$$

By considering the Coulomb cost function

$$c(x_1, \ldots, x_N) = \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|}$$

and by viewing $|\psi|^2$ as the density of a probability measure γ on $(\mathbb{R}^d)^N$, we get

$$\begin{aligned} F_0^{\mathcal{S}} &= \inf \left\{ \int c(X) d\gamma(X) \mid \gamma \in \Pi(\rho, \ldots, \rho), \sqrt{\gamma} \in H^1((\mathbb{R}^d)^N) \right\} \\ &\ge \inf \left\{ \int c(X) d\gamma(X) \mid \gamma \in \Pi(\rho, \ldots, \rho) \right\} = C(\rho). \end{aligned}$$

Thus, the two functionals look almost the same, except for the regularity request on the transport plan. However, as we studied in detail in Chapter 4, every transport plan can be approximated by a regular one, so the two functionals are actually the same.[4]

[4] This still requires some work, since the Coulomb cost is not a continuous and bounded function. Details will be made precise in the following.

This raises the question of whether $F_\hbar^{\mathcal{A}}, F_\hbar^{\mathcal{S}} \to C$ in a suitable sense of convergence. This connection between the Levy-Lieb functional and the multi-marginal optimal transport functional was not realized until recent years — see for instance [10, 18]. Since then, the semiclassical limit of the Levy-Lieb functional was studied in [4, 32, 19], and finally now we have a complete picture. We will devote the rest of this section to the proof of the following results.

Theorem 5.3.1. *For every $\rho \in \mathcal{R}$,*

$$\lim_{\hbar \to 0} F_\hbar^{\mathcal{S}}(\rho) = C(\rho).$$

Theorem 5.3.2. *For every $\rho \in \mathcal{R}$,*

$$\lim_{\hbar \to 0} F_\hbar^{\mathcal{S}}(\rho) = C(\rho).$$

For the proof of Theorem 5.3.1 we will closely follow [4], and is based on our smoothing procedure of Chapter 4.

A proof of Theorem 5.3.2 is given in [4] for $N = 2, 3$, and a very similar idea was later developed in [19] to get the same result for any number of marginals. However, in our opinion, the proof of Theorem 5.3.2 given by M. Lewin in [32] is much more physically meaningful and elegant. Moreover, it exploits once more the ideas coming from Chapter 4, and this give everything a nice unique frame. For the main estimates, however, we will refer to the original paper, because we want to focus on the link between Lewin's contruction and Chapter 4.

For both theorem, we will prove the result in the spinless case. The spin dependence can be handled easily, *e.g.*, by letting all the particles in the same spin state with probability one.

Without further ado, let us move to the proofs. For clarity of presentation, we separate the argument in two sections: the first one will show how to deal with the bosonic case. In the second part, we will present how Lewin generalized this construction to mixed quantum states to get Theorem 5.3.2.

Proof of Theorem 5.3.1

Let us fix $\rho \in \mathcal{R}$. For every $\psi \in \mathcal{S}$, $\psi \mapsto \rho$ we have

$$\hbar^2 T(\psi) + V_{int}(\psi) \geq V_{int}(\psi) \geq C(\rho),$$

since the transport plan μ defined by $d\mu(X) = |\psi(X)|^2 \, dX$ is an admissible competitor for the multi-marginal optimal transport problem $C(\rho)$.

By passing to the infimum on the left-hand side we get

$$F_\hbar^{\mathcal{S}}(\rho) \geq C(\rho),$$

hence

$$\liminf_{\hbar\to 0} F_\hbar^{\mathcal{S}}(\rho) \geq C(\rho).$$

Let now $\mu \in \mathcal{P}((\mathbb{R}^d)^N)$ be an optimal transport plan for $C(\rho)$, *i.e.*,

$$\int \sum_{1\leq i<j\leq N} \frac{d\mu(X)}{|x_i - x_j|} = C(\rho).$$

First of all, notice that we can assume μ to be a symmetric measure: indeed, if it is not the case, let

$$\operatorname{Sym}\mu := \frac{1}{N!}\sum_{\sigma\in\mathfrak{S}_N} (T_\sigma)_\#(\mu),$$

where $T_\sigma\colon (\mathbb{R}^d)^N \to (\mathbb{R}^d)^N$ is the permutation of coordinates defined by $T_\sigma(x_1,\ldots,x_N) = (x_{\sigma(1)},\ldots,x_{\sigma(N)})$. Because of the symmetry of the cost function, $\operatorname{Sym}\mu$ will still be an optimal transport plan for $C(\rho)$.

Now the idea of the proof is to let

$$\psi_\hbar := \sqrt{\Theta^{\sqrt{\hbar}}[\mu]},$$

where Θ is the smoothing operator defined in Chapter 4. In this case we take a compactly supported kernel η, since we need to apply Theorem 4.3.1. Observe that this defines $\psi_\hbar \in H^1((\mathbb{R}^d)^N)$, symmetric and positive.

Lemma 5.3.3. *With the notation above,*

$$\lim_{\hbar\to 0} \hbar^2 T(\psi_\hbar) = 0.$$

Proof. We apply the $p = 2$ version of the energy bound for Θ given by Theorem 4.2.4 to get

$$T(\psi_\hbar) = \mathcal{E}_2(\Theta^\hbar[\mu]) \leq \frac{Nc(\eta)}{\hbar} + N\int_{\mathbb{R}^d} |\nabla\sqrt{\rho}|^2 .$$

Since $\rho \in \mathcal{R}$, the last integral is finite, and we get the thesis. □

Lemma 5.3.4. *With the notation above,*

$$\lim_{\hbar\to 0} V_{int}(\psi_\hbar) = C(\rho).$$

Proof. By a result of Buttazzo et al. [9, Theorem 2.4], there exists $\alpha > 0$ such that

$$\mu(D_{2\alpha}) = 0,$$

where $D_{2\alpha}$ denotes the "enlarged diagonal"

$$D_{2\alpha} = \left\{(x_1,\ldots,x_N) \in (\mathbb{R}^d)^N \mid |x_i - x_j| \leq 2\alpha \text{ for some } i \neq j\right\}.$$

We define the truncated Coulomb cost

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

where

$$f_\alpha(x,y) = \begin{cases} \frac{1}{|x-y|} & \text{if } |x-y| \ge \alpha \\ \frac{1}{\alpha} & \text{otherwise.} \end{cases}$$

Observe that c_α is a bounded and continuous function which differ from the Coulomb cost only on $D_\alpha \subseteq D_{2\alpha}$, hence

$$C(\rho) = \int c(X) d\mu(X) = \int c_\alpha(X) d\mu(X).$$

By Corollary 4.2.6, we know that $\Theta^{\sqrt{\hbar}}[\mu] \rightharpoonup \mu$, so in particular

$$\lim_{\hbar \to 0} \int c_\alpha(X) \left|\psi_\hbar(X)\right|^2 dX = \lim_{\hbar \to 0} \int c_\alpha(X) d\Theta^{\sqrt{\hbar}}[\mu](X) = \int c_\alpha(X) d\mu(X) = C(\rho).$$

On the other hand by Theorem 4.3.1 with $\Omega = D_\alpha$, we know that, for $\hbar$ sufficiently small, $\psi_\hbar = 0$ on D_α, hence

$$V_{int}(\psi_\hbar) = \int c(X) \left|\psi_\hbar(X)\right|^2 dX = \int c_\alpha(X) \left|\psi_\hbar(X)\right|^2 dX,$$

and we get the thesis. □

This concludes the proof of Theorem 5.3.1, since we get

$$\limsup_{\hbar \to 0} F_\hbar^{\mathcal{S}}(\rho) \le \limsup_{\hbar \to 0} \left(\hbar^2 T(\psi_\hbar) + V_{int}(\psi_\hbar)\right) = C(\rho).$$

Proof of Theorem 5.3.2

We get easily that

$$\liminf_{\hbar \to 0} F_\hbar^{\mathcal{S}}(\rho) = C(\rho)$$

as in the proof of Theorem 5.3.1.

In order to get the (much harder) inequality, we give a brief quantum-oriented introduction and notation, in order to clarify the similarities of Lewin's argument with our construction. As in quantum chemistry is often done, we extend the set $\mathcal{A}$ of wave-function to the set of N-particles density matrix, as for the following

Definition 10. The set of *N-particles density matrices* is the set of linear operators $\Gamma \colon \mathcal{A} \to \mathcal{A}$ such that:

(i) Γ is trace-class, $\operatorname{Tr} \Gamma = 1$;

(ii) $\Gamma = \Gamma^*$, $\Gamma \geq 0$.

The set of N-particles density matrices is convex. Observe also that any $\psi \in \mathcal{A}$ provides an N-density matrix Γ_ψ given by the orthogonal projection on the linear subspace generated by ψ., so the notion of N-particles density matrix effectively extends that of wave-function. To any given Γ, we can associate a single-particle density ρ_Γ defined by duality by

$$\int_{\mathbb{R}^d} \phi(x) d\rho(x) = \mathrm{Tr}(\phi(x_1)\Gamma)$$

for every $\phi \in C_b(\mathbb{R}^d)$. Notice that, consistently, if $\Gamma = \Gamma_\psi$, then the density ρ_Γ defined in this way coincides with the single particle density of the wave-function ψ.

A crucial result in Density Functional Theory is the following by H. Lieb [34, Section 4.B].

Theorem 5.3.5. *Let $\rho \in \mathcal{R}$. Then*

$$F_\hbar^{\mathcal{A}}(\rho) = \inf \{\mathrm{Tr}\, H_0\Gamma \mid \Gamma \text{ } N\textit{-particle density matrix, } \rho_\Gamma = \rho\}$$

Here H_0 denotes the fundamental hamiltonian

$$H_0 = -\sum_{k=1}^{N} \Delta_{x_k} + c(x_1, \ldots, x_N),$$

which is a non-negative self-adjoint operator on $\mathcal{A}$, since the Coulomb cost function $c(x_1, \ldots, x_N)$ is non-negative.[5]

Let us take a symmetric optimal transport plan $\mu \in \mathcal{P}((\mathbb{R}^d)^N)$ for the multi-marginal optimal transport problem. The idea to complete the proof of Theorem 5.3.2 is thus to construct a suitable sequence $\Gamma_\hbar$ such that:

(A) $\mathrm{Tr}(-\Delta_{x_k})\Gamma_\hbar$ is controlled in terms of $\hbar$ for every $k = 1, \ldots, N$;

(B) $\Gamma_\hbar \rightharpoonup \mu$ and $\mathrm{supp}\, \Gamma_\hbar \cap D_\alpha = \emptyset$ for $\hbar$ sufficiently small.

This will allow to conclude as in the previous section.

Lewin's nice idea is to modify the smoothing operator of Chapter 4 in order to get *not* a wave-function, *but* an N-particle denisty matrix. We take a compactly supporte kernel $\eta_\varepsilon(z) \in C^\infty(\mathbb{R}^d)$, and we start as usual by regularizing μ by convolution. Let as in Chapter 4

$$\Lambda^\varepsilon[\mu](Y) = \int \prod_{k=1}^{N} \eta^\varepsilon(y_k - z_k)\, \mathrm{d}\mu(Z).$$

[5] Here the Laplacian is naturally defined by duality using the Sobolev derivatives.

Recall that we got back the original marginals by letting

$$\Theta^\varepsilon[\mu](X) = \int \prod_{k=1}^{N} \frac{\eta^\varepsilon(y_k - x_k)\rho(x_k)}{(\rho * \eta^\varepsilon)(y_k)} \Lambda^\varepsilon[\mu](Y)\, \mathrm{d}Y,$$

and in this way we get a symmetric (bosonic) wave-function, because we are averaging on Y the symmetric wave-functions

$$\psi_Y(X) = \prod_{k=1}^{N} \frac{\eta^\varepsilon(y_k - x_k)\rho(x_k)}{(\rho * \eta^\varepsilon)(y_k)}.$$

If we want to get an N-particle density matrix, we should average on Y suitable operators Γ_Y. A standard way to construct fermionic wave functions is to take the so-called *Slater determinant*: given $\varphi_1, \dots, \varphi_N \in H^1(\mathbb{R}^d)$, the wave-function

$$S(\varphi_1, \dots, \varphi_N)(x_1, \dots, x_N) := \det \begin{pmatrix} \varphi_1(x_1) & \dots & \varphi_1(x_N) \\ \vdots & \ddots & \vdots \\ \varphi_N(x_1) & \dots & \varphi_N(x_N) \end{pmatrix}$$

belongs to the set $\mathcal{A}$.

Recalling that to each wave-function ψ we can associate $\Gamma = \Gamma_\psi$, we define the wave-function

$$\psi^{\varepsilon,Y}(x_1, \dots, x_N) := \det \begin{pmatrix} \frac{\eta^\varepsilon(y_1 - x_1)\rho(x_1)}{(\rho*\eta^\varepsilon)(y_1)} & \dots & \frac{\eta^\varepsilon(y_1 - x_N)\rho(x_N)}{(\rho*\eta^\varepsilon)(y_1)} \\ \vdots & \ddots & \vdots \\ \frac{\eta^\varepsilon(y_N - x_1)\rho(x_1)}{(\rho*\eta^\varepsilon)(y_N)} & \dots & \frac{\eta^\varepsilon(y_N - x_N)\rho(x_N)}{(\rho*\eta^\varepsilon)(y_N)} \end{pmatrix}$$

and let $\Gamma^\varepsilon_Y := \Gamma_{\psi^{\varepsilon,Y}}$. When we average over Y, since $\Lambda^\varepsilon[\mu]$ is a probability measure, by convexity of the N-particles density matrices we get an admissible trial state

$$\Gamma^\varepsilon := \int \Gamma^\varepsilon_Y \Lambda^\varepsilon[\mu](Y)\, \mathrm{d}Y.$$

For the properties of Γ^ε, which lead to the result and also to an additional bound on the rate of convergence,we refer to [32] and in particular Theorem 1 therein.

Bibliography

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Springer Science & Business Media, 2008.

[2] Ugo Bindini. Marginals with finite repulsive cost. *Canadian Journal of Mathematics*, 72:373–391, 2018.

[3] Ugo Bindini. Smoothing operators in multi-marginal optimal transport. *Mathematical Physics, Analysis and Geometry*, 23:1–27, 2020.

[4] Ugo Bindini and Luigi De Pascale. Optimal transport with coulomb cost and the semiclassical limit of density functional theory. *Journal de l'École polytechnique*, 4:909–934, 2017.

[5] Ugo Bindini and Luigi De Pascale. From wave-functions to single electron densities. *arXiv preprint arXiv:1907.02024*, 2019.

[6] Guy Bouchitté, Giuseppe Buttazzo, Thierry Champion, and Luigi De Pascale. Relaxed multi-marginal costs and quantization effects. *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 2020.

[7] Andrea Braides. *Gamma-convergence for Beginners.* Oxford University Press, 2002.

[8] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[9] Giuseppe Buttazzo, Thierry Champion, and Luigi De Pascale. Continuity and estimates for multimarginal optimal transportation problems with singular costs. *Applied Mathematics & Optimization*, 78(1):185–200, 2018.

[10] Giuseppe Buttazzo, Luigi De Pascale, and Paola Gori-Giorgi. Optimal-transport formulation of electronic density-functional theory. *Physical Review A*, 85(6):062502, 2012.

[11] Patrick Cattiaux and Myriam Fradon. Entropy, reversible diffusion processes, and Markov uniqueness. *journal of functional analysis*, 138(1):243–272, 1996.

[12] James A Clarkson. Uniformly convex spaces. *Transactions of the American Mathematical Society*, 40(3):396–414, 1936.

[13] Maria Colombo, Luigi De Pascale, and Simone Di Marino. Multimarginal optimal transport maps for 1-dimensional repulsive costs. *Canad. J. Math*, 67:350–368, 2013.

[14] 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.

[15] Maria Colombo and Simone Di Marino. Equality between monge and kantorovich multimarginal problems with coulomb cost. *Annali di Matematica Pura ed Applicata (1923-)*, 194(2):307–320, 2015.

[16] 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.

[17] 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.

[18] 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.

[19] Codina Cotar, Gero Friesecke, and Claudia Klüppelberg. Smoothing of transport plans with fixed marginals and rigorous semiclassical limit of the hohenberg–kohn functional. *Archive for Rational Mechanics and Analysis*, 228(3):891–922, 2018.

[20] Codina Cotar and Mircea Petrache. Equality of the jellium and uniform electron gas next-order asymptotic terms for coulomb and riesz potentials. *arXiv preprint arXiv:1707.07664*, 2017.

[21] Codina Cotar and Mircea Petrache. Next-order asymptotic expansion for n-marginal optimal transport with coulomb and riesz costs. *Advances in Mathematics*, 344:137–233, 2019.

[22] Gianni Dal Maso. *An introduction to* Γ*-convergence*, volume 8. Springer Science & Business Media, 2012.

[23] Luigi De Pascale. Optimal transport with coulomb cost. approximation and duality. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1643–1657, 2015.

[24] Luigi De Pascale. On c-cyclical monotonicity for optimal transport problem with coulomb cost. *European Journal of Applied Mathematics*, 30(6):1210–1219, 2019.

[25] Simone Di Marino, Augusto Gerolin, and Luca Nenna. Optimal transportation theory with repulsive costs. *Topological optimization and optimal transport*, 17:204–256, 2015.

[26] Persi Diaconis and David Freedman. Finite exchangeable sequences. *The Annals of Probability*, pages 745–764, 1980.

[27] Reiner M. Dreizler and Eberhard K.U. Gross. *Density Functional Theory.* Springer-Verlag Berlin Heidelberg, 1990.

[28] Wilfrid Gangbo and Andrzej Święch. Optimal maps for the multidimensional Monge-Kantorovich problem. *Communications on Pure and Applied Mathematics*, 51(1):23–45, 1998.

[29] Nassif Ghoussoub and Bernard Maurey. Remarks on multi-marginal symmetric Monge-Kantorovich problems. *Discrete and Continuous Dynamical Systems - A*, 34, 2014.

[30] Nassif Ghoussoub and Abbas Moameni. Symmetric Monge-Kantorovich problems and polar decompositions of vector fields. *Geometric and Functional Analysis*, 24(4):1129–1166, 2014.

[31] Young-Heon Kim and Brendan Pass. A general condition for Monge solutions in the multi-marginal optimal transport problem. *SIAM Journal on Mathematical Analysis*, 46(2):1538–1550, 2014.

[32] Mathieu Lewin. Semi-classical limit of the Levy-Lieb functional in density functional theory. *Comptes Rendus Mathematique*, 356(4):449–455, 2018.

[33] Mathieu Lewin, Paola Gori-Giorgi, and Brendan Pass. Optimal transport methods in density functional theory. http://www.birs.ca/workshops/2019/19w5035/report19w5035.pdf, 2019.

[34] Elliott H. Lieb. Density functionals for Coulomb systems. In *Inequalities*, pages 269–303. Springer, 2002.

[35] Brendan Pass. Uniqueness and Monge solutions in the multimarginal optimal transportation problem. *SIAM Journal on Mathematical Analysis*, 43(6):2758–2775, 2011.

[36] Brendan Pass. Optimal transportation with infinitely many marginals. *Journal of Functional Analysis*, 264(4):947–963, 2013.

[37] Brendan Pass. Remarks on the semi-classical Hohenberg-Kohn functional. *Nonlinearity*, 26(9):2731, 2013.

[38] Mircea Petrache and Simona Rota Nodari. Equidistribution of jellium energy for Coulomb and Riesz interactions. *Constructive Approximation*, 47(1):163–210, 2018.

[39] Aldo Pratelli. On the equality between Monge's infimum and Kantorovich's minimum in optimal mass transportation. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, volume 43, pages 1–13, 2007.

[40] Michael Seidl, Paola Gori-Giorgi, and Andreas Savin. Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities. *Physical Review A*, 75(4):042511, 2007.

[41] Cédric Villani. *Topics in optimal transportation.* American Mathematical Society, 2003.

[42] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[43] Constantin Zalinescu. *Convex analysis in general vector spaces.* World scientific, 2002.