Density functional theory and optimal transportation with Coulomb cost

Codina Cotar, Gero Friesecke and Claudia Klüppelberg
Department of Mathematics
Technische Universität München
cotar@ma.tum.de, gf@ma.tum.de, cklu@ma.tum.de

Abstract. We present here novel insight into exchange-correlation functionals in density functional theory, based on the viewpoint of optimal transport. We show that in the case of two electrons and in the semiclassical limit, the exact exchange-correlation functional reduces to a very interesting functional of novel form, which depends on an optimal transport map T associated with a given density ρ. Since the above limit is strongly correlated, the limit functional yields insight into electron correlations. We prove the existence and uniqueness of such an optimal map for any number of electrons and each ρ, and determine the map explicitly in the case when ρ is radially symmetric.

Contents

1 Introduction

The precise modelling of electron correlations continues to constitute the major obstacle in developing high-accuracy, low-cost methods for electronic structure computations in molecules and solids. In this article we shed new light on the longstanding problem of how to accurately incorporate electron correlation into density functional theory (DFT), by deriving and analyzing the semiclassical limit of the exact Hohenberg-Kohn functional with the single-particle density ρ held fixed. In this limit, we find that the exact functional reduces formally to a very interesting functional of novel form which depends on an *optimal transport map* T associated with a given density ρ. Our work thereby links DFT, which is a large and very active research area in physics and chemistry [PY95, FNM03, Ra09], for the first time to optimal transportation theory, which has recently become a very active area in mathematics [GM96, Ru96, Vill09].

In optimal transportation theory the goal is to transport "mass" from an initial density ρ_A to a target density ρ_B in such a way that the "cost" $c(x, y)$ for transporting mass from x to y is minimized. Mathematically, this means that one minimizes a cost functional $\int c(x, y) d\gamma(x, y)$ over a set of joint measures γ (in physics terminology: pair densities) subject to fixed marginals (single-particle densities). See below for a precise formulation. The main mathematical novelty of the optimal transportation problem arising from DFT,

$$\text{Minimize } \int_{\mathbb{R}^6} \frac{1}{|x - y|} d\gamma(x, y) \text{ subject to equal marginals } \rho, \tag{1.1}$$

is that the cost, which is given by the Coulomb law $c(x, y) = 1/|x - y|$, *decreases* rather than increases with distance and has a *singularity* on the diagonal.

Our goals in this paper are

(i) to prove that for any given single-particle density ρ, the optimal transportation problem with Coulomb cost possesses a unique minimizer which is given by an optimal transport map T_ρ associated with ρ. (It is well known that uniqueness is *false* for the seemingly simpler cost function $c(x, y) = |x - y|$.)

(ii) to derive an explicit formula for the optimal map in the case when ρ is radially symmetric. (Note that in physics, radial densities arise as atomic ground state densities for many elements such as He, Li, N, Ne, Na, Mg, Cu.)

(iii) to prove that DFT with electron interaction energy given by the optimal transportation cost $E_{OT}[\rho]$ (defined as the minimum cost in (1.1)) is the semiclassical limit of exact Hohenberg-Kohn DFT in case of two electrons, and establish basic properties such as that it is a rigorous lower bound to exact DFT for any number of electrons.

We do not know whether our semiclassical limit result remains true for a general number of electrons. As explained in Section 5, this question is related to the representability problem for two-particle density matrices [CY02].

To prove (i) and (ii) we adapt geometric methods as developed in [GM96], [Ru96], and [SK92], for cost functions that increase with distance. In our case of decreasing cost functions, one can still geometrically construct a potential which specifies both in which direction and how far to move the mass of ρ_A which sits near x. To prove (iii) we will need to make modifications to the optimal transport plan which yields $E_{OT}[\rho]$, since any wave function whose pair density is given by the optimal plan has infinite kinetic energy. The main technical idea here is a construction to re-instate the original marginals after smoothing.

The optimal transportation functional $E_{OT}[\rho]$ which emerges as a limit of the Hohenberg-Kohn functional

should be viewed as a natural "opposite" of the well known mean field functional $J[\rho] = \frac{1}{2} \int_{\mathbb{R}^6} \frac{\rho(x)\,\rho(y)}{|x-y|} dx\,dy$: it arises in a strongly correlated rather than a de-correlated limit, thereby yielding valuable qualitative insight into electron correlations. We also believe that $E_{OT}[\rho]$ has a role to play in the design of quantitative competitors to existing exchange-correlation functionals: it provides an alternative starting point of novel functional form for designing approximations, and – just like the mean field functional – could be incorporated as an ingredient into hybrid functionals. Basic quantitative issues are addressed in the companion paper [CFK11].

This paper is structured as follows. In Section 2 we discuss density functional theory from a mathematical perspective. In section 3 we introduce optimal transport theory, prove in Theorem 3.6 the main result of the section, the uniqueness of the optimal transport map, and establish some of its general properties. In section 4 we give in Theorem 4.10 an explicit formula for the optimal map for equal, radially symmetric marginals. In section 5 we compare the optimal transportation cost $E_{OT}[\rho]$ to the exact Hohenberg-Kohn functional, and show that it is its semiclassical limit in the case of two electrons (Theorem 5.2), as well as a lower bound for any number of electrons (Theorem 5.1).

2 Density functional theory

Density functional theory (DFT) was introduced by Hohenberg, Kohn and Sham in the 1960s in two fundamental papers [HK64, KS65], as an approximate computational method for solving the many-electron Schrödinger equation whose computational cost remains feasible even for large systems. In this theory, one only computes the single-particle density instead of the full wave function. In order to obtain a closed equation, a closure assumption is needed which expresses the pair density in terms of the single-particle density. A simple "independence" ansatz has turned out to be far too crude in practice. A huge effort has gone into developing corrections which account for the failure of independence [PY95, FNM03, Ra09].

Our goal in this section is to discuss DFT from a mathematical perspective.

The key quantity DFT aims to predict is the ground state energy E_0 of a molecule as a function of the atomic positions. From this, further properties can be readily extracted, for instance, in order to determine a molecule's stable equilibrium shapes one minimizes E_0 (locally or globally) over atomic positions.

Starting point for developing DFT models is the "exact" (non-relativistic, Born-Oppenheimer) quantum mechanical ground state energy E_0^{QM}. The definition contains some details which may look a bit complicated to readers not familiar with many-particle quantum theory, but the basic mathematical structure relevant to developing DFT models is simple. E_0^{QM} is the minimum value of a suitable energy functional $\mathcal{E}$ over a suitable class of functions $\mathcal{A}$.

2.1 Exact ground state energy

The detailed definition is as follows. Consider a molecule with atomic nuclei at positions $R_1, .., R_M \in \mathbb{R}^3$, with charges $Z_1, .., Z_M \in \mathbb{N}$ ($Z = 1$ means hydrogen, $Z = 2$ Helium, $Z = 3$ Lithium, $Z = 4$ Beryllium, $Z = 5$ Boron, $Z = 6$ Carbon, $Z = 7$ Nitrogen, $Z = 8$ Oxygen, and so on), and with N electrons. The energy functional depends on the positions and charges of the nuclei only through the ensuing Coulomb potential

$$v(x) = -\sum_{\alpha=1}^{M} \frac{Z_\alpha}{|x - R_\alpha|}, \quad x \in \mathbb{R}^3. \tag{2.1}$$

For the discussion below, the potential v can more generally be any function in the space $L^{3/2}(\mathbb{R}^3) + L^\infty(\mathbb{R}^3) = \{v_1 + v_2 \,|\, v_1 \in L^{3/2}(\mathbb{R}^3),\ v_2 \in L^\infty(\mathbb{R}^3)\}$. The class of functions $\mathcal{A}$ is given by

$$\mathcal{A} = \{\Psi \in L^2((\mathbb{R}^3 \times \mathbb{Z}_2)^N; \mathbb{C}) \,|\, \nabla\Psi \in L^2,\ \Psi \text{ antisymmetric},\ ||\Psi||_{L^2} = 1\}. \tag{2.2}$$

(Here antisymmetric means $\Psi(z_{\sigma(1)}, .., z_{\sigma(N)}) = sgn(\sigma)\Psi(z_1, .., z_N)$ for all permutations σ, where $z_1, ..., z_N \in \mathbb{R}^3 \times \mathbb{Z}_2$ are space-spin-coordinates for the N electrons. Spin will not play a big role in the sequel, but the

fact that the functions Ψ depend on all the positions of all the electrons is important. It leads to the fact that we will have to deal with "N-point distributions".) Elements of $\mathcal{A}$ are called (N-electron) *wave functions*.

The energy functional is given by

$$\mathcal{E}^{QM}[\Psi] = T[\Psi] + V_{ne}[\Psi] + V_{ee}[\Psi] \tag{2.3}$$

where (employing the notation $z_i = (x_i, s_i) \in \mathbb{R}^3 \times \mathbb{Z}_2$, $\int dz_i = \sum_{s_i \in \mathbb{Z}_2} \int_{\mathbb{R}^3} dx_i$)

$$T[\Psi] = \frac{1}{2} \int ... \int \sum_{i=1}^{N} |\nabla_{x_i} \Psi(x_1, s_1, .., x_N, s_N)|^2 dz_1 .. dz_N$$

is the kinetic energy,

$$V_{ne}[\Psi] = \int ... \int \sum_{i=1}^{N} v(x_i) |\Psi(x_1, s_1, .., x_N, s_N)|^2 dz_1 .. dz_N$$

is the electron-nuclei interaction energy, and

$$V_{ee}[\Psi] = \int ... \int \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|} |\Psi(x_1, s_1, .., x_N, s_N)|^2 dz_1 .. dz_N$$

is the electron-electron interaction energy.

The quantum mechanical ground state energy is defined as

$$E_0^{QM} = \inf_{\Psi \in \mathcal{A}} \mathcal{E}^{QM}[\Psi]. \tag{2.4}$$

In the usual case where v is given by (2.1) and $N = \sum_{\alpha=1}^{M} Z_\alpha$ (neutral molecules), it is a basic theorem due to Zhislin that the infimum is attained. For a simple proof see [Fr03].

Since the energy functional is a quadratic form, we could equivalently have defined E_0^{QM} as the lowest eigenvalue of the corresponding linear partial differential operator $-\frac{1}{2} \sum_{i=1}^{N} \Delta_{x_i} + \sum_{i=1}^{N} v(x_i) + \sum_{i<j} \frac{1}{|x_i - x_j|}$. This formulation is useful for many other purposes, but – unlike (2.4) – does not play an important role in DFT.

2.2 Probabilistic interpretation; marginals

The absolute value squared of Ψ can be interpreted as an N-point probability distribution,

$|\Psi(x_1, s_1, .., x_N, s_N)|^2 = $ probability density that the electrons are at positions x_i with spins s_i.

In quantum mechanics this is known as the Born interpretation. Note that the above function is nonnegative and integrates to 1, due to the requirement that Ψ has L^2 norm 1. (In fact this was the physical motivation for this requirement.)

Various partial marginals will play an important role. First, by integrating out the spins we obtain the N-point position density:

$$\rho_N^{\Psi}(x_1, .., x_N) = \sum_{s_1, .., s_N \in \mathbb{Z}_2} |\Psi(x_1, s_1, .., x_N, s_N)|^2. \tag{2.5}$$

Next, by integrating out all but two respectively one electron positions we obtain the pair density and the single particle density:

$$\rho_2^{\Psi}(x_1, x_2) = \binom{N}{2} \int_{\mathbb{R}^{3(N-2)}} \rho_N^{\Psi}(x_1, .., x_N) dx_3 .. dx_N, \tag{2.6}$$

$$\rho^{\Psi}(x_1) = N \int_{\mathbb{R}^{3(N-1)}} \rho_N^{\Psi}(x_1, .., x_N) dx_2 .. dx_N. \tag{2.7}$$

For the rest of this section, we drop the superscript Ψ from ρ^Ψ and ρ_2^Ψ.

The normalization factors are a convention in quantum mechanics so that ρ integrates to the number of particles and ρ_2 to the number of pairs in the system. (In Section 3 we find it convenient to work with the corresponding probability densities, normalized so as to integrate to 1.) With the above conventions, the important fact that ρ is a marginal distribution of ρ_2 takes the form

$$\frac{N}{\binom{N}{2}} \int \rho_2(x,y)\,dy = \rho(x), \quad \frac{N}{\binom{N}{2}} \int \rho_2(x,y)\,dx = \rho(y). \tag{2.8}$$

The relevance of ρ and ρ_2 for determining the ground state energy (2.4) come from the fact that the electron-nuclei energy V_{ne} and the electron-electron energy V_{ee} in (2.3) depend only on these.

Lemma 2.1. *With the above definitions, for any $\Psi \in \mathcal{A}$ we have*

$$V_{ne}[\Psi] = \int_{\mathbb{R}^3} v(x)\rho(x)\,dx, \quad V_{ee}[\Psi] = \int_{\mathbb{R}^6} \frac{1}{|x-y|}\rho_2(x,y)\,dx\,dy. \tag{2.9}$$

Proof. This follows from definitions (2.6), (2.7) and the fact that due to the antisymmetry of Ψ, $|\Psi(z_1,..,z_N)|^2$ is a symmetric function of the z_i. $\square$

In the sequel we write $V_{ne}[\rho]$, $V_{ee}[\rho_2]$ instead of $V_{ne}[\Psi]$, $V_{ee}[\Psi]$.

Also, we note that the space of densities arising from functions $\Psi \in \mathcal{A}$,

$$\mathcal{R} := \{\rho \,:\, \mathbb{R}^3 \to \mathbb{R} \,|\, \rho \text{ is the density (2.7) of some } \Psi \in \mathcal{A}\}, \tag{2.10}$$

is known explicitly: by a result of Lieb [Li83],

$$\mathcal{R} = \{\rho \,:\, \mathbb{R}^3 \to \mathbb{R} \,|\, \rho \geq 0, \sqrt{\rho} \in H^1(\mathbb{R}^3), \int_{\mathbb{R}^3} \rho(x)\,dx = N\}, \tag{2.11}$$

where $H^1(\mathbb{R}^3)$ is the usual Sobolev space $\{u \in L^2(\mathbb{R}^3) \,|\, \nabla u \in L^2(\mathbb{R}^3)\}$.

2.3 Universal Hohenberg-Kohn functional

The expression for V_{ne} derived in Lemma 2.1 leads to the following well known partitioning of the minimization (2.4) into a double minimization (first minimize over Ψ subject to fixed ρ, then over ρ):

$$E_0^{QM} = \inf_{\rho \in \mathcal{R}} \left\{ F_{HK}[\rho] + \int_{\mathbb{R}^3} v(x)\,\rho(x)dx \right\} \tag{2.12}$$

with

$$F_{HK}[\rho] := \inf_{\Psi \in \mathcal{A},\, \Psi \mapsto \rho} \left\{ T[\Psi] + V_{ee}[\rho_2] \right\}. \tag{2.13}$$

Here and below, the notation $\Psi \mapsto \rho$ means that Ψ has single-particle density ρ. Note that F_{HK} is a universal functional of ρ, in the sense that it does not depend on the external potential v, and is called the *Hohenberg-Kohn functional*. It is defined on the admissible set (2.10). The above constrained-search definition of F_{HK} is due to Levy and Lieb [Le79, Li83]; in the original Hohenberg-Kohn paper [HK64] the functional was constructed in a more indirect and slightly less general way, requiring that ρ be the density of some Ψ which is a non-degenerate ground state of $\mathcal{E}^{QM}$ for some potential v.

2.4 Exchange-correlation functionals

The problem with definition (2.4) of E_0^{QM}, as well as definition (2.13) of the 'exact' density functional F_{HK}, is that it is unfeasible in practice except when the number N of particles is very small. This is due to the so-called problem of exponential scaling: the functions over which one minimizes are functions on $\mathbb{R}^{3N}$ and the discretization of $\mathbb{R}^{3N}$ requires a K^N-point grid if the single-particle space $\mathbb{R}^3$ is discretized by a K-point grid.

This problem would disappear if we could accurately approximate V_{ee} in the variational principle (2.4) by a functional $\tilde{V}_{ee}$ of ρ instead of ρ_2,

$$V_{ee}[\rho_2] \approx \tilde{V}_{ee}[\rho]. \tag{2.14}$$

(Why this is so is not completely trivial, since there remains T to deal with, but see eq. (2.17) below.) Thus in DFT one approximates the variational principle for the ground state energy E_0^{QM} by:

$$E_0^{DFT} = \inf_{\Psi \in \mathcal{A}} \left\{ T[\Psi] + V_{ne}[\rho] + \tilde{V}_{ee}[\rho] \right\}. \tag{2.15}$$

Physically, this means (in the light of Lemma 2.1) that in DFT, interactions of electrons with an external environment, such as the Coulomb forces exerted by an array of atomic nuclei, are included exactly, but electron-electron interactions have to be suitably "modelled". By partitioning the minimization in (2.15) analogously to (2.12), (2.13), E_0^{DFT} can be obtained by minimization of a functional of ρ alone,

$$E_0^{DFT} = \inf_{\rho \in \mathcal{R}} \left\{ T_{QM}[\rho] + V_{ne}[\rho] + \tilde{V}_{ee}[\rho] \right\}, \quad T_{QM}[\rho] := \inf_{\Psi \in \mathcal{A},\, \Psi \mapsto \rho} T[\Psi]. \tag{2.16}$$

The minimization over the "large" space $\mathcal{A}$ of functions on $(\mathbb{R}^3 \times \mathbb{Z}_2)^N$ in (2.16) can now be replaced by a minimization over a much "smaller" space. As can be shown with the help of reduced density matrices [CY02], and abbreviating $\int = \sum_{s \in \mathbb{Z}_2} \int_{\mathbb{R}^3}$,

$$\inf_{\Psi \in \mathcal{A},\, \Psi \mapsto \rho} T[\Psi] = \inf \Big\{ \sum_{i=1}^{\infty} \frac{\lambda_i}{2} \int |\nabla \phi_i|^2 \quad | \quad 0 \le \lambda_i \le 1, \sum_{i=1}^{\infty} \lambda_i = N,$$

$$\phi_i \in H^1(R^3 \times \mathbb{Z}_2), \int \phi_i \overline{\phi_j} = \delta_{ij}, \sum_{i=1}^{\infty} \sum_{s \in \mathbb{Z}_2} |\phi_i(x,s)|^2 = \rho(x) \Big\}. \tag{2.17}$$

After truncating the sum after an appropriate number i_{max} of terms (the standard truncation being $i_{max} = N$, yielding the Kohn-Sham kinetic energy functional [KS65]) and discretizing $\mathbb{R}^3$ by a K-point grid, the number of degrees of freedom of the right hand side scales linearly instead of exponentially in N.

By means of this fact, the task of eliminating the exponential complexity of (2.4) is reduced to the following

Fundamental problem of DFT *Design accurate approximations of the form (2.14). In other words, approximate a simple explicit functional of the pair density ρ_2, a function on $\mathbb{R}^6$, by a functional of its joint right and left marginal ρ, a function on $\mathbb{R}^3$. Note that the approximations only need to be accurate for single-particle densities and pair densities of ground states of molecules, not arbitrary states. Elsewhere it suffices that the approximations give a reasonably good lower bound so as to avoid spurious minimizers.*

Example 2.2. (statistical independence) The simplest idea would be to assume statistical independence,

$$\rho_2(x,y) \approx \frac{1}{2}\rho(x)\rho(y) \tag{2.18}$$

(the factor 1/2 coming from the normalization factors in (2.7), (2.6)), and substitute this ansatz into the formula for V_{ee} derived in Lemma 2.1. This leads to taking

$$\tilde{V}_{ee}[\rho] = \frac{1}{2} \int_{\mathbb{R}^6} \frac{1}{|x-y|} \rho(x)\rho(y)\, dx\, dy =: J[\rho], \tag{2.19}$$

6

i.e. V_{ee} is replaced by the Coulomb self-repulsion of the single-particle density. The above mean field functional appears, for instance, in Thomas-Fermi-theory.

In modern DFT, the very naive ansatz (2.18) was never used, but – without this being natural from a probabilistic point of view – the convention is to include corrections to it additively, i.e. one makes an ansatz

$$\tilde{V}_{ee}[\rho] = J[\rho] + E_{xc}[\rho], \tag{2.20}$$

the additive correction being called an *exchange-correlation functional*. This notational convention should not, of course, prevent us from contemplating non-additive modifications of (2.18) and (2.19). □

Example 2.3. (correctly normalized mean field) Let $\rho_2 = \binom{N}{2}\gamma$ and $\rho = N\mu$, so that γ and μ have integral 1. Then

$$\gamma \approx \mu \otimes \mu \Leftrightarrow \frac{\rho_2}{\binom{N}{2}} \approx \frac{\rho \otimes \rho}{N^2},$$

which is equivalent to

$$\rho_2 \approx \frac{1}{2}\left(1 - \frac{1}{N}\right)\rho \otimes \rho,$$

where here and below we use the notation $(\rho \otimes \rho')(x,y) = \rho(x)\rho'(y)$, corresponding to the product measure when interpreting ρ, ρ' as measures.

Note that physicists and chemists use

$$\rho_2 \approx \frac{1}{2}\rho \otimes \rho,$$

which is justified in the context of macroscopic systems such as an electron gas (where one has taken a limit $N \to \infty$), but less natural in the context of DFT for atoms and molecules. □

Example 2.4. (local density approximation) In a model system, the so-called free electron gas, the pair density can be determined explicitly [Fr97]. In this case the single-particle density is a constant,

$$\rho(x) \equiv \overline{\rho}, \tag{2.21}$$

and the pair density can be determined to be

$$\rho_2(x,y) = \frac{1}{2}\overline{\rho}^2\left(1 - \frac{1}{q}h\left((3\overline{\rho}\pi^2)^{1/3}|x-y|\right)^2\right), \tag{2.22}$$

where

$$h(s) = \frac{3(\sin s - s\cos s)}{s^3}.$$

In particular, at long range $|x-y| \to \infty$ statistical independence is correct, but at short range $|x-y| \to 0$, ρ_2 tends to zero in the case of a single spin state, i.e. it vanishes on the diagonal $x = y$, and to half the size of a statistically independent sample in the (physical) case of two spin states. Substituting the result (2.22) into the formula for $V_{ee}[\rho_2]$ leads to the so-called *local density approximation*

$$\tilde{V}_{ee}[\rho] = J[\rho] - c_x \int_{\mathbb{R}^3} \rho(y)^{4/3}dy, \tag{2.23}$$

where $c_x = \frac{3}{4}(\frac{3}{\pi})^{1/3}$. As a heuristic approximation to V_{ee}, this formula goes back to Dirac and Bloch (for a rigorous justification see [Fr97]). It was widely used in the early days of DFT, following [KS65]. □

Example 2.5. (B3LYP) Current functionals used in practice, e.g. the 'B3LYP' functional of Becke, Lee, Yang and Parr [Be93, LYP88], rely on – from a mathematical point of view questionable – guesses of functional forms (e.g. local in ρ, or local in ρ and $\nabla\rho$), additional terms depending non-locally on the orbitals in (2.17), and careful fitting of parameters to experimental or high-accuracy-computational data. The resulting expressions are a little too complicated to write down here. They have led to an accuracy improvement for E_0^{QM} over the local density approximation of an order of magnitude or so, but not more, with little progress in the last decade despite continuing effort. □

7

3 Optimal transportation for DFT

We begin with a basic observation. The weight factor in front of ρ_2 in V_{ee} in (2.9) is always positive, and largest on the diagonal $x = y$, so even "complete anticorrelation" might be a better ansatz than independence (keeping in mind that the states on which the ansatz needs to be good are the minimizers of a functional which includes V_{ee}).

In fact, such a complete anticorrelation is exactly what emerges when one starts from the exact Hohenberg-Kohn functional F_{HK}, inserts a semiclassical factor $\hbar^2$ in front of the kinetic energy functional T in (2.13) and passes to the semiclassical limit $\hbar \to 0$. In this limit, the Hohenberg-Kohn functional F_{HK} reduces to the following functional obtained by a minimization over pair densities instead of wave functions,

$$\tilde{F}[\rho] = \inf_{\rho_2 \in \mathcal{R}_2, \, \rho_2 \mapsto \rho} \int_{\mathbb{R}^6} \frac{1}{|x - y|} \, \rho_2(x, y) \, dx \, dy. \tag{3.1}$$

Here $\rho_2 \mapsto \rho$ means that ρ_2 satisfies eq. (2.8), i.e. it has right and left marginal ρ, and the set $\mathcal{R}_2$ of admissbile pair densities is the image of $\mathcal{A}$ under the map $\Psi \mapsto \rho_2$. Unlike the corresponding admissible set of single-particle densities, $\mathcal{R}_2$ is not known explicitly (this is a variant of the representability problem for two-particle density matrices [CY02]).

Formally, ignoring this point and discarding in particular the smoothness restriction (which can be proved analogously to the proof in [Li83] of (2.11)) that

$$\rho_2 \in \mathcal{R}_2 \Longrightarrow \sqrt{\rho_2} \in H^1(\mathbb{R}^6), \tag{3.2}$$

the above functional $\tilde{F}$ reduces to the functional

$$E_{OT}[\rho] = \inf_{\rho_2 \in \mathcal{M}_+, \, \rho_2 \mapsto \rho} C[\rho_2], \quad C[\rho_2] := \int_{\mathbb{R}^6} \frac{1}{|x - y|} d\rho_2(x, y), \tag{3.3}$$

where $\mathcal{M}_+$ denotes the set of (nonnegative) Radon measures on $\mathbb{R}^6$. For a rigorous justification that (3.3) is indeed the correct semiclassical limit of F_{HK} in case $N = 2$ see section 5.

The variational problem that has appeared here, to minimize a "cost functional" C over a set of joint measures on $\mathbb{R}^6$ subject to fixed marginals, is an *optimal transport problem*. This type of problem, with "cost functions" such as $|x - y|$ or $|x - y|^2$ instead of $|x - y|^{-1}$, dates back to Monge in 1781 and has a famous history, which is nicely summarized in the very readable paper [GM96], which was our main source when studying the problem (3.3).

Before formulating our particular problem we give some notations and definitions.

For a set $Z \subset \mathbb{R}^d$, we denote by $\mathcal{P}(Z)$ the set of probability measures on Z. If Z is a closed subset of $\mathbb{R}^d$ and $\gamma \in \mathcal{P}(Z)$, then the support of γ is the smallest closed set $supp \, \gamma \subset Z$ of full mass, that is such that $\gamma(supp \, \gamma) = \gamma(Z) = 1$.

Suppose $X, Y \subset \mathbb{R}^d$ are closed sets. If $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, we denote by $\Gamma(\mu, \nu)$ the joint probability measures γ on $\mathbb{R}^d \times \mathbb{R}^d$ which have μ and ν as their marginals, that is with $\mu(U) = \gamma(U \times \mathbb{R}^d)$ and $\nu(U) = \gamma(\mathbb{R}^d \times U)$ for Borel $U \subset \mathbb{R}^d$. In fact, if $\gamma \in \Gamma(\mu, \nu)$, then $supp \, \gamma \subset X \times Y$. Typically μ has density u_1 and ν has density u_2, where u_1, u_2 are L^1 functions, in which case we write $\Gamma(u_1, u_2)$.

Remark 3.1. *In this section it is convenient to eliminate the prefactors in ρ_2 and ρ, and to consider the cost functional C on probability measures γ, with equal marginals μ (which are again probability measures), with ρ_2 in (3.3) corresponding to $\binom{N}{2}\gamma$ and ρ corresponding to $N\mu$ as in Example 2.3.*

We first formulate the general problem, which is now called *the Kantorovich problem*. For some *cost function* $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ we are interested in minimizing the transport cost

$$C[\gamma] := \int c(x, y) d\gamma(x, y), \tag{3.4}$$

among joint measures $\gamma \in \Gamma(\mu, \nu)$, called *transport plans*, to obtain

$$\inf_{\gamma \in \Gamma(\mu, \nu)} C[\gamma]. \tag{3.5}$$

Let $\mathcal{T}(\mu, \nu)$ be the set of Borel maps $T : \Omega \subset \mathbb{R}^d \to \mathbb{R}^d$ that *push μ forward to ν*, i.e. $T_{\#}\mu[V] := \mu(T^{-1}(V)) = \nu(V)$ for Borel $V \subset \mathbb{R}^d$. The so-called *Monge problem* is to minimize

$$I[T] = \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x) \tag{3.6}$$

over maps T in $\mathcal{T}(\mu, \nu)$, called *transport maps*, to obtain

$$\inf_{T_{\#}\mu=\nu} I[T]. \tag{3.7}$$

There is a natural embedding which associates to each transport map $T \in \mathcal{T}(\mu, \nu)$ a transport map $\gamma_T := (\mathrm{id} \times T)_{\#}\mu \in \Gamma(\mu, \nu)$ or, in physics notation, $\gamma_T(x, y) := \delta_{T(x)}(y)\mu(x)$, where id: $\mathbb{R} \to \mathbb{R}$ is the identity map. Since $C[\gamma_T] = I[T]$, we conclude that

$$\inf_{\gamma \in \Gamma(\mu, \nu)} C[\gamma] \leq \inf_{T \in \mathcal{T}(\mu, \nu)} I[T]. \tag{3.8}$$

Let $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ and endow $\bar{\mathbb{R}}$ with the usual topology so that $c \in C(\mathbb{R}^d \times \mathbb{R}^d, \bar{\mathbb{R}})$ means that $\lim_{(x,y)\to(\bar{x},\bar{y})} c(x, y) = c(\bar{x}, \bar{y})$. In particular, if $c(\bar{x}, \bar{y}) = +\infty$, then $c(x, y)$ tends to $+\infty$ as (x, y) tends to $(\bar{x}, \bar{y})$, and similarly for limit $-\infty$.

Throughout we will work with cost functions c such that $c(x, y) = h(x - y)$ for all $x, y \in \mathbb{R}^d$. We recall the definition of the dual of a convex function and refer to Rockafellar [Ro72] for standard definitions and further background.

Definition 3.2. *The dual (or Legendre transform) $h^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ of a convex function $h : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is given by*

$$h^*(y) := \sup_{x \in \mathbb{R}^d} \{< x, y > -h(x)\}. \tag{3.9}$$

Since our cost functions of interest are not convex, we need to work with the following more subtle definition of Legendre transform.

Definition 3.3. *(generalized Legendre transform) Suppose that $l : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and convex. Define $k : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ by $k(\lambda) = l(\lambda)$ if $\lambda \geq 0$ and $k(\lambda) = +\infty$ otherwise. Define $h^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ by*

$$h^*(x) = k^*(-|x|) := \sup_{\beta \in \mathbb{R}} \{-\beta|x| - k(|x|)\}, \quad x \in \mathbb{R}^d. \tag{3.10}$$

We define $l^\circ(\lambda) = k^*(-|\lambda|)$ for $\lambda \in \mathbb{R}$.

Definition 3.4. *A function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$, not identically $-\infty$, is said to be c-concave if it is the infimum of a family of translates and shifts of $h(x)$: i.e, there is a set $A \subset \mathbb{R}^d \times \mathbb{R}$ such that*

$$\psi(x) := \inf_{(y,\lambda)\in A} \{c(x, y) + \lambda\}, \quad x \in \mathbb{R}^d. \tag{3.11}$$

Let $\Delta := \{(x, x) \mid x \in \mathbb{R}^d\}$ and $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be such that $c(x, y) := h(x - y) := l(|x - y|) \geq 0$ and with c and l such that

(A1) $l : [0, \infty] \to [0, \infty]$ is strictly convex, strictly decreasing and C^1 on $(0, \infty)$;

(A2) $c \in C^1(\mathbb{R}^d \times \mathbb{R}^d \setminus \Delta, \mathbb{R})$;

(A3) $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty]$ is lower semi-continuous;

(A4) for every $x_0 \in \mathbb{R}^d, c(x_0, x_0) = +\infty$.

Remark 3.5. *Note that [GM96] only assume that l is continuous and not C^1; using their arguments, we could replace (A1) and (A2) by*

(A1') $l : [0, \infty] \to [0, \infty]$ *is strictly convex, decreasing and continuous on $(0, \infty)$;*

(A2') $c \in C(\mathbb{R}^d \times \mathbb{R}^d \setminus \Delta, \mathbb{R})$.

The main result of this section, obtained by combining the results from Theorems 3.25 and 3.27, is

Theorem 3.6. *Assume that c and l satisfy (A1)-(A4), and let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be absolutely continuous with respect to the Lebesgue measure. Then there exists a unique minimizer $\gamma \in \Gamma(\mu, \nu)$ of the functional C defined in (3.4), and a unique transport map T pushing μ forward to ν such that $\gamma = (id, T)_{\#}\mu$ (or, in physics notation, $\gamma(x, y) = \delta_{T(x)}(y)\mu(y)$). This map is of the form $T(x) = x - \nabla h^*(\nabla \psi(x))$ for some c-concave function ψ on $\mathbb{R}^d$.*

We next give a simple example, which illustrates the emergence of such an optimal map T when the marginals consist of two Dirac delta functions, and explicitly compute the function h^* appearing above for our cost function of interest.

Example 3.7. Let $a, b \in \mathbb{R}$. For c and l as in (A1)-(A4) we are interested in minimizing

$$\int c(x, y)d\gamma(x, y), \tag{3.12}$$

subject to

$$\int \gamma(x, y)dx = \delta_a(y) + \delta_b(y) \text{ and } \int \gamma(x, y)dy = \delta_a(x) + \delta_b(x), \tag{3.13}$$

where δ_a is the Dirac function such that, for all Borel subsets of $\mathbb{R}$, we have $\int_\Omega \delta_a(y) = 1$ if $a \in \Omega$, and $\int_\Omega \delta_a(y) = 0$ otherwise. We claim that the minimum in (3.13) is attained for

$$\gamma(x, y) = \delta_a(x)\delta_b(y) + \delta_b(x)\delta_a(y) =: \gamma_0(x, y). \tag{3.14}$$

To show this, note first that

$$\gamma(x, y) = c_{aa}\delta_a(x)\delta_a(y) + c_{ab}\delta_a(x)\delta_b(y) + c_{ba}\delta_b(x)\delta_a(y) + c_{bb}\delta_b(x)\delta_b(y),$$

with $c_{aa}, c_{ab}, c_{ba}, c_{bb} \geq 0$ and $c_{aa} + c_{ab} + c_{ba} + c_{bb} = 2$. Due to the constraints on γ from (3.13) we have $c_{ba} = c_{ab}$ and $c_{aa} = c_{bb}$. Hence

$$\gamma(x, y) = \alpha \left(\delta_a(x)\delta_a(y) + \delta_b(x)\delta_b(y) \right) + \beta \left(\delta_a(x)\delta_b(y) + \delta_b(x)\delta_a(y) \right), \text{ with } \alpha, \beta \geq 0 \text{ and } \alpha + \beta = 1.$$

Minimizing (3.12) subject to (3.13) is then equivalent to the following problem:

Minimize $2\alpha \, l(0) + 2\beta \, l(|b - a|)$ subject to the constraints $\alpha, \beta \geq 0, \, \alpha + \beta = 1$.

Since by (A1) we have $l(0) > l(|a - b|)$, the minimum in the above is attained for $\alpha = 0$ and $\beta = 1$, which proves the claim.

Formula (3.14) for the minimizer admits a very interesting interpretation which motivates the notion of *optimal transport map* and foresees the structure of general minimizers as given in Theorem 4.8. Denote the single-particle density $\delta_a(x) + \delta_b(x)$ in (3.13) by $\rho(x)$, and introduce the map $T : \{a, b\} \to \{a, b\}$ which maps a to b and b to a. Then T pushes ρ forward to ρ, and the optimal measure γ_0 has the form

$$\gamma_0(x, y) = \delta_{T(x)}(y)\rho(x),$$

or, in measure-theoretic notation,

$$\gamma_0 = (id, T)_{\#}\rho.$$

$\square$

Example 3.8. (generalized Legendre transform for Coulomb cost) Let $h(x) = k(|x-y|)$ with $k(\lambda) = \lambda^{-1}$ for $\lambda > 0$ and $+\infty$ for $\lambda \leq 0$. Note that $k : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous and convex. The ordinary Legendre transform of k is given by $k^*(s) = \sup_{\beta \in \mathbb{R}} \{\beta s - k(\beta)\}$. Since $k(\beta) = +\infty$ for $\beta \leq 0$, negative values of β do not contribute to the above supremum (the term in brackets then being $-\infty$). Consequently,

$$k^*(s) = \sup_{\beta > 0} \{\beta s - k(\beta)\}.$$

When $s > 0$, we infer $k^*(s) = +\infty$. It thus remains to calculate

$$k^*(-s) = \sup_{\beta > 0} \{-\beta s - k(\beta)\}, \quad s \geq 0.$$

Recall now that $k(\beta) = 1/\beta$. The elementary calculus problem of maximising the function in brackets on $(0, \infty)$ has the unique solution $\beta = 1/\sqrt{s}$, whence $k^*(-s) = -2\sqrt{s}$ for $s \geq 0$. Altogether it follows that the generalized Legendre transform of h is

$$h^*(x) = -2\sqrt{|x|}, \quad x \in \mathbb{R}^d. \tag{3.15}$$

Consequently, if $c(x, y) = \frac{1}{|x-y|}$ then the optimal map T in Theorem 3.6 is of the form $T(x) = x + \frac{\nabla \psi(x)}{|\nabla \psi(x)|^{3/2}}$. $\square$

Subsections 3.1-3.4 are devoted to the proof of Theorem 3.6. The proofs follow partially the proofs in [GM96], [KM07], [GM95] and [GO07]. In the final subsection, 3.5, we derive some general properties of the unique optimal measure and the optimal cost under the assumption of equal marginals, $\mu = \nu$.

3.1 Definitions and notation

In the following we present some definitions which are needed throughout the remainder of the section.

Definition 3.9. (1) *Let $V \subset \mathbb{R}^d$. A c-concave function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is said to be the c-transform on V of a function ϕ if (3.11) holds with $A \subset V \times \mathbb{R}$. Moreover,*

$$\psi(x) = \inf_{y \in V} \{c(x, y) - \phi(y)\},$$

for some function $\phi : V \to \mathbb{R} \cup \{-\infty\}$.

(2) *The c-transform of a function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is the function $\psi^c : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ defined by*

$$\psi^c(y) = \inf_{x \in \mathbb{R}^d} \{c(x, y) - \psi(x)\}.$$

(3) *A subset $S \subset X \times Y$ is called c-cyclically monotone, if for any finite number of points $(x_j, y_j) \in S, j = 1, \ldots, n$ and permutations $\sigma : \{1, .., n\} \to \{1, .., n\}$*

$$\sum_{j=1}^{n} c(x_j, y_j) \leq \sum_{j=1}^{n} c(x_{\sigma(j)}, y_j). \tag{3.16}$$

Remark 3.10. *Remark 3.4 of [GO07] proves the following: If $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is not identically $-\infty$, and is given by (3.11), then we have*
(i) $\psi(y) \geq -\lambda > -\infty$ for all $(y, \lambda) \in A$, where A is the set in (3.11). Hence, ψ^c is not identically $-\infty$.
(ii) $\psi^{cc} = \psi$.

Theorem 3.11 (optimal measures have c-cyclically monotone support: Proposition 3.2 of [GO07]). *Assume that $X, Y \in \mathbb{R}^d$ are closed sets, that $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and that $c \geq 0$ is lower semicontinuous on $X \times Y$. Then the following hold:*

(a) *There is at least one optimal measure $\gamma \in \Gamma(\mu, \nu)$.*

(b) *Suppose that in addition $c \in C(X \times Y, \bar{\mathbb{R}})$. Unless $C \equiv +\infty$ throughout $\Gamma(\mu, \nu)$, there is a c-cyclically monotone set $S \subset \mathbb{R}^d \times \mathbb{R}^d$ containing the support of all optimal measures in $\Gamma(\mu, \nu)$.*

Definition 3.12. (1) *A function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is superdifferentiable at $x \in \mathbb{R}^d$, if $\psi(x)$ is finite and there exists $y \in \mathbb{R}^d$ such that*

$$\psi(x+z) \le \psi(x) + <z, y> +o(|z|) \text{ as } |z| \to 0; \tag{3.17}$$

here $o(\lambda)$ means terms $\eta(\lambda)$ such that $\eta(\lambda)/\lambda$ tends to zero with λ.

(2) *A pair (x, y) belongs to the superdifferential $\partial^{\cdot}\psi \subset \mathbb{R}^d \times \mathbb{R}^d$ of ψ, if $\psi(x)$ is finite and (3.17) holds, in which case y is called a supergradient of ψ at x. Such supergradients y comprise the set $\partial^{\cdot}\psi(x) \subset \mathbb{R}^d$, while for $V \subset \mathbb{R}^d$ we define $\partial^{\cdot}\psi(V) := \cup_{x \in V} \partial^{\cdot}\psi(x)$.*

(3) *The analogous notions of* subdifferentiability, subgradients *and the* subdifferential $\partial_{\cdot}\psi$ *are defined by reversing inequality (3.17).*

(4) *A real-valued function ψ will be differentiabale at x precisely if it is both super- and subdifferentiable there; then*

$$\partial^{\cdot}\psi(x) = \partial_{\cdot}\psi(x) = \{\nabla\psi(x)\}.$$

Definition 3.13. *The c-superdifferential $\partial^c\psi$ of $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$, not identical $-\infty$, consists of the pairs $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ for which $c(x, y) - \psi(x) \le c(z, y) - \psi(z)$ for all $z \in \mathbb{R}^d$.*

Lemma 3.14 (relating c-superdifferentials to subdifferentials: Lemma 3.1 of [GM96]). *Let $h : \mathbb{R}^d \to \mathbb{R}^d$ and $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$. If $c(x, y) = h(x - y)$, then $(x, y) \in \partial^c\psi$ implies $\partial_{\cdot}\psi(x) \subset \partial_{\cdot}h(x - y)$. When h and ψ are differentiable, then $\nabla\psi(x) = \nabla h(x - y)$.*

Definition 3.15. *A function $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is said to be* locally semi-concave (locally semi-convex) *at $p \in \mathbb{R}^d$, if there is a constant $\lambda < \infty$, which makes $\psi(x) - \lambda|x|^2$ concave (convex) on some (small) open ball centered at p.*

Remark 3.16. *Suppose that $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ have no atoms and that γ^* minimizes $C[\gamma]$ over $\Gamma(\mu, \nu)$ and that $C(\gamma^*) < \infty$. Then $\gamma^*(\Delta) = 0$ and so, $\text{supp } \gamma^* \setminus \Delta$ contains at least one element, say (x_0, y_0). Also $\gamma^*(E) = 0$, where $E = (x_0 \times Y) \cup (X \times y_0)$. Hence the set $X \times Y \setminus (E \cap \Delta)$ is non-empty, so it contains an element $(\bar{x}_0, \bar{y}_0)$. Note that $x_0, y_0 \notin \{\bar{x}_0, \bar{y}_0\}$.*

We will need the non-atomic property of the marginal measures μ and ν (and the resulting remark above) from the uniqueness section 3.3 onwards, as a means to bypass the singularity of c on the diagonal. We will use Remark 3.16 in Lemma 3.20 and in Lemma 3.21 below.

Remark 3.17. *Suppose that $S \subset X \times Y$ is c-cyclically monotone and contains two pairs $(x_0, y_0), (\bar{x}_0, \bar{y}_0)$ such that $x_0 \ne y_0$, $\bar{x}_0 \ne \bar{y}_0$ and $\bar{x}_0, \bar{y}_0 \notin \{x_0, y_0\}$. Then for all $(x, y) \in S$, we have $x \ne y$.*

Proof. Let us assume that $(x, y) \in S$, with $x = y$. There are the following possibilities to consider: $x = y = x_0$, $x = y = y_0$, $x = y = \bar{x}_0$ $x = y = \bar{y}_0$ and $x, y \notin \{x_0, y_0, \bar{x}_0, \bar{y}_0\}$.

We present a proof for one of the cases, the other cases being treated analogously. Consider $x = y = y_0$. Then $x, y \notin \{x_0, \bar{x}_0, \bar{y}_0\}$ and from (3.16), we get

$$c(x, y) \le c(x, \bar{y}_0) + c(x_0, y) + c(\bar{x}_0, y_0) - c(x_0, y_0) - c(\bar{x}_0, \bar{y}_0) < +\infty,$$

which leads to a contradiction as $c(x, y) = +\infty$. $\qquad\square$

3.2 Existence of an optimal measure with c-cyclical monotone support

The main issue in this section is not the existence of an optimal measure, as the existence of an optimal measure is assured by Theorem 3.11 (a), but the existence of an optimal measure with c-cyclical monotone support. In order to use Theorem 3.11 (b) and construct such an optimal measure, we need to first construct a joint measure γ, with marginals μ and ν and with $C[\gamma] < \infty$. This is done in the Lemma below.

Lemma 3.18. *Suppose that $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and are absolutely continuous with respect to the Lebesgue measure. Then there exists $\gamma \in \Gamma(\mu, \nu)$ and $\epsilon > 0$ such that for all $(x, y) \in supp\ \gamma$,*

$$|x - y| > \epsilon$$

Proof. The proof follows similar arguments as the proof of Proposition 4.1 from [GO07] adapted to our situation.

Let $b, c \in \mathbb{R}_+$ and let $S(0; b) := \{x \in \mathbb{R}^d : |x| \leq b\}$, $S^c(0; b) := \{x \in \mathbb{R}^d : |x| > b\}$ and $S(b; c) := \{x \in \mathbb{R}^d : b < |x| \leq c\}$. Since μ and ν are absolutely continuous with respect to the Lebesgue measure, the functions

$$t \mapsto \mu|_{S(0;t)}, \ t \mapsto \nu|_{S(0;t)}, \ t \mapsto \mu|_{S^c(0;t)}, \ t \mapsto \nu|_{S^c(0;t)} \tag{3.18}$$

are continuous.

Step 1. We assume first that there exists $b \in \mathbb{R}_+$ such that $supp\ \mu \subset S(0; b)$ and $supp\ \nu \subset S^c(0; b)$. Then, because of (3.18), we may choose $\epsilon_1, \epsilon_2 > 0$ such that

$$\mu(S(0; b - \epsilon_1)) = \nu(S(b; b + \epsilon_2)) = \frac{1}{2} \text{ with } \epsilon_1 < b.$$

Let

$$\mu^- = \mu|_{S(0;b-\epsilon_1)}, \ \mu^+ = \mu|_{S(b-\epsilon_1;b)}, \ \nu^- = \nu|_{S(b;b+\epsilon_2)} \text{ and } \nu^+ = \nu|_{S^c(0;b+\epsilon_2)}.$$

Set

$$\gamma := 2(\mu^- \otimes \nu^- + \mu^+ \otimes \nu^+).$$

Note that $\gamma \in \Gamma(\mu, \nu)$ and for all $(x, y) \in supp\ \gamma$, we have

$$|x - y| \geq \min\{\epsilon_1, \epsilon_2\} > 0.$$

Step 2. Assume that μ and ν are arbitrary in $\mathcal{P}(\mathbb{R}^d)$. We use (3.18) to choose $b \in \mathbb{R}_+$ such that

$$\mu(S(0; b)) = \nu(S^c(0; b)) = m. \tag{3.19}$$

If $m = 0$, then we reduce the discussion to Step 1. Similarly if $m = 1$. We can therefore assume that $0 < m < 1$. More precisely, if we denote by $f(b) := \mu(S(0; b)) - \nu(S^c(0; b))$ for all $b \in \mathbb{R}_+$, then $f(b)$ is an increasing and continuous function of b, going from negative values to positive ones as b goes from 0 to $+\infty$. Therefore, there exists $b_0 \in \mathbb{R}_+$ such that $f(b_0) = 0$, which is equivalent to $\mu(S(0; b_0)) = \nu(S^c(0; b_0))$. Set

$$\mu^- = \mu|_{S(0;b_0)}, \ \mu^+ = \mu|_{S^c(0;b_0)}, \ \nu^- = \nu|_{S(0;b_0)} \text{ and } \nu^+ = \nu|_{S^c(0;b_0)}.$$

By (3.19), $\frac{\mu^-}{m}$ and $\frac{\nu^+}{m}$ are probability measures. They satisfy

$$supp\ \mu^- \subset S(0; b_0) \text{ and } supp\ \nu^+ \subset S^c(0; b_0).$$

Therefore, they satisfy the assumptions of Step 1, so there exists $\delta_1 > 0$ and $\gamma_1 \in \Gamma(\frac{\mu^-}{m}, \frac{\nu^+}{m})$ such that for all $(x, y) \in supp\ \gamma_1$, we have

$$|x - y| > \delta_1.$$

Similarly, there exists $\delta_2 > 0$ and $\gamma_2 \in \Gamma(\frac{\mu^+}{1-m}, \frac{\nu^-}{1-m})$ such that for all $(x, y) \in supp\ \gamma_2$, we have

$$|x - y| > \delta_2.$$

Set $\gamma := m\gamma_1 + (1-m)\gamma_2$. Then $\gamma \in \Gamma(\mu, \nu)$ and for all $(x,y) \in supp\ \Gamma(\mu, \nu)$, we have

$$|x - y| \geq \min\{\delta_1, \delta_2\}.$$

$\square$

Theorem 3.19 (existence of optimal measure with c-cyclical monotone support)**.** *Assume that $c(x,y) = h(x - y) := l(|x - y|)$ satisfies (A1)–(A4) and let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be absolutely continuous with respect to the Lebesgue measure. Then there exists a measure γ_0 with c-cyclically monotone support which minimizes the functional $C(\gamma)$ introduced in (3.4) over $\Gamma(\mu, \nu)$.*

Proof. By Lemma 3.18, there exists $\gamma \in \Gamma(\mu, \nu)$ and $\epsilon > 0$ such that

$$|x - y| > \epsilon \tag{3.20}$$

for all $x, y \in supp\ \gamma$. Since l is strictly decreasing on $(0, \infty)$, (3.20) together with (A1) ensure that c is uniformly bounded from above on $supp\ \gamma$ by $l(\epsilon)$. This proves that $C[\gamma] < \infty$. The statement follows now immediately from Theorem 3.11. $\square$

3.3 Geometrical characterization of the optimal measure

It is well known that a set is cyclically monotone if and only if it is contained in the subdifferential of a c-concave function; this result was proved for general cost functions $c : X \times Y \to \mathbb{R}$ in [SK92]. The following theorem is a further extension that is needed to deal with cost functions which satisfy (A1)-(A3) (and are allowed to take the value $+\infty$ somewhere).

Lemma 3.20. *Suppose that $X, Y \subset \mathbb{R}^d$ are closed sets.*

(1) *For $S \subset X \times Y$ to be c-cyclically monotone, it is necessary and sufficient that $S \subset \partial^c \psi$ for some c-concave $\psi : X \to \mathbb{R} \cup \{-\infty\}$.*

(2) *Suppose that $S \subset X \times Y$ is c-cyclically monotone and contains two pairs $(x_0, y_0), (\bar{x}_0, \bar{y}_0)$ such that $\bar{x}_0, \bar{y}_0 \notin \{x_0, y_0\}$. Let $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ be the c-concave function from (1). Then*

    (2a) $\psi(x_0)$ and $\psi(\bar{x}_0)$ are finite,

    (2b) whenever $(x, y) \in \partial^c \psi$, we have that $\psi(x) > -\infty$ and $x \neq y$,

    (2c) we have μ-a.e.

$$\psi(x) = \inf_{y \in Y} \{c(x, y) - \psi^c(y), x \neq y\}, \quad x \in X.$$

Proof. Part (1) has been proved in Theorem 2.7 of [GM96] or Lemma 2.1 of [Ru96].
(2) As in the proof of Lemma 2.1 in [Ru96], we define

$$\psi(x) = \inf_{n \in \mathbb{N}} \inf_{\{(x_j, y_j)\}_{j=0}^n \subset S} \left\{ c(x, y_n) + \sum_{j=0}^{n-1} c(x_{j+1}, y_j) - \sum_{j=0}^{n} c(x_j, y_j) \right\}, \quad x \in X. \tag{3.21}$$

(2a) The construction above yields a c-concave ψ on X with $S \subset \partial^c \psi$. Standard arguments (see for example [Ru96]) give that $\partial^c \psi$ is c-cyclically monotone in the sense that

$$\sum_{j=0}^{n} c(x_j, y_j) \leq \sum_{j=0}^{n-1} c(x_{n+1}, y_n) + c(x_{n+1}, y_n) \quad \text{with } x_{n+1} = x_0. \tag{3.22}$$

and so $\psi(x_0) \geq 0$. Taking $n = 1$, $x_1 = \bar{x}_0$ and $y_1 = \bar{y}_1$ in (3.21) gives that $\psi(x_0) \leq 0$. We conclude that $\psi(x_0) = 0$ and so ψ is finite and not identically $-\infty$. Note that by construction, ψ is the c-concave function from (1). Recall now that $(x, y) \in \partial^c \psi$ is equivalent to

$$c(x, y) - \psi(x) \leq c(z, y) - \psi(z), \quad z \in X. \tag{3.23}$$

14

Now recall Remark 3.16. Setting $(x, y) = (\bar{x}_0, \bar{y}_0)$, $z = x_0$ in (3.23), using the fact that $\psi(x_0) = 0$ and that $x_0 \neq \bar{y}_0$, we obtain that $\psi(\bar{x}_0)$ is finite.

(2b) Next, if $(u_0, y) \in S$, setting $z = u_0$, we have that

$$c(x, y) - \psi(x) \leq c(u_0, y) - \psi(u_0). \tag{3.24}$$

If $y \neq x_0$, we set $u_0 = x_0$ in (3.24) to obtain the claim. If $y = x_0$, we set $u_0 = \bar{x}_0$ and we use the fact that $\psi(\bar{x}_0)$ is finite, to obtain the claim.

(2c) This representation is a simple consequence of the construction. $\qquad\square$

We have proved the following. If we define

$$\partial_0^c \psi := \{(x, y) \in \partial^c \psi \mid x \neq y\} \quad \text{and} \quad \partial_0^c \psi(x) := \partial^c \psi(x) \setminus \{x\}, \tag{3.25}$$

then $\partial_0^c \psi = \partial^c \psi$ μ-a.e. and we will focus on the off-diagonal elements from now on. We also assume for the rest of this chapter that $X = Y = \mathbb{R}^d$.

Lemma 3.21 (μ-a.e. differentiability of c-transforms)**.** *Let c and l satisfy (A1)-(A4). Then the function ψ from Lemma 3.20 is μ-a.e. differentiable on $\mathbb{R}^d$.*

Proof. Recall that

$$\psi(x) = \inf_{y \in \mathbb{R}^d} \{c(x, y) - \psi^c(y), x \neq y\}.$$

We will prove that ψ is μ-a.e. differentiable on $\mathbb{R}^d$.

Step 0. Let $r > 0$ be such that ψ takes finite values at two or more points in $S(0; r)$ (this is possible due to Remark 3.16 and Lemma 3.20). Let $0 < a < r$ arbitrarily fixed, which means that $S(0; a) \subset S(0; r)$. We will show in Steps $1 - 3$ below that ψ is μ-a.e. differentiable on $S(0; a)$, from which we will derive in Step 4 the corresponding property on $\mathbb{R}^d$. The reason for the choice of an $0 < a < r$ such that $S(0; a) \subset S(0; r)$, will become apparent in Step 2 below.

In order to prove that ψ is μ-a.e. differentiable on $S(0; a)$, take $x \in S(0; a)$ arbitrarily fixed. Then

$$\psi(x) = \min\{\psi_1(x), \psi_2(x)\},$$

where

$$\psi_1(x) = \inf_{y \in S(0;r)} \{c(x, y) - \psi^c(y), x \neq y\} \quad \text{and} \quad \psi_2(x) = \inf_{y \in S^c(0;r)} \{c(x, y) - \psi^c(y), x \neq y\}.$$

Due to the fact that ψ takes finite values at two or more points in $S(0; r)$, it follows from the definition that ψ_1 and ψ_2 also take finite values at two or more points in $S(0; r)$.

Step 1. ψ_1 is locally Lipschitz and semi-concave on $S(0; a)$:

As ψ_1 takes finite values at two or more points, the proof follows the same reasoning as the proof of Proposition A.6 in [KM07] and will be omitted. Note also that by Proposition C.6 of [GM96], differentiability of ψ_1 can only fail on a set of μ-measure zero.

Step 2. ψ_2 is locally Lipschitz and μ-a.e. differentiable on $S(0; a)$:

Let $\delta := r - a > 0$. Define $\xi \leq 0$ by using the right derivative $2\delta\xi := l'(\delta^+)$ of l at δ. Then the function $l_\delta(\lambda) = l(\lambda) - \xi\lambda^2$ is strictly convex on $[\delta, \infty)$ and non-decreasing since $l'_\delta(\delta^+) = 0$. Extend this function to $\lambda \leq \delta$ by making $l_\delta(\lambda)$ constant-valued there. Then $h_\delta(x) := l_\delta(|x|)$ will be convex on $\mathbb{R}^d$: taking $x, y \in \mathbb{R}^d$ and $0 < t < 1$ implies

$$h_\delta((1 - t)x + ty) \leq l_\delta((1 - t)|x| + t|y|) \leq (1 - t)h_\delta(x) + th_\delta(y). \tag{3.26}$$

Note that $h(x) = h_\delta(x) + \xi|x|^2$ whenever $|x| \geq \delta$. Take $x \in S(0; a)$ and $y \in S^c(0; r)$; then $|x - y| \geq \delta$ and the definition of ψ_2 yields

$$\psi_2(x) - \xi|x|^2 = \inf_{y \in S^c(0;r)} \left\{ h_\delta(x - y) + 2\xi < x, y > + \xi y^2 - \psi^c(y) \right\}.$$

15

Note now that h_δ satisfies conditions (H1)-(H3) of [GM96]. Since l is continuous on $(0, \infty)$, we can apply Theorem 3.3 of [GM96] to h_δ and ψ_2 (see also Proposition C.2 in [GM96]) and thus, $\psi_2(x) - \xi|x|^2$ will be locally Lipschitz. Using the fact that $\psi_2(x) - \xi|x|^2$ is locally Lipschitz, Rademacher's theorem shows that the gradient $\nabla\psi_2$ is defined μ-a.e. everywhere on $S(0; r)$.

Step 3. ψ is μ-a.e. differentiable on $S(0; a)$:

Since $\psi(x) = \min(\psi_1(x), \psi_2(x))$, the μ-a.e. differentiability of ψ on $S(0; a)$ follows immediately as ψ is the minimum of two μ-a.e. differentiable functions.

Step 4. ψ is μ-a.e. differentiable on $\mathbb{R}^d$:

Let $(a_n)_{n \in \mathbb{N}}$ be an increasing sequence of positive real numbers tending to infinity as $n \to \infty$ and $\mathbb{R}^d = \cup_{n \in \mathbb{N}} S(0; a_n)$. Let $A := \{x \in \mathbb{R}^d : \psi \text{ is differentiable at } x\}$. Then $\mu(A) = \lim_{n \to \infty} \mu(A \cap S(0; a_n))$. The statement follows now immediately by means of Step 3. □

The following is a version of Lemma 5.2 of [GM96] for strictly convex and decreasing l.

Lemma 3.22 (the c-superdifferential lies in the graph of a map). *Let c satisfy (A1)-(A4). Suppose that $\psi : \mathbb{R}^d \to \mathbb{R}$ is differentiable at some $x \in \mathbb{R}^d$. Then $y \in \partial_0^c \psi(x)$ implies that h^* is differentiable at $\nabla\psi(x)$ and that $y = x - \nabla h^*(\nabla\psi(x))$.*

Proof. We recall first that by (A1), ∇h is injective off the diagonal; the injectivity of ∇h off the diagonal is **crucial** for the argument of this lemma, as will become apparent below. The proof follows similar steps as the proof of Lemma 5.2 in [GM96]. Let $y \in \partial_0^c \psi(x)$. Then $x \neq y$, so h is differentiable at $x - y$. From Lemma 3.14 (see also Lemma 3.1 of [GM96]) we have that $\partial.\psi(x) \in \nabla h(x - y)$. As ψ is differentiable at x, we have $\partial.\psi(x) = \nabla\psi(x)$ and $\nabla\psi(x) = \nabla h(x - y)$. Since $x \neq y$ and since l is strictly convex and strictly decreasing, the gradient $\nabla\psi(x)$ does not vanish μ-a.e. Lemma A.4 (ii)-(iii) implies both $(\nabla\psi(x), x - y) \in \partial.h^*$ and differentiability of h^* at $\nabla\psi(x)$, whence $\nabla h^*(\nabla\psi(x)) = x - y$. □

Lemma 3.23. *Let c and l satisfy (A1)-(A4) and let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$. Suppose that a joint measure $\gamma \in \Gamma(\mu, \nu)$ has $supp\ \gamma \subset \partial_0^c \psi = \partial^c \psi$, where $\psi : \mathbb{R}^d \to R$ is the c-transform of a function on $supp\ \nu$. The map $T(x) := x - \nabla h^*(\nabla\psi(x))$ pushes μ forward to ν. In fact, $\gamma = (id, T)_{\#}\mu$ and $T_{\#}\mu = \nu$.*

Proof. The proof follows the same steps of Theorem 5.4 from [GM96]. For the reader's convenience, we provide the reasoning below.

To begin, one would like to know that the map $T(\cdot)$ is Borel and defined μ-a.e.. By Lemma 3.21, differentiability of ψ can only fail on a set $\mathcal{N}$ of μ-measure zero in $\mathbb{R}^d$. Thus $\mu(\mathcal{N}) = 0$, $\gamma(N \times Y) = 0$ and so the map $\nabla\psi$ is defined μ-a.e.. Since by Remark 3.16, $\gamma(\Delta) = 0$ and since $supp\ \gamma \subset \partial^c \psi = \partial_0^c \psi$, we have $\gamma(\partial_0^c \psi) = 1$. Therefore, define $S := \{(x, y) \in \partial_0^c \psi | x \in dom \nabla\psi\}$, where $dom \nabla\psi$ denotes the subset of $\mathbb{R}^d$ on which ψ is differentiable. Lemma 3.21 shows that ψ is μ-a.e. differentiable on $dom\ \psi$. Since its gradient is obtained as the pointwise limit of a sequence of continuous approximants (finite differences), $\nabla\psi$ is Borel measurable on the (Borel) set $dom\ \nabla\psi$ where it can be defined. Lemma A.4 shows that ∇h^* is a Borel map. For $(x, y) \in S$, Lemma 3.22 implies that T is defined at x and $y = T(x)$. Thus T is defined on the projection of S onto $\mathbb{R}^d$ by $\pi(x, y) := x$; it is a Borel map since $\nabla\psi$ and ∇h^* are. Moreover, the set $\pi(S)$ is Borel and has full measure for μ: both $\partial_0^c \psi$ and $\pi(\partial_0^c \psi)$ are σ-compact, so $\pi(S) = \pi(\partial_0^c \psi) \cap dom \nabla\psi$ is the intersection of two Borel sets with full measure. Thus $\gamma(Z \cap S) = \gamma(Z)$ for $Z \subset \mathbb{R}^d \times \mathbb{R}^d$. It remains to check that $(id, T)_{\#}\mu = \gamma$, from which $T_{\#}\mu = \nu$ follows immediately.

It suffices now to show that the measure $(id, T)_{\#}\mu$ coincides with γ on products $U \times V$ of Borel sets $U, V \in \mathbb{R}^d$; the semi-algebra of such products generates all Borel sets in $\mathbb{R}^d \times \mathbb{R}^d$. Define $S := \{(x, y) \in \partial_0^c \psi | x \in dom\ \nabla\psi\}$. Therefore, since $y = T(x)$ if $(x, y) \in S$, we have

$$(U \times V) \cap S = ((U \cap T^{-1}(V)) \times \mathbb{R}^d) \cap S. \qquad (3.27)$$

Being the intersection of two sets having full measure for γ – the closed set $\partial_0^c \psi$ and the Borel set $dom\ \nabla\psi \times \mathbb{R}^d$ – the set S is Borel with full measure. Thus $\gamma(Z \cap S) = \gamma(Z)$ for $Z \subset \mathbb{R}^d \times \mathbb{R}^d$. Applied to (3.27), this yields

$$\gamma(U \times V) = \gamma((U \cap T^{-1}(V)) \times \mathbb{R}^d) = \mu(U \cap T^{-1}(V)) = (id, T)_{\#}\mu(U \times V).$$

16

$\gamma \in \Gamma(\mu, \nu)$ implies the second equation, Definition 3.6 implies the third. $\qquad\square$

Remark 3.24. *Note that by Lemma 3.20 and Lemma 3.22, we have*

$$\mu(\{x \in \mathbb{R}^d : T(x) = x\}) = 0.$$

Theorem 3.25 (geometric representation of optimal solution)**.** *Assume that c and l satisfy (A1)-(A4), and let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be absolutely continuous with respect to the Lebesgue measure. Then the following hold:*

  (a) *There exists an optimal measure $\gamma \in \Gamma(\mu, \nu)$ with c-cyclical monotone support;*

  (b) *For any such γ, there is a function $\psi : \mathbb{R}^d \to \mathbb{R}$ which is the c-transform of some function on supp γ such that the map $T := id - \nabla h^*(\nabla \psi)$ pushes μ forward to ν and satisfies $\gamma = (id, T)_{\#}\mu$.*

  (c) *$\gamma := (S \times id)_{\#}\nu$ for some inverse map $S : \mathbb{R}^d \to \mathbb{R}^d$.*

  (d) *$S(T(x)) = x$ μ-a.e., while $T(S(y)) = y$ ν-a.e.*

Proof. For completeness, we sketch the arguments used in [GM96].

(a) and (b) The existence of an optimal γ with c-cyclical monotone support is guaranteed by Theorem 3.19. Then *supp* $\gamma \subset \partial_0^c \psi$. The map $\pi(x, y) = x$ on $\mathbb{R}^d \times \mathbb{R}^d$ pushes γ forward to $\mu = \pi_{\#}\gamma$ while projecting the closed set $\partial_0^c \psi$ to a σ-compact set of full measure for μ. From Lemma 3.21, we know that ψ is differentiable μ-a.e. Lemma 3.23 shows that $T(\cdot)$ pushes μ forward to ν while γ coincides with the measure $(id, T)_{\#}\mu$. The proofs of (c) and (d) follow the same reasoning as the proofs of (iv) and (v) from Theorem 4.6 in [GM96] and will be omitted.

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.4   Uniqueness of the optimal measure

Lemma 3.26 (c-superdifferentiability of c-transforms)**.** *Let c and l satisfy (A1)-(A4) and let $V \subset \mathbb{R}^d$ be a closed set. Let $\psi : \mathbb{R}^d \to \mathbb{R}$ be the c-transform of a function on V and suppose that $T := id - \nabla h^*(\nabla \psi)$ can be defined at some $p \in \mathbb{R}^d$ (i.e., ψ is differentiable at p and ∇h^* exists at $\nabla\psi(p)$). Then $\partial_0^c \psi(p) = \{T(p)\}$.*

Proof. The proof follows similar arguments as the proof of Proposition 6.1 of [GM96]. From Lemma 3.22, it is clear that $\partial_0^c \psi(p) \subset \{T(p)\}$. Therefore, we only need to prove that $\partial_0^c \psi(p)$ is non-empty. Assume that $T(p)$ is defined for some $p \in \mathbb{R}^d$. By c-concavity of ψ, there is a sequence $(y_n, \alpha_n)_{n=1}^{\infty} \subset A \subset V \times \mathbb{R}$ such that

$$\psi(p) = \lim_{n \to +\infty} [c(p, y_n) + \alpha_n]. \qquad (3.28)$$

As is shown below, $(|y_n|)_{n=1}^{\infty}$ must be bounded. We first assume this bound to complete the proof. Since $(y_n)_{n=1}^{\infty}$ is bounded, a subsequence must converge to a limit $y_n \to y$ in the closed set V. On the other hand, $y \in \partial_0^c \psi(p)$ since for all $x \in \mathbb{R}^d$, (3.11) and (3.28) imply

$$\psi(x) \leq \inf_{n \in \mathbb{N}}\{c(x, y_n) + \alpha_n\} \leq c(x, y) + \psi(p) - c(p, y),$$

with both $\psi(x), \psi(p) > -\infty$, as shown by Lemma 3.20 (2b). Thus, $p \neq y$ and $y \in \partial_0^c \psi(p)$. It remains only to prove that the sequence $(|y_n|)_{n=1}^{\infty}$ is bounded, which means that we can extract a convergent subsequence that converges to a point $y \in Y$. To produce a contradiction, suppose that a subsequence diverges in a direction $\hat{y}_n \to \hat{y}$. Then $|p - y_n|$ is bounded away from zero by $\delta > 0$. Since for each arbitrary small $x \in \mathbb{R}^d$ we have $|p - y_n - x| > 0$, it follows that ∇h exists at $p - y_n - x$. More precisely, for each n the uniform subdifferentiability in Lemma A.3 gives

$$h(p - y_n) \geq h(p - y_n - x) - <x, w_n(x)> + O_\delta(|x|^2), \text{ where } w_n(x) = \frac{p - y_n - x}{|p - y_n - x|}l'(|p - y_n - x|) \quad (3.29)$$

17

for arbitrary small $x \in \mathbb{R}^d$ and $O_\delta(|x|^2)$ independent of n. We re-write

$$< x, w_n(x) >=< x, \frac{p - y_n}{|p - y_n|} > \frac{|p - y_n|}{|p - y_n - x|} l'(|p - y_n - x|)- < x, x > \frac{l'(|p - y_n - x|)}{|p - y_n - x|}.$$

Equation (3.29) now becomes

$$h(p-y_n) \geq h(p-y_n-x)- < x, \frac{p - y_n}{|p - y_n|} > \frac{|p - y_n|}{|p - y_n - x|} l'(|p-y_n-x|)-\frac{1}{\delta}|x|^2 l'(|p-y_n-x|)+O_\delta(|x|^2). \quad (3.30)$$

Recall now that the derivative of l is negative and increasing and, therefore, $|l'(|p - y_n - x|)| \leq |l'((\delta + |x|))|$. The sequence $(|y_n|)_{n=1}^\infty$ can only diverge if $|l'(|p - y_n - x|)|$ tends to $|l'(\infty)| := \inf_\lambda |l'(\lambda)|$. Therefore, $\frac{|p-y_n|}{|p-y_n-x|} l'(|p - y_n - x|) \to l'(\infty)$. Taking a subsequence if necessary ensures that $\frac{p-y_n}{|p-y_n|}$ converge to a limit $w \in \mathbb{R}^d$, with $|w| = 1$. Thus $\frac{p-y_n}{|p-y_n-x|} l'(|p - y_n - x|)$ converges to $wl'(\infty)$, which is independent of x. Combining (3.30) with the definition of a c-concave function, this yields

$$\psi(p) \geq \psi(p - x)- < x, w > l'(\infty) + O_\delta(|x|),$$

where the large n limit has been taken using (3.28). By taking $z = -x$ in the above equation, we get

$$\psi(p)+ < z, w_1 > +O_\delta(|x|) \geq \psi(p + z), \text{ where } w_1 := -wl'(\infty) \text{ and } |w_1| = |l'(\infty)|.$$

Thus $w_1 \in \partial\cdot\psi(p)$. On the other hand, differentiability of ψ at p implies $\partial\cdot\psi(p) = \{\nabla\psi(p)\}$, whence $w_1 = \nabla\psi(p)$. Now $(w_1, p - T(p)) \in \partial\cdot h^*$ follows from the definition of $T(p)$. Assume now that $p = T(p)$. It follows that $(w_1, 0) \in \partial\cdot h^*$. If h^* is non-constant, Lemma A.2 gives $(-|y|, 0) \in \partial\cdot l$ – a result which is obvious when h^* is constant. Thus $(-|y|, 0) \in \partial\cdot l)$ by Lemma A.1 (i). This conclusion contradicts the fact that l is not differentiable at 0. Therefore $T(p) \neq p$ and $(\nabla h)(p - T(p)) = w_1$. Lemma A.2 gives $(|p - T(p)|, -|w_1|) \in \partial\cdot l$. Since l is strictly convex, $|w_1| > |l'(\infty)|$, which produces a contradiction. Therefore, the sequence $(y_n)_{n=1}^\infty$ is bounded. $\qquad\square$

Theorem 3.27 (uniqueness of the optimal map). *Assume that c and l satisfy (A1)-(A4) and let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be such that they are absolutely continuous with respect to the Lebesgue measure. Then an optimal map T pushing μ forward to ν is uniquely determined μ-a.e. by the requirement that it is of the form $T(x) = x - \nabla h^*(\nabla\psi(x))$ for some c-concave ψ on $\mathbb{R}^d$.*

Proof. For completeness' sake, we will sketch the main idea of the proof, as given in Theorem 4.4 of [GM96].

We will prove by contradiction that T is unique. That is, we assume that there exists, in addition to T and ψ, a second c-concave function ψ' for which $T'(x) := x - \nabla h^*(\nabla\psi'(x))$ pushes μ forward to $T_\#\mu = T'_\#\mu = \nu$. Recall now that ψ and ψ' are μ-a.e. differentiable. T and T' are defined μ-almost everywhere, and unless they coincide, there exists some $p \in \mathbb{R}^d$ at which both ψ and ψ' are differentiable but $T(p) \neq T'(p)$. From this, it is clear that $\nabla\psi(p) \neq \nabla\psi'(p)$.

Let $U := \{x \in \mathbb{R}^d | \psi(p) > \psi'(p)\}$. A contradiction will be derived by showing that the push-forwards $T_\#\mu = \nu$ and $T'_\#\mu = \nu$–alleged to coincide–must differ on $V := \partial^c\psi(U)$. We will show that

$$\mu(T^{-1}(V)) < \mu(U) \leq \mu(T'^{-1}(V)).$$

The main ingredient necessary to prove the last equation is the fact that $\partial_0^c\psi(p) = \{T(p)\}$, which is proved in Lemma 3.26. By using this together with the fact that μ is absolutely continuous with respect to the Lebesgue measure, the proof follows via the same arguments as Theorem 4.4 from [GM96] and will be omitted. $\qquad\square$

3.5 Some general properties of the optimal measure and the optimal cost for equal marginals

We will investigate in this subsection the case of equal marginals $\mu = \nu$ with common density ρ, and assume throughout that the cost function c satisfies conditions (A1)-(A4). We also introduce the optimal cost

$$E_{OT}^{norm}[\rho] := \inf_{\gamma \in \Gamma(\rho,\rho)} C[\gamma].$$

$E_{OT}^{norm}[\rho]$ corresponds to the non-normalized functional E_{OT} introduced in Section 3 via

$$E_{OT}^{norm}[\rho] = \frac{1}{\binom{N}{2}}E_{OT}[N\rho].$$

For the proof of Theorem 3.29 below, we will need the following lemma.

Lemma 3.28. *For any marginals μ and ν, and any map T which pushes μ forward to ν, μ, ν and T satisfy the following equation*

$$\mu(x) = \nu(T(x))|det\ DT(x))|,$$

where DT is the approximate gradient of T (for a definition see for example Definition 10.2 of [Vill09]).

Proof. We have that $\mu(T^{-1}(A)) = \nu(A)$ for all Borel sets $A \subseteq \mathbb{R}^d$. Then for any such A we have

$$\int_{y\in T(A)} d\nu(y) = \int_{x\in A} d\mu(x). \tag{3.31}$$

With the change of variables $y = T(x)$ the left-hand side of (3.31) becomes

$$\int_{y\in T(A)} d\nu(y) = \int_{y\in A} |det\ DT(x)|\nu(T(x))dx. \tag{3.32}$$

From (3.31) and (3.32), the claim follows. □

Theorem 3.29. *Assume that $\mu = \nu$ with common density ρ. Then*

(a) *The optimal measure γ_T which minimizes $C[\gamma]$ is symmetric, that is*

$$\gamma_T(A \times B) = \gamma_T(B \times A) \text{ for all Borel } A, B \in \mathbb{R}^d.$$

(b) *The optimal cost $E_{OT}^{norm}[\rho]$ is strictly convex in ρ.*

(c) *Let $c(x,y) = 1/|x-y|$. Then for all $\alpha > 0$ we have the following dilation behaviour*

$$E_{OT}^{norm}[\alpha^d \rho(\alpha \cdot)] = \alpha E_{OT}^{norm}[\rho(\cdot)].$$

Proof. (a) Recall from Theorem 3.25 that $\gamma_T = (id, T)_\# \mu$, where T is the optimal map which pushes μ forward to ν, i.e. $\mu(T^{-1}(A)) = \nu(A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$. Then $\gamma_T(x,y) = \mu\left((id,T)^{-1}(x,y)\right) = \delta_{T(x)}(y)\mu(x)$. Using this we get

$$\gamma_T(A \times B) = \int_{x\in A}\int_{y\in B} \delta_{T(x)}(y)\mu(dx) = \int_{x\in A} \chi_B(T(x))\mu(dx), \tag{3.33}$$

where χ denotes the indicator function. We now use Lemma 3.28 and the fact that $\mu = \nu$, then the right hand-side of (3.33) becomes

$$\begin{aligned}
\int_{x\in A} \chi_B(T(x))\mu(dx) &= \int_{x\in A} \chi_B(T(x))\mu(T(x))|det\ DT(x)|dx = \int_{y\in T^{-1}(A)} \chi_B(y)\mu(y)dy \\
&= \int_{y\in B} \chi_{T^{-1}(A)}(y)\mu(y)dy = \int_{y\in B} \chi_A(T(y))\mu(y)dy \\
&= \gamma_T(B \times A).
\end{aligned}$$

(b) The convexity is immediate from the definition of $E_{OT}^{norm}[\rho]$, and strict convexity follows from uniqueness.

(c) Fix any $\alpha > 0$. Then

$$
\begin{aligned}
E_{OT}^{norm}[\alpha^d \rho(\alpha \cdot)] &= \inf_{\tilde{\gamma} \in \Gamma(\alpha^d \rho(\alpha \cdot), \alpha^d \rho(\alpha \cdot))} \int \frac{1}{|x-y|} \tilde{\gamma}(x, y) dx dy \\
&= \inf_{\alpha^{2d} \gamma(\alpha x, \alpha y) : \gamma \in \Gamma(\rho, \rho)} \int \frac{1}{|x-y|} \alpha^{2d} \gamma(\alpha x, \alpha y) dx dy \\
&= \inf_{\gamma \in \Gamma(\rho, \rho)} \int \frac{1}{|x-y|} \alpha^{2d} \gamma(\alpha x, \alpha y) dx dy \\
&= \inf_{\gamma \in \Gamma(\rho, \rho)} \alpha \int \frac{1}{|x-y|} \gamma(x', y') dx' dy' \\
&= \alpha E_{OT}^{norm}[\rho(\cdot)],
\end{aligned}
$$

where for the second equality we used the fact that for $\gamma \in \Gamma(\rho, \rho)$ we have $\int \alpha^{2d} \gamma(\alpha x, \alpha y) dy = \alpha^d \rho(\alpha x)$ and for the penultimate equality we used the change of variables $x' = \alpha^d x, y' = \alpha^d y$. $\qquad\square$

Remark 3.30. *(0) Trivially, the statements in Theorem 3.29 also hold for the non-normalized functional* E_{OT}.

(1) Recall from Lemma 2.1 that the exact energy interaction is of form

$$
V_{ee}[\Psi] = \int_{\mathbb{R}^6} \frac{1}{|x-y|} \rho_2(x, y) \, dx \, dy,
$$

with the exact pair density ρ_2 being symmetric due to the antisymmetry condition on the underlying Ψ in (2.2). Property (a) of Theorem 3.29 shows that the approximate interaction energy

$$
E_{OT}[\rho] = \int_{\mathbb{R}^6} \frac{\rho_2^{opt}(x, y)}{|x-y|} \, dx \, dy,
$$

is of the same form, with the arising ρ_2^{opt} being automatically symmetric as a consequence of optimality coupled with the weaker symmetry condition that ρ_2^{opt} has equal marginals.

(2) Property (c) of Theorem 3.29 is a scaling property of the exact electron-electron energy V_{ee} not shared by many approximate density functionals used in the physics literature, such as the local density approximation (2.23).

(3) The dilation behaviour of E_{OT} equals that of the exact V_{ee}, as well as that of approximations like (2.23).

4   Explicit example - equal radially symmetric marginals

As in the last subsection, we continue to investigate the case of equal marginals $\mu = \nu$ with common density $\rho : \mathbb{R}^d \to [0, \infty)$. Moreover, we assume that $\rho(x) > 0$ for all $x \in \ supp \ \mu$. We will also assume throughout that the cost function c satisfies conditions (A1)-(A4).

Throughout this section, for any dimension $d \in \mathbb{N}$ we will denote the optimal map by $T^{(d)}$. In subsection 4.1 we will explicitly compute $T^{(1)}$, and in subsection 4.2 we use the one-dimensional analysis to explicitly compute $T^{(d)}$ when ρ is radially symmetric, that is to say when $\rho(x) = \lambda(|x|)$ for all $x \in \mathbb{R}^d$ and some function λ.

As turns out, in the above situations the optimal map is *universal* with respect to all cost functions satisfying (A1)-(A4), but the fact that $c(x, y)$ decreases with the distance $|x - y|$ is essential.

4.1   Explicit solution for equal marginals in one dimension

Let $\mu = \nu \in \mathcal{P}(\mathbb{R})$ be equal marginals on $\mathbb{R}$. Moreover, we define $I := supp \ \mu$. Recall from Theorem 3.11 that the unique optimal measure has c-cyclically monotone support. This will help us to characterize the optimal map $T^{(1)}$ in the following lemmas.
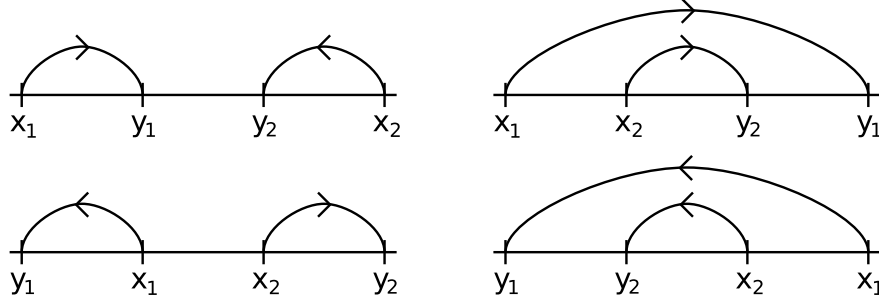
Figure 1: Configurations excluded by Lemma 4.1

.

Lemma 4.1. *Let (x_1, y_1) and (x_2, y_2) be two points in the support of the optimal map γ, that is $y_1 = T^{(1)}(x_1)$ and $y_2 = T^{(1)}(x_2)$. The possible configurations (not counting the symmetries between (x_1, y_1) and (x_2, y_2)) are: $x_1 < x_2 \leq y_1 \leq y_2$, $x_1 \leq y_2 \leq y_1 \leq x_2$, $y_1 \leq x_2 < x_1 \leq y_2$, $y_1 \leq y_2 \leq x_1 < x_2$, $x_1 \leq y_2 \leq x_2 \leq y_1$, $y_1 \leq x_2 \leq y_2 \leq x_1$, $x_1 \leq y_1 \leq x_2 \leq y_2$ and $y_1 \leq x_1 \leq y_2 \leq x_2$ (see also Figure 1 for the excluded configurations).*

Proof. If (x_1, y_1) and (x_2, y_2) are two points in the support of the optimal map γ, then

$$c(x_1, y_1) + c(x_2, y_2) \leq c(x_1, y_2) + c(x_2, y_1).$$

Let us consider the excluded cases one by one.

(i) $x_1 \leq y_1 < y_2 \leq x_2$.

   Then, due to the fact that l is strictly decreasing, it follows that

   $$c(x_1, y_1) + c(x_2, y_2) > c(x_1, y_2) + c(x_2, y_1),$$

   which contradicts the c-cyclically monotonicity property of the optimal solution.

(ii) $y_1 \leq x_1 < x_2 \leq y_2$.

   Similar to (i).

(iii) $x_1 < x_2 \leq y_2 < y_1$.

   We have $y_2 - x_2 < y_2 - x_1 < y_1 - x_1$. Therefore, $y_2 - x_1 = t(y_2 - x_2) + (1 - t)(y_1 - x_1)$ and $y_1 - x_2 = (1 - t)(y_2 - x_2) + t(y_1 - x_1)$, where $t \in [0, 1]$. Thus, using the strict convexity of h, we have

   $$c(x_2, y_1) + c(x_1, y_2) < tc(x_2, y_2) + (1 - t)c(x_1, y_1) + (1 - t)c(x_2, y_2) + tc(x_1, y_1) = c(x_1, y_1) + c(x_2, y_2),$$

   which contradicts the c-cyclically monotonicity property of the optimal solution.

(iv) $y_1 < y_2 \leq x_2 < x_1$

   Similar to (iii).

$\square$

Lemma 4.2. *Assume $(x_1, y_1), (x_2, y_2) \in supp\, \gamma$ are such that one of the following four configurations holds: $x_1 < y_2 < x_2 < y_1$ or $y_1 < x_2 < y_2 < x_1$ or $x_1 < y_1 < x_2 < y_2$ or $y_1 < x_1 < y_2 < x_2$. Then, if (x_3, y_3) is another point in the supp γ, none of the following configurations are possible: $x_i < x_k < y_j < x_j < y_k < y_i$, $x_i < y_k < y_j < x_j < x_k < y_i$, $y_i < x_k < x_j < y_j < y_k < x_i$, $y_i < y_k < x_j < y_j < x_k < x_i$, $x_i < y_i < x_j < y_j < x_k < y_k$ and $y_i < x_i < y_j < x_j < y_k < x_k$, where $i, j, k \in \{1, 2, 3\}$ (see also Figure 2).*
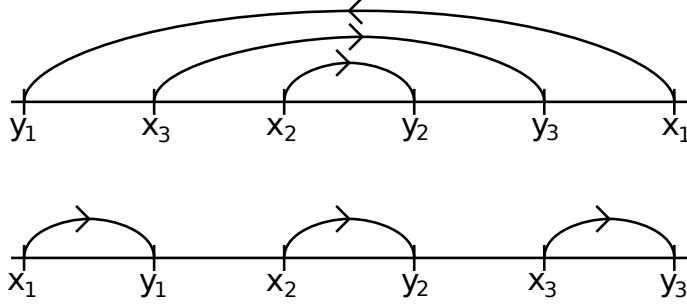
21

Figure 2: Example of configurations excluded by Lemma 4.2
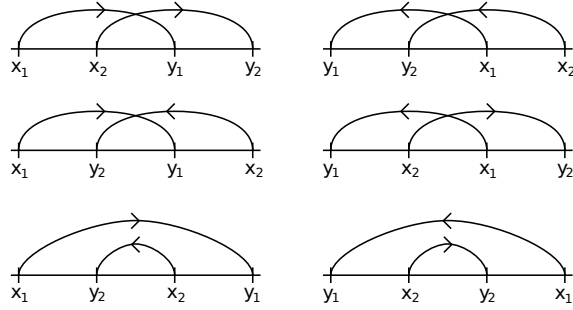
.

Figure 3: Possible configurations by Remark 4.4

Proof. Note that the first 4 configurations are immediately excluded by Lemma 4.1. Let us focus on the penultimate configuration. From the c-cyclically monotonicity property, we have that

$$c(x_1, y_1) + c(x_2, y_2) + c(x_3, y_3) \leq c(x_j, y_k) + c(x_i, y_j) + c(x_k, y_i).$$

But due to the fact that h is strictly decreasing, we have that $c(x_i, y_i) > c(x_i, y_j)$, $c(x_j, y_j) > c(x_k, y_i)$ and $c(x_k, y_k) > c(x_j, y_k)$, which gives rise to a contradiction. The last configuration can be dealt with in a similar way. $\square$

Remark 4.3. *Note that by Lemma 4.1, if $(x_1, y_1), (x_2, y_2)$ and $(x_3, y_3) \in supp\ \gamma$, with $x_1 < x_2 < x_3$ the following configurations are also not possible: $x_1 < y_2 < y_1 < x_2 < x_3 < y_3$ and $x_1 < y_2 < y_1 < y_3 < x_2 < x_3$. Similarly, the configurations $y_1 < x_1 < x_2 < y_3 < y_2 < x_3$ and $x_1 < x_2 < y_1 < y_3 < y_2 < x_3$ are not possible.*

Remark 4.4. *From Lemma 4.1, Lemma 4.2 and Remark 4.3, it follows that the configurations μ-a.e. possible are of form: $x_1 < x_2 < y_1 < y_2$, $x_1 < y_2 < y_1 < x_2$, $y_1 < y_2 < x_1 < x_2$, $y_1 < x_2 < x_1 < y_2$, $x_1 \leq y_2 < x_2 \leq y_1$ and $y_1 \leq x_2 < y_2 \leq x_1$ (see also Figure 3).*

Definition 4.5. *We say that $T^{(1)}$ has no points of decrease on $A \subseteq I$ if $\mu(\{x \in A : \exists x' \in A, x' > x, T^{(1)}(x) > T^{(1)}(x')\}) = 0$. We say that $T^{(1)}$ has points of decrease on A with positive measure if $\mu(\{x \in A : \exists x' \in A, x' > x, T^{(1)}(x) > T^{(1)}(x')\}) > 0$. We say that $T^{(1)}$ is μ-a.e. decreasing on $A \subseteq I$ if $\mu(\{x \in A : \exists x' \in A, x' > x, T^{(1)}(x) > T^{(1)}(x')\}) = 1$. We define similarly for $T^{(1)}$ the notions of points of increase on A and μ-increasing on A.*

Note that we can assume that the set B of such x' above such that $T^{(1)}$ has points of decrease (respectively points of increase) on A with positive measure, is also a set of positive measure. Otherwise, if the set of such x' is of μ-measure zero, we may consider the set $A \setminus B$, on which $T^{(1)}$ is μ-a.e. decreasing (respectively μ-a.e. increasing).
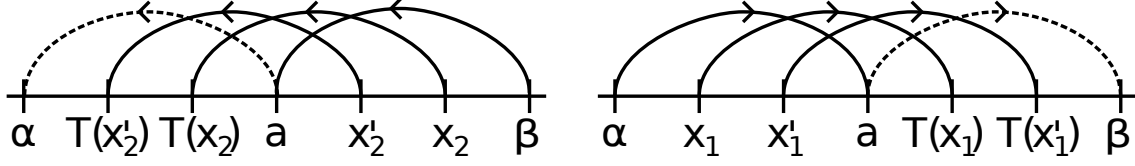
Figure 4: Optimal map configurations

Lemma 4.6. *The map $T^{(1)}$ cannot be μ-a.e. decreasing on any subset $A \subseteq I$.*

Proof. Asssume that $T^{(1)}$ is μ-a.e. decreasing on a subset $A \subseteq I$. Let $(x_i, y_i)_{i=1}^3 = (x_i, T^{(1)}(x_i))_{i=1}^3$, where $x_i \in A$ for $i \in \{1, 2, 3\}$. Recall now from Remark 4.4 the possible configurations by which $T^{(1)}$ is decreasing.

Assume first that $x_1 < y_2 < x_2 < y_1$. Then the possibilities for (x_3, y_3) such that $T^{(1)}$ is strictly decreasing, are: $x_3 < x_1 < y_2 < x_2 < y_1 < y_3$, $x_1 < x_3 < y_2 < y_3 < x_2 < y_1$, $x_1 < y_2 < y_3 < x_3 < x_2 < y_1$, $x_1 < y_2 < x_3 < y_3 < x_2 \leq y_1$, $y_3 < x_1 < y_2 < x_2 < x_3 < y_1$, $x_1 < y_3 < y_2 < x_2 < x_3 < y_1$, $y_3 < x_1 < y_2 < x_2 < y_1 < x_3$ and $x_1 < y_3 < y_2 < x_2 < y_1 < x_3$. In view of Lemmas 4.1 and 4.2, each these possibilities can only happen on a set of μ-a.e. measure 0. The case $y_1 < x_2 < y_2 < x_1$ can be treated similarly.

Assume next that $x_1 < y_2 < y_1 < x_2$. Then the possibilities for (x_3, y_3) such that $T^{(1)}$ is strictly decreasing, are: $x_3 < x_1 < y_2 < y_1 < y_3 < x_2$, $x_3 < x_1 < y_2 < y_1 < x_2 < y_3$, $x_1 < x_3 < y_2 < y_3 < y_1 < x_2$, $x_1 < y_2 < x_3 < y_3 < y_1 < x_2$, $x_1 < y_2 < y_3 < x_3 < y_1 < x_2$, $y_3 < x_1 < y_2 < y_1 < x_2 < x_3$ and $x_1 < y_3 < y_2 < y_1 < x_2 < x_3$. In view of Lemmas 4.1 and 4.2, each these possibilities can only happen on a set of μ-a.e. measure 0. The case $y_1 < x_2 < x_1 < y_2$ can be treated in a similar way. $\square$

Remark 4.7. *We have $supp\ \gamma = I \times I$; in particular, $Im(T^{(1)}) = I$, where we denoted by $Im(T^{(1)})$ the image of $T^{(1)}$.*

Proof. Note first that $\gamma(I \times I) = \gamma(I \times \mathbb{R}) = \mu(I) = 1$. Let us now assume that $supp\ \gamma = I \times J$, with $I, J \subseteq \mathbb{R}$ and $J \subset I$, with $\mu(I \setminus J) > 0$. Then $1 = \gamma(I \times J) = \gamma(I \times \mathbb{R}) = \mu(I) = \gamma(\mathbb{R} \times J) = \mu(J)$, which contradicts the definition of the support I of the marginals. $\square$

For the proof of the next theorem, we will use the results of Theorems 3.25 and 3.27; in particular, we will use the properties of the map $T^{(1)}$, as given in those two theorems.

Theorem 4.8. *Let $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$ and let $I = [\alpha, \beta]$. There exists $a \in I$ such that $T^{(1)}$ is μ-a.e. increasing on $[\alpha, a]$ and μ-a.e. increasing on $(a, \beta]$, with $T^{(1)}(\alpha) = T^{(1)}(\beta) = a$, $T^{(1)}(a_-) = \beta$ and $T^{(1)}(a_+) = \alpha$, with discontinuity at a. Except on a set of μ-measure zero, we have*

(a) *For all $x_1, x_1' \in (\alpha, a)$ with $x_1 < x_1'$, we have $x_1 < x_1' < T^{(1)}(x_1) < T^{(1)}(x_1')$, with $T^{(1)}(x) \in (a, \beta)$;*

(b) *For all $x_2, x_2' \in (a, \beta)$ with $x_2 < x_2'$, we have $T^{(1)}(x_2) < T^{(1)}(x_2') < x_2 < x_2'$, with $T^{(1)}(x) \in (\alpha, a)$;*

(c) *For every interval $(l_1 l_2) \subseteq I$, we have $T((l_1, l_2)) = (r_1, r_2) \subseteq I$.*

Moreover, a is such that $\mu([\alpha, a]) = \mu([a, \beta]) = \frac{1}{2}$ (see also Figure 4).

Proof. Recall first from Remark 3.24 that

$$\mu(\{x \in \mathbb{R}^d : T^{(1)}(x) = x\}) = 0.$$

Recall also from Theorem 3.25 that T is a bijective map; in particular, $\mu(x \in I : \exists y \in I \setminus \{x\}, T^{(1)}(x) = T^{(1)}(y)) = 0$.

Step 1. $T^{(1)}$ cannot be μ-a.e. increasing on I:

Assume that $T^{(1)}$ is μ-a.e. increasing on I. Then $T^{(1)}$ is μ-a.e. strictly increasing and it can only be increasing as described in Remark 4.4. Suppose first that for μ-a.e. all $x_1, x_2 \in I$, we have $x_1 < x_2 < T^{(1)}(x_1) < T^{(1)}(x_2)$. Then for each $x \in [\alpha, \beta]$, we have two possibilities: $T^{(1)}(\alpha, x) = (\alpha, T^{(1)}(x))$ or $T^{(1)}(\alpha, x) \subseteq (c, T^{(1)}(x))$, with $\alpha < c$. If $T^{(1)}(\alpha, x) = (\alpha, T^{(1)}(x))$, using $\mu = \mu \circ (T^{(1)})^{-1}$ and the fact that $\rho > 0$, we get that $T^{(1)}(x) = x$ μ-a.e., which contradicts Remark 3.24. If $T^{(1)}(\alpha, x) \subseteq (c, T^{(1)}(x))$, then $T^{(1)}(\alpha, \beta) \subseteq (c, T^{(1)}(\beta))$. Using again $\mu = \mu \circ (T^{(1)})^{-1}$, we get that $\mu(\alpha, \beta) = 1 \leq \mu(c, T^{(1)}(\beta))$, where $\alpha < c$, which would contradict the definition of $supp\, \mu = [\alpha, \beta]$. The case with $T^{(1)}(x_1) < T^{(1)}(x_2) < x_1 < x_2$ for μ-a.e. all $x_1, x_2 \in I$, can be discounted the same way.

Step 2. $T^{(1)}$ cannot have both points of increase and points of decrease on every interval $A \subseteq I$ with positive measure:

Assume that there exists an interval $A \subseteq I$, such that $T^{(1)}$ has both points of increase and points of decrease on A with positive measure. Take an arbitray point $x_1 \in A$ such that $\exists x_1' \in A, x_1' > x_1$ and $T^{(1)}(x_1') > T^{(1)}(x_1)$. By assumption, for a subset of such x_1 (and for a subset of x_1') in A of positive measure, $\exists x_2, x_2' \in (x_1, x_1')$ with $x_2' > x_2$ and $T^{(1)}(x_2') < T^{(1)}(x_2)$. Assume now that $x_1 < x_1' < T^{(1)}(x_1) < T^{(1)}(x_1')$. Due to the μ-a.e. possible configurations as given by Remark 4.4, x_2' is such that $T^{(1)}(x_2') < x_2'$. Then the c-cyclical monotonicity of the optimal support fails for $(x_1', T^{(1)}(x_1'))$ and $(x_2', T^{(1)}(x_2'))$. If we denote by B the set of such x_1', we have $\gamma(B, T^{(1)}(B)) = \mu(B) > 0$, which contradicts the assumption on the optimal support. The case with $T^{(1)}(x_1) < T^{(1)}(x_1') < x_1 < x_1'$ can be treated similarly so its proof will be omitted.

Step 3. There exists $(\alpha, \alpha_1), (\beta_1, \beta) \subset I$, with $\alpha_1 \leq \beta_1$, such that μ-a.e. for all $x_1 \in (\alpha, \alpha_1)$ we have $T^{(1)}(x_1) > x_1$ and μ-a.e. for all $x_2 \in (\beta_1, \beta)$ we have $T^{(1)}(x_2) < x_2$:

Recall first the possible configurations, as given by Remark 4.4. Note now that in view of Step 1, Step 2 and of Lemma 4.6, there exists $(\alpha, \alpha_1), (\beta_1, \beta) \subset I$, with $\alpha_1 \leq \beta_1$ on which $T^{(1)}$ is an increasing function. It remains to show that the optimal map can only be such that μ-a.e. for all $x_1 \in (\alpha, \alpha_1)$, we have $T^{(1)}(x_1) > x_1$ and μ-a.e. for all $x_2 \in (\beta_1, \beta)$, we have $T^{(1)}(x_2) < x_2$.

Let us consider the alternatives one by one. Suppose to begin with that $\mu(x_1 \in (\alpha, \alpha_1) : T^{(1)}(x_1) < x_1) > 0$ and $\mu(x_2 \in (\beta_1, \beta) : x_2 < T^{(1)}(x_2)) > 0$. By Lemma 4.1, if $x_1 \in (\alpha, \alpha_1)$ with $T^{(1)}(x_1) < x_1$ and $x_2 \in (\beta_1, \beta)$ with $x_2 < T^{(1)}(x_2)$, c-cyclical monotonicity of the support fails for $(x_1, T^{(1)}(x_1))$ and $(x_2, T^{(1)}(x_2))$. Assume next that $\mu(x_1 \in (\alpha, \alpha_1) : T^{(1)}(x_1) < x_1) > 0$ and $\mu(x_2 \in (\beta_1, \beta) : T^{(1)}(x_2) < x_2) > 0$. In view of Step 1, of Lemma 4.6 and of Lemma 4.7, there exists then some subset $(\alpha_1', \beta_1') \subseteq (\alpha_1, \beta_1)$ such that with positive measure, $T^{(1)}$ has both points of increase and points of decrease on (α_1', β_1'). But this contradicts the conclusion of Step 2 and therefore our assumption has to be wrong. The case with $\mu(x_1 \in (\alpha, \alpha_1) : x_1 < T^{(1)}(x_1)) > 0$ and $\mu(x_2 \in (\beta_1, \beta) : x_2 < T^{(1)}(x_2)) > 0$ can be reasoned similarly, so its proof will be omitted.

Step 4. $T^{(1)}(\alpha) \geq T^{(1)}(\beta)$:

Assume that $T^{(1)}(\alpha) < T^{(1)}(\beta)$. By Step 3, there exist $(\alpha, \alpha') \subset (\alpha, T^{(1)}(\alpha))$ and $(\beta', \beta) \subset (T^{(1)}(\beta), \beta)$ on which $T^{(1)}$ is as described in (a) and (b). Then, if $x_1 \in (\alpha, \alpha')$ and $x_2 \in (\beta', \beta)$, c-cyclical monotonicity of $supp\, \gamma$ would fail for $(x_1, T^{(1)}(x_1))$ and $(x_2, T^{(1)}(x_2))$, as shown in Lemma 4.1 (i). Therefore, $T^{(1)}(\alpha) \geq T^{(1)}(\beta)$.

Step 5. There exists $b \in I$ such that $T^{(1)}$ is as described in (a) and (b), for $(\alpha, b) \subset I$ and for $(b, \beta) \subset I$, respectively:

Note that by Lemma 4.6, Step 1 and Step 2, $T^{(1)}$ has to be μ-a.e. increasing on a certain number of sub-intervals of (α_1, β_1). On any such sub-intervals, either $T^{(1)}(x) < x$ μ-a.e. or $T^{(1)}(x) > x$ μ-a.e. In both these cases, due to the form of $T^{(1)}$ on (α, α_1) and (α, α_1), as proved in Step 3, the c-cyclical monotonicity of the support would fail on a set of positive measure unless $\mu((\alpha_1, \beta_1)) = 0$.

Step 6. For every interval $(l_1 l_2) \subseteq I$, we have $T((l_1, l_2)) = (r_1, r_2) \subseteq I$:

This is a simple consequence of Step 5 and of Remark 4.7.

Step 7. $b = T^{(1)}(\alpha) = T^{(1)}(\beta)$, $T^{(1)}(b_-) = \beta$ and $T^{(1)}(b_+) = \alpha$:

Note first that $T^{(1)}(b_-) = \beta$ and $T^{(1)}(b_+) = \alpha$ or else (α, a) and (a, β) will be mapped into a smaller interval

than I, which contradicts Remark 4.7. By Step 4, we have $T^{(1)}(\alpha) \geq T^{(1)}(\beta)$. It remains to prove that $b = T^{(1)}(\alpha) = T^{(1)}(\beta)$. If this does not hold, the alternatives are: $T^{(1)}(\alpha) \geq b > T^{(1)}(\beta)$, $T^{(1)}(\alpha) > T^{(1)}(\beta) \geq b$ and $T^{(1)}(\beta) < T^{(1)}(\alpha) \leq b$. We will only show the reasoning for the case $T^{(1)}(\alpha) \geq b > T^{(1)}(\beta)$, as the other two possibilities can be dealt with in a similar way. In this first case, we map (α, b) to $(T^{(1)}(\alpha), \beta)$ and (b, β) to $(\alpha, T^{(1)}(\beta))$. Therefore, $supp\, \gamma \subset I \times ((\alpha, T^{(1)}(\beta)) \cup (T^{(1)}(\alpha), \beta) \neq I \times I$, which contradicts Remark 4.7.

Step 8. $\boldsymbol{\mu([\alpha, a]) = \mu([a, \beta]) = \frac{1}{2}}$:

(α, a) is mapped into (a, β) and (a, β) is mapped into (α, a). Therefore

$$\gamma((\alpha, a) \times (a, \beta)) + \gamma((a, \beta) \times (\alpha, a)) = 1.$$

But

$$\gamma((\alpha, a) \times (a, \beta)) = \gamma((\alpha, a) \times I) = \mu((\alpha, a)),$$

as $\gamma((\alpha, a) \times (\alpha, a)) = 0$. Similarly,

$$\gamma((a, \beta) \times (\alpha, a)) = \mu((\alpha, a)).$$

It follows that $2\mu((\alpha, a)) = 1$ or $\mu((\alpha, a)) = \frac{1}{2}$. □

Theorem 4.9. *Assume that $\mu = \nu$ with density $\rho(x) > 0$ on $I = [\alpha, \beta]$, where $\alpha, \beta \in \mathbb{R} \cup \{\pm\infty\}$. Let $\mu_1(x) := \mu((\alpha, x))$, $\bar{\mu}_1 := \mu((x, a))$ for $x \in (\alpha, a)$, $\mu_2(x) := \mu((x, \beta))$ and $\bar{\mu}_2 := \mu((a, x))$, for $x \in (a, \beta)$. If $x \in (\alpha, a)$, we have $T^{(1)}(x) = \bar{\mu}_2^{-1}(\mu_1(x))$ and if $x \in (a, \beta)$, we have $T^{(1)}(x) = \bar{\mu}_1^{-1}(\mu_2(x))$.*

Proof. We will use the fact that $\mu \circ T^{(1)} = \mu$ to find $T^{(1)}$. Let $x \in (\alpha, a)$. Then due to the properties of $T^{(1)}$ from Theorem 4.8, it follows that $T^{(1)}((\alpha, x) = (a, T^{(1)}(x))$. Therefore,

$$\mu_1(x) = \mu((\alpha, x)) = \mu((a, T^{(1)}(x)) = \bar{\mu}_2(T^{(1)}(x)).$$

We know that $\rho(x) > 0$. Due to the fact that $T^{(1)}(x)$ is increasing on (α, a) and with $T^{(1)}(a_-) = \beta$, $\mu_2(T^{(1)}(x))$ is a a strictly increasing function. We can take inverses and have

$$T^{(1)}(x) = \bar{\mu}_2^{-1}(\mu_1(x)).$$

A similar reasoning holds for $x \in (a, \beta)$. □

4.2 Equal radially symmetric marginals in dimension d

We assume in this subsection that the marginals μ and ν are radially symmetric and in $\mathcal{P}(\mathbb{R}^d)$. As before, we suppose that the cost function is given by $c(x, y) = \ell(|x - y|) \geq 0$, with c and ℓ satisfying (A1)–(A4).

Theorem 4.10. *Suppose that $\mu = \nu$, with common density $\rho(x) = \lambda(|x|)$ for all $x \in supp\, \mu$. Moreover, we assume that $\rho(x) > 0$ for all $x \in supp\, \mu$. Then the optimal transport map $T^{(d)}$ has to be radially symmetric itself, that is*

$$T^{(d)}(x) = g(|x|)\frac{x}{|x|}, \quad x \in \mathbb{R}^d, \tag{4.1}$$

for some function $g : [0, \infty) \to \mathbb{R}$. Moreover $g \leq 0$, and g is an increasing function with $g(0_+) = -\infty$ and $g(+\infty) = 0$.

Proof. **Step 1. $\boldsymbol{T(Rx) = RT(x)}$ for all $\boldsymbol{R \in O(d)}$ and all $\boldsymbol{x \in supp\, \mu}$:** (Here $O(d)$ denotes the group of orthogonal $d \times d$ matrices):

Let γ_T be a minimizer of C on $\Gamma(\rho, \rho)$. Then $(R \times R)_\sharp \gamma$ is also a minimizer, for any $R \in O(n)$, since it belongs to $\Gamma(\rho, \rho)$ by the radial symmetry of ρ, and has the same cost C as γ_T by the invariance of the cost function $c(x, y)$ under $(x, y) \mapsto (R^{-1}x, R^{-1}y)$. Hence by uniqueness, $\gamma_T = (R \times R)_\sharp \gamma_T$. But an elementary calculation shows that the latter is equivalent to $T(z) = RT(R^{-1}z)$ for all $z \in supp\, \mu$. Left, respectively right hand side, evaluated on a set $A \times B$ give $\int \chi_A(x)\chi_B(T(x))\rho(x)dx$ respectively $\int \chi_A(Rx)\chi_B(RT(x))\rho(x)dx$.

A change of variables together with the radial symmetry of ρ shows that the latter expression equals $\int \chi_A(z)\chi_B(RT(R^{-1}z))\rho(z)dz$. Comparing with the former expression yields the assertion.

Step 2. T is radial and direction reversing, i.e. $T(x) = g(|x|)\frac{x}{|x|}$ for some $g \leq 0$:

Let e_1 be a fixed unit vector in $\mathbb{R}^d$ and $r > 0$. By Step 1, for all $R \in O(n)$ we have

$$T(Rre_1) = RT(re_1) = Rf(r)v(r) \text{ with } f(r) := |T(re_1)| \text{ and } v(r) := T(re_1)/|T(re_1)|. \tag{4.2}$$

Hence

$$I[T] = \int \ell(|x - T(x)|)\rho(x)\,dx = \int \ell\big(|x|e_1 - f(|x|)v(|x|)\big)\,\rho(x)\,dx.$$

But ℓ is by assumption strictly decreasing and $|re_1 - f(r)v(r)|$ is maximized among unit vectors $v(r)$ if and only if $v(r) = -e_1$. Hence, since T minimizes I, $v(r) = -e_1$. Substituting into (4.2) yields the assertion, with $g(r) = -f(r) = -|T(re_1)|$.

Step 3. g solves a one-dimensional mass transportation problem:

For any Borel map T on $\mathbb{R}^d$, abbreviate $I_\mu[T] := \int \ell(|x - T(x)|)\rho(x)dx$ (Monge functional with map T and equal marginals $\rho(x) = \lambda(|x|)$). If T is a radial map, i.e. of form $T(x) = g(|x|)\frac{x}{|x|}$ for some Borel $g : [0, \infty) \to \mathbb{R}$, and $\tilde{g}$ denotes the antisymmetric extension of g to $\mathbb{R}$, such that, in particular, $T(x, 0, 0) = (\tilde{g}(x), 0, 0)$ for all $x \in \mathbb{R}$, then using polar coordinates (with $|S^{d-1}|$ denoting the Hausdorff measure of the unit sphere in $\mathbb{R}^d$)

$$
\begin{aligned}
I_\rho[T] &= \int_{r=0}^{\infty} \ell(|r - g(r)|)|S^{d-1}|r^{d-1}\lambda(r)dr \\
&= \int_{s=-\infty}^{\infty} \ell(|s - \tilde{g}(s)|)\underbrace{\tfrac{1}{2}|S^{d-1}||s|^{d-1}\lambda(|s|)}_{=:\rho_1(s)}\,ds = I_{\rho_1}[T].
\end{aligned}
$$

Hence the d-dimensional Monge problem of minimizing I_ρ over radial maps is equivalent to the one-dimensional Monge problem of minimizing I_{ρ_1} over antisymmetric maps, and – because of Step 1 (with d=1 and $R = -I$) – to the one-dimensional Monge problem of minimizing I_{ρ_1} over arbitrary maps. It follows that the function g in (4.1), antisymmetrically extended to $\mathbb{R}$, is a minimizer of I_{ρ_1}. The asserted properties of g now follow immediately from Theorem 4.8 and the fact that ρ_1 (being symmetric) has median 0. □

Corollary 4.11. *Suppose that $\mu = \nu$ are as in Theorem 4.10. Let $t \in (0, \infty)$ and denote by*

$$F_1(t) = |S^{d-1}|\int_0^t \lambda(s)s^{d-1}ds \text{ and } F_2(-t) = |S^{d-1}|\int_t^\infty \lambda(s)s^{d-1}ds.$$

Then the function g in (4.1) *is given by*

$$g(t) = F_2^{-1}(F_1(t)).$$

Proof. We have already shown that g, antisymmetrically extended to $\mathbb{R}$, minimizes the one-dimensional functional I_{ρ_1}, with ρ_1 as in Step 3 above. The assertion is now a direct consequence of the representation formula given in Theorem 4.9. □

Example 4.12. (exponential radially symmetric distribution) Assume now that $\rho(x) = \frac{1}{Z}e^{-|x|}$ for $x = (x_{1,2}, x_3) \in \mathbb{R}^3$, where Z is the normalizing constant, that is, $Z = \int e^{-|x|}dx_1dx_2dx_3$. Then for $t \in (0, \infty)$, we have

$$F_1(t) = 1 - \left(1 + t + \frac{t^2}{2}\right)e^{-t}, \ F_2(-t) = e^{-t}\left(1 + t + \frac{t^2}{2}\right) \text{ and } g(t) = F_2^{-1}(F_1(t)).$$
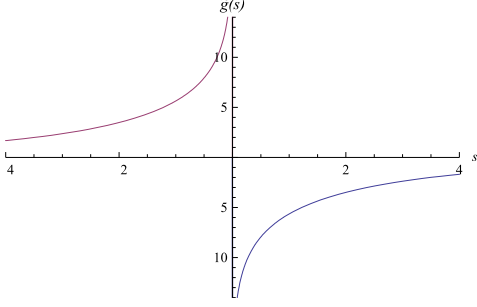
□

Figure 5: Optimal transport map T for the density $\rho(x) = const \times e^{-|x|}$. As shown in the text, T leaves lines through the origin invariant, e.g. $T(x,0,0) = (\tilde{g}(x),0,0)$ for all x, and the figure shows the function $\tilde{g}$.
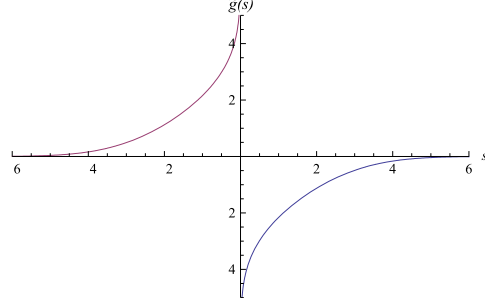
Figure 6: Optimal transport map T for the density $\rho(x) = const \times e^{-|x|^2/2}$.

Example 4.13. (normal radially symmetric distribution) Assume now that $\rho(x) = \frac{1}{Z}e^{-|x|^2/2}$ for $x \in \mathbb{R}^3$, where Z is the normalizing constant. Then for $t \in (0,\infty)$, we have

$$F_1(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-s^2/2} s^2 ds, \ F_2(-t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-s^2/2} s^2 ds \text{ and } g(t) = F_2^{-1}(F_1(t)).$$

□

5 Asymptotic exactness of the optimal transport functional in the semiclassical limit

Our goal in this section is to compare the *exact* quantum mechanical ground state energy to the *approximate* DFT ground state energy obtained by replacing V_{ee} by the optimal transportation functional. Recall from (2.3) and (2.4)) that the exact ground state energy of an N-electron system is defined as

$$E_0^{QM} = \inf_{\Psi \in \mathcal{A}} \left\{ T[\Psi] + V_{ne}[\rho^\Psi] + V_{ee}[\rho_2^\Psi] \right\} \tag{5.1}$$

and the approximate ground state energy is (recall the DFT formalism in (2.15), (2.16))

$$E_0^{DFT-OT} = \inf_{\Psi \in \mathcal{A}} \left\{ T[\Psi] + V_{ne}[\rho^\Psi] + E_{OT}[\rho^\Psi] \right\} = \inf_{\rho \in \mathcal{R}} \left\{ T_{QM}[\rho] + V_{ne}[\rho] + E_{OT}[\rho] \right\}. \tag{5.2}$$

In the above, ρ_2^Ψ and ρ^Ψ denote the pair density, respectively the single particle density of Ψ (see (2.5), (2.6) and (2.7)), and E_{OT} is the optimal transportation functional with Coulomb cost from (3.3)

$$E_{OT}[\rho] = \inf_{\gamma \in \Gamma(\rho,\rho)} \int_{\mathbb{R}^6} \frac{1}{|x-y|} d\gamma(x,y). \tag{5.3}$$

Due to the fact that ρ^Ψ is the marginal of ρ_2^Ψ, we have

$$V_{ee}[\rho_2^\Psi] \geq E_{OT}[\rho^\Psi] \text{ for every } \Psi \in \mathcal{A}. \tag{5.4}$$

Taking the infimum over Ψ gives

Theorem 5.1. *For every N, and any potential $v \in L^{3/2}(\mathbb{R}^3) + L^\infty(\mathbb{R}^3)$, the density functional with electron-electron interaction energy given by the mass transportation functional is a rigorous lower bound:*

$$E_0^{QM} \geq E_0^{DFT-OT}.$$

Now consider the kinetic energy functional $T[\Psi]$ from Section 2.1 with physical constants inserted,

$$T_\hbar[\Psi] = \frac{\hbar^2}{2m} \int ... \int \sum_{i=1}^{N} |\nabla_{x_i} \Psi(x_1, s_1, .., x_N, s_N)|^2 dz_1..dz_N. \tag{5.5}$$

Here m is the mass of the electron and $\hbar$ is Planck's constant h divided by 2π. We are interested in the limit $\hbar \to 0$ (*semiclassical limit*). Define now $E_0^{QM}(\bar{h})$ and $E_0^{DFT-OT}(\bar{h})$ as in (5.1) and (5.2), but with $T[\Psi]$ replaced by $T_\hbar[\Psi]$. Note now that the statement

$$\frac{E_0^{QM}(\hbar)}{E_0^{DFT-OT}(\hbar)} \to 1 \text{ as } \hbar \to 0 \tag{5.6}$$

is in general false. The reason is that when $\hbar$ gets small, then (for typical V_{ne}) the ground state densities of both models contract, and the approximation is not uniformly good on families of contracting densities.

This has nothing particular to do with the use of E_{OT}, but (5.6) fails for *any* DFT model (2.15) whose electron interaction functional $\tilde{V}_{ee}$ has the correct scaling under dilations,

$$\tilde{V}_{ee}[\alpha^3 \rho(\alpha \cdot)] = \alpha \tilde{V}_{ee}[\rho(\cdot)], \tag{5.7}$$

such as the mean field functional (2.18) or the local density approximation (2.23). A counterexample is already given by atoms, $v(x) = -Z/|x|$ (eq. (2.1) in Section 2, with $\alpha = 1$). Very remarkably, in this case

$$E_0^{QM}(\hbar) = \frac{E_0^{QM}(1)}{\hbar^2} \text{ and } E_0^{DFT-OT}(\hbar) = \frac{E_0^{DFT-OT}(1)}{\hbar^2}, \tag{5.8}$$

and hence the quotient $E_0^{QM}(\hbar)/E_0^{DFT-OT}(\hbar)$ is independent of $\hbar$! To prove this, use that the four functionals involved, T, V_{ne}, V_{ee}, and $\tilde{V}_{ee}$, all have a definite scaling behaviour with respect to dilations. For a given $\Psi \in \mathcal{A}$, consider its L^2-norm-preserving dilation

$$\Psi_\hbar(x_1, .., x_N) := (\hbar^2)^{-3N/2} \Psi(\hbar^{-2} x_1, .., \hbar^{-2} x_N).$$

Then (with $T_\hbar$ being the kinetic energy with prefactor $\hbar^2/2m$ from (5.5))

$$(T_\hbar + V_{ne} + V_{ee})[\Psi_\hbar] = \hbar^{-2}(T_1 + V_{ne} + V_{ee})[\Psi].$$

Taking the infimum over Ψ gives the first assertion in (5.8). The second assertion follows analogously after noting that

$$\rho^{\Psi_\hbar}(x) = (\hbar^2)^{-3} \rho^\Psi(\hbar^{-2} x), \quad E_{OT}[\rho^{\Psi_\hbar}] = \hbar^{-2} E_{OT}[\rho^\Psi]$$

(or more generally $\tilde{V}_{ee}[\rho^{\Psi_\hbar}] = \hbar^{-2} \tilde{V}_{ee}[\rho^{\Psi_\hbar}]$ for every $\tilde{V}_{ee}$ satisfying (5.7)).

What we can prove is the following "pointwise" statement in which we only minimize out Ψ at fixed ρ:

Theorem 5.2. *Let $N = 2$. Then*

$$\lim_{\hbar \to 0} F_{HK}[\rho] = E_{OT}[\rho] \text{ for every } \rho \in \mathcal{R}.$$

Here $\mathcal{R}$ is the natural class of densities given by the image of $\mathcal{A}$ under the map $\Psi \mapsto \rho$ ($\mathcal{R}$ is defined in (2.11)), and F_{HK} is the Hohenberg-Kohn functional (2.13) with kinetic energy functional $T_\hbar$ in place of T.

This case for $N = 2$ already contains the main analytic issue, namely that the optimal transport measure γ is singular and so its square root fails to be in L^2 and fails to have an L^2 gradient. But the case allows to avoid the quantum mechanical issues of spin and antisymmetry, which would enter on top of this when $N \geq 3$.

An interesting challenge raised by the above theorem is to derive higher order corrections to E_{OT} in the semiclassical limit.

5.1 Re-instating the constraint

In order to show that $\lim_{\hbar \to 0} F_{HK}[\rho] = E_{OT}[\rho]$, we will need to make modifications to the optimal plan γ which yields $E_{OT}[\rho]$, since any Ψ which represents γ has $T[\Psi] = +\infty$. Therefore we cannot use these Ψ's as trial states in the variational principle for $F_{HK}[\rho]$. Hence, we will need to modify the optimal γ. *But the modifications that one would like to use, e.g. smoothing, lead to modified marginals.*

Hence we need to be able to control the change in $E_{OT}[\rho]$ induced by a small change in ρ. This is not trivial, due to the rigid infinite-dimensional constraint in the variational principle for E_{OT} that the trial states must have marginals exactly equal to ρ, and is achieved in Theorem 5.3 below.

The main technical idea behind this theorem is the following construction to "re-instate the constraint", i.e. to deform a given trial plan into a nearby one with prescribed marginals. Suppose we are given an arbitrary transport plan $\gamma_{A \to A}$ with equal marginals ρ_A, and an arbitrary second density ρ_B. We assume that $\rho_A, \rho_B \in L^1 \cap L^3(\mathbb{R}^3)$, $\rho_A, \rho_B \geq 0$, and $\int_{\mathbb{R}^3} \rho_A(x)\,dx = \int_{\mathbb{R}^3} \rho_B(x)\,dx = 1$. Our interest is in the case when ρ_B is near ρ_A, but the construction works for general ρ_B.

Intuitively, the plan $\gamma_{B \to B}$ with equal marginals ρ_B we have in mind is the following.

- First transport ρ_B to ρ_A by a transport plan $\gamma_{B \to A}$ that does not move much mass around when ρ_B is close to ρ_A.

- Then apply the plan $\gamma_{A \to A}$.

- Finally transport ρ_A back to ρ_B.

First, let us construct a suitable plan $\gamma_{B \to A}$. Let $f(x) := \min\{\rho_A(x), \rho_B(x)\}$. Take $f_A := (\rho_A - f)_+$ and $f_B := (\rho_B - f)_+$. Then $\rho_A = f + f_A$ and $\rho_B = f + f_B$.

On f we "do nothing", i.e. we let:

$$\gamma_{f \to f}(x, y) = f(x)\delta_x(y).$$

On f_A we transport to f_B via a convenient plan which allows simple estimates (note that $\int_{\mathbb{R}^3} f_A(x)\,dx = \int_{\mathbb{R}^3} f_B(x)\,dx$, due to the fact that $\int_{\mathbb{R}^3} \rho_A(x)\,dx = \int_{\mathbb{R}^3} \rho_B(x)\,dx$):

$$\gamma_{f_A \to f_B}(x, y) = \frac{f_A(x)f_B(y)}{\int_{\mathbb{R}^3} f_B(x)\,dx}.$$

We then set

$$\gamma_{A \to B}(x, y) = \gamma_{f \to f}(x, y) + \gamma_{f_A \to f_B}(x, y) = f(x)\delta_x(y) + \frac{f_A(x)f_B(y)}{\int_{\mathbb{R}^3} f_B(y)\,dy}. \tag{5.9}$$

Note that $\int_{\mathbb{R}^3} \gamma_{A \to B}(x, y)\,dy = f(x) + f_A(x) = \rho_A(x)$ and $\int \gamma_{A \to B}(x, y)\,dx = f(y) + f_B(y)\frac{\int_{\mathbb{R}^3} f_A(x)\,dx}{\int_{\mathbb{R}^3} f_B(x)\,dx} = f(y) + f_B(y) = \rho_B(y)$, as required. We will also need the reverse plan

$$\gamma_{B \to A}(x, y) = f(x)\delta_x(y) + \frac{f_B(x)f_A(y)}{\int_{\mathbb{R}^3} f_A(y)\,dy}, \tag{5.10}$$

which satisfies $\int_{\mathbb{R}^3} \gamma_{B \to A}(x, y)\,dy = \rho_B(x)$ and $\int_{\mathbb{R}^3} \gamma_{B \to A}(x, y)\,dx = \rho_A(y)$. Finally we introduce the combined plan

$$P(x, w) := \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \gamma_{B \to A}(x, y) \frac{\chi_{\rho_A > 0}(y)}{\rho_A(y)} \gamma_{A \to A}^{opt}(y, z) \frac{\chi_{\rho_A > 0}(z)}{\rho_A(z)} \gamma_{A \to B}(z, w)\,dy\,dz. \tag{5.11}$$

We now claim that

$$\int_{\mathbb{R}^3} P(x, w)\,dw = \rho_B(x) \quad \text{and} \quad \int_{\mathbb{R}^3} P(x, w)\,dx = \rho_B(w). \tag{5.12}$$

To prove the first claim, we begin by integrating over w. This yields

$$\int_{\mathbb{R}^3} P(x, w)\,dw = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \gamma_{B \to A}(x, y) \frac{\chi_{\rho_A > 0}(y)}{\rho_A(y)} \gamma_{A \to A}^{opt}(y, z) \frac{\chi_{\rho_A > 0}(z)}{\rho_A(z)} \rho_A(z)\,dy\,dz.$$

29

Noting that $\frac{\chi_{\rho_A>0}(z)}{\rho_A(z)}\rho_A(z) = 1$ whenever $\gamma^{opt}_{A\to A}(y,z) > 0$ and recalling that $\int_{\mathbb{R}^3}\gamma^{opt}_{A\to A}(y,z)\,dz = \rho_A(y)$, integrating over z yields

$$\int_{\mathbb{R}^3} P(x,w)\,dw = \int_{\mathbb{R}^3}\gamma_{B\to A}(x,y)\frac{\chi_{\rho_A>0}(y)}{\rho_A(y)}\rho_A(y)\,dy.$$

Since $\frac{\chi_{\rho_A>0}(y)}{\rho_A(y)}\rho_A(y) = 1$ whenever $\gamma_{B\to A}(x,y) > 0$ and since $\int_{\mathbb{R}^3}\gamma_{B\to A}(x,y)\,dy = \rho_B(x)$, the right hand side becomes equal to $\rho_B(x)$ after integrating over y. The second marginal condition can be derived analogously.

5.2 Continuity of the optimal transport functional

By combining the techique introduced above with appropriate estimates, we are able to control the change in $E_{OT}[\rho]$ induced by a small change in ρ.

Theorem 5.3. *There exists a $c_* > 0$ such that for any $\rho_A,\rho_B \in L^1\cap L^3(\mathbb{R}^3)$, with $\rho_A,\rho_B \geq 0$ and $\int_{\mathbb{R}^3}\rho_A(x)\,dx = \int_{\mathbb{R}^3}\rho_A(x)\,dx = 1$, the optimal transport functional with Coulomb cost (5.3) satisfies*

$$|E_{OT}[\rho_A] - E_{OT}[\rho_B]| \leq c_* \left(||\rho_A||_{L^1\cap L^3(\mathbb{R}^3)} + ||\rho_B||_{L^1\cap L^3(\mathbb{R}^3)}\right)||\rho_A - \rho_B||_{L^1(\mathbb{R}^3)\cap L^3(\mathbb{R}^3)},$$

where $||\rho_i||_{L^1\cap L^3(\mathbb{R}^3)} := \max\{||\rho_i||_{L^1(\mathbb{R}^3)}, ||\rho_i||_{L^3(\mathbb{R}^3)}\}$ for $i\in\{A,B\}$.

Proof. Fix arbitrarily two marginals $\rho_A,\rho_B \in L^1\cap L^3(\mathbb{R}^3)$, with $\rho_A,\rho_B \geq 0$ and $\int_{\mathbb{R}^3}\rho_A(x)\,dx = \int_{\mathbb{R}^3}\rho_A(x)\,dx = 1$. Let $\gamma_{A\to A} = \gamma^{opt}_{A\to A}$ be an optimal transport plan of C subject to the constraint that $\gamma_{A\to A}$ has equal marginals ρ_A. The main idea is to consider the associated plan $\gamma_{B\to B} = P$ introduced in (5.11) and show that

$$C(\gamma_{B\to B}) \leq C(\gamma^{opt}_{A\to A}) + c_* \left(||\rho_A||_{L^1\cap L^3(\mathbb{R}^3)} + ||\rho_B||_{L^1\cap L^3(\mathbb{R}^3)}\right)||\rho_A - \rho_B||_{L^1(\mathbb{R}^3)\cap L^3(\mathbb{R}^3)}. \quad (5.13)$$

By the variational principle for $E_{OT}[\rho_B]$ and the optimality of $\gamma^{opt}_{A\to A}$ this implies

$$E_{OT}[\rho_B] \leq E_{OT}[\rho_A] + c_* \left(||\rho_A||_{L^1\cap L^3(\mathbb{R}^3)} + ||\rho_B||_{L^1\cap L^3(\mathbb{R}^3)}\right)||\rho_A - \rho_B||_{L^1(\mathbb{R}^3)\cap L^3(\mathbb{R}^3)},$$

as required.

Step 1: $C(P) \leq C(\gamma^{opt}_{A\to A}) + 3M$ with $M = \sup_{y\in\mathbb{R}^3}\int_{\mathbb{R}^3} c(y,w)f_B(w)\,dw$:

By substituting the expressions (5.9) and (5.10) into (5.11), we get

$$C(P) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} c(x,w)\left[f(x)\delta_x(y) + \frac{f_B(x)f_A(y)}{\int_{\mathbb{R}^3} f_A(y)\,dy}\right]\frac{\chi_{\rho_A>0}(y)}{\rho_A(y)}\gamma^{opt}_{A\to A}(y,z)\frac{\chi_{\rho_A>0}(z)}{\rho_A(z)}$$
$$\left[f(z)\delta_z(w) + \frac{f_A(z)f_B(w)}{\int_{\mathbb{R}^3} f_B(w)\,dw}\right]dx\,dy\,dz\,dw. \quad (5.14)$$

This is a sum of four terms, which arise by picking one term from each square bracket and carrying out the integrals over the delta functions:

$$W_1 = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} c(y,z)f(y)\frac{\chi_{\rho_A>0}(y)}{\rho_A(y)}\gamma^{opt}_{A\to A}(y,z)\frac{\chi_{\rho_A>0}(z)}{\rho_A(z)}f(z)\,dy\,dz,$$

$$W_2 = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\int_{R^3} c(y,w)f(y)\frac{\chi_{\rho_A>0}(y)}{\rho_A(y)}\gamma^{opt}_{A\to A}(y,z)\frac{\chi_{\rho_A>0}(z)}{\rho_A(z)}\frac{f_A(z)f_B(w)}{\int_{\mathbb{R}^3} f_B(w)\,dw}\,dy\,dz\,dw,$$

$$W_3 = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} c(x,z)\frac{f_B(x)f_A(y)}{\int_{\mathbb{R}^3} f_A(y)\,dy}\frac{\chi_{\rho_A>0}(y)}{\rho_A(y)}\gamma^{opt}_{A\to A}(y,z)\frac{\chi_{\rho_A>0}(z)}{\rho_A(z)}f(z)\,dx\,dy\,dz$$

and

$$W_4 = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} c(x,w)\frac{f_B(x)f_A(y)}{\int_{\mathbb{R}^3} f_A(y)\,dy}\frac{\chi_{\rho_A>0}(y)}{\rho_A(y)}\gamma^{opt}_{A\to A}(y,z)\frac{\chi_{\rho_A>0}(z)}{\rho_A(z)}\frac{f_A(z)f_B(w)}{\int_{\mathbb{R}^3} f_B(w)\,dw}\,dx\,dy\,dz\,dw.$$

Next, we will estimate each of these four terms. For the first term, we use the simple estimate that $f\chi_{\rho_A > 0} \le \rho_A$, which gives

$$W_1 \le \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} c(y, z) \gamma_{A \to A}^{opt}(y, z) \, dy \, dz = C(\gamma_{A \to A}^{opt}).$$

For the second term, we estimate $\int_{\mathbb{R}^3} c(y, w) \, dw$ by the constant M defined in Step 2 and $f(y) \frac{\chi_{\rho_A > 0}(y)}{\rho_A(y)}$ by 1, and we get

$$
\begin{aligned}
W_2 &\le \frac{M}{\int_{\mathbb{R}^3} f_B(w) \, dw} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \gamma_{A \to A}^{opt}(y, z) \frac{\chi_{\rho_A > 0}(z)}{\rho_A(z)} f_A(z) \, dy \, dz \\
&= \frac{M}{\int_{\mathbb{R}^3} f_B(w) \, dw} \int_{\mathbb{R}^3} \chi_{\rho_A > 0}(z) f_A(z) \, dz = M.
\end{aligned}
$$

Analogously, by the change of variables $(y, z, w) \mapsto (z, y, x)$ we have

$$W_3 = W_2 \le M.$$

Finally, to bound W_4 we estimate $\int_{\mathbb{R}^3} c(x, w) f_B(w) \, dw$ by M and $f_A(y) \frac{\chi_{\rho_A > 0}(y)}{\rho_A(y)}$ by 1, and we obtain

$$
\begin{aligned}
W_4 &\le \frac{M}{\left(\int_{\mathbb{R}^3} f_B(w) \, dw\right)^2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} f_B(x) \gamma_{A \to A}^{opt}(y, z) f_A(z) \frac{\chi_{\rho_A > 0}(z)}{\rho_A(z)} \, dx \, dy \, dz \\
&\le \frac{M}{\left(\int_{\mathbb{R}^3} f_B(w) \, dw\right)^2} \left[\int_{\mathbb{R}^3} f_B(x) \, dx\right] \left[\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \gamma_{A \to A}^{opt}(y, z) \, dy f_A(z) \frac{\chi_{\rho_A > 0}(z)}{\rho_A(z)} \, dz\right] = M.
\end{aligned}
$$

Plugging the above bounds for W_1, W_2, W_3, W_4 into (5.14) yields the assertion.

Step 2. For $g \in L^1 \cap L^2(\mathbb{R}^3)$, we have:

$$\sup_{x \in \mathbb{R}^3} \left| \int_{\mathbb{R}^3} \frac{1}{|x - y|} g(y) \, dy \right| \le c_0 \max\{||g||_{L^1(\mathbb{R}^3)}, ||g||_{L^3(\mathbb{R}^3)}\}, \tag{5.15}$$

with $c_0 = 2 \left(\frac{8\pi}{3}\right)^{1/3}$.

To prove this, we split $\frac{1}{|x - y|}$ into a short-range and a long-range part,

$$\frac{1}{|z|} = \frac{\chi_{|z| < a}}{|z|} + \frac{\chi_{|z| \ge a}}{|z|} =: h_s(z) + h_l(z),$$

with the obvious definitions for h_s and h_l, and with cut-off parameter $a > 0$ to be chosen later. Note that $h_s \in L^{3/2}(\mathbb{R}^3)$ and $h_l \in L^{\infty}(\mathbb{R}^3)$. By Hölder's inequality we have

$$
\begin{aligned}
\left| \int_{\mathbb{R}^3} \frac{1}{|x - y|} g(y) \, dy \right| &= \left| \int_{\mathbb{R}^3} h_s(x - y) g(y) \, dy + \int_{\mathbb{R}^3} h_l(x - y) g(y) \, dy \right| \le ||h_s||_{L^{3/2}(\mathbb{R}^3)} ||g||_{L^3(\mathbb{R}^3)} \\
&+ ||h_l||_{L^{\infty}(\mathbb{R}^3)} ||g||_{L^1(\mathbb{R}^3)} \le ||g||_{L^1 \cap L^3(\mathbb{R}^3)} \left(||h_s||_{L^{3/2}(\mathbb{R}^3)} + ||h_l||_{L^{\infty}(\mathbb{R}^3)}\right).
\end{aligned}
$$

Explicitly,

$$||h_s||_{L^{3/2}(\mathbb{R}^3)} + ||h_l||_{L^{\infty}(\mathbb{R}^3)} = \left(4\pi \int_0^a r^2 \frac{1}{r^{3/2}} \, dr\right)^{2/3} + \frac{1}{a} = \left(\frac{8\pi}{3}\right)^{2/3} a + \frac{1}{a}.$$

Minimizing over a in the above gives $a = \left(\frac{8\pi}{3}\right)^{-1/3}$, leading to the value of c_0 in the assertion.

Step 3. Putting it all together:

By Steps 1 and 2 we have

$$C(P) \le C(\gamma_{A \to A}^{opt}) + 3c_0 ||f_B||_{L^1 \cap L^3(\mathbb{R}^3)}.$$

But $0 \le f_B \le |\rho_A - \rho_B|$, so $||f_B||_{L^1 \cap L^3(\mathbb{R}^3)} \le ||\rho_A - \rho_B||_{L^1 \cap L^3(\mathbb{R}^3)}$. This establishes (5.13) and Theorem 5.3, with $c_* = 3c_0 = 6\left(\frac{8\pi}{3}\right)^{1/3}$. $\qquad \square$

31

5.3 Finiteness of kinetic energy

In this section we investigate the behaviour of derivatives of the combined plan $\gamma_{B\to B} = P$ introduced in (5.11) when the original plan $\gamma_{A\to A}$ is differentiable.

Recall that $\gamma_{A\to A}$ is a transport plan of C subject to the constraint of equal marginals ρ_A, $\gamma_{A\to B}$ was defined in (5.9), and $\gamma_{B\to A}$ is the reverse plan (5.10). Unlike in the previous section, here $\gamma_{A\to A}$ does not need to be optimal. Due to the fact that wave functions correspond, up to integrating out variables, to square roots of pair densities, and the kinetic energy of a wave function is $\frac{\hbar^2}{2}\int |\nabla\Psi|^2$, we have to show that $\nabla\sqrt{\gamma_{B\to B}} \in L^2$, in order to be able to construct an admissible trial function with pair density γ in the variational definition of the Hohenberg-Kohn density functional F_{HK}. The following result gives hypotheses under which this is true. Before stating the result we introduce the following notion which we call *strong positivity*.

Definition 5.4. *A transportation plan $\gamma \in \mathcal{P}(\mathbb{R}^{2d})$ with marginals μ, $\nu \in \mathcal{P}(\mathbb{R}^d)$ is called strongly positive if there exists a constant $\beta > 0$ such that*

$$\gamma \geq \beta\mu \otimes \nu.$$

We note that strong positivity implies, in particular, that the support of γ is the product of the supports of its marginals, $supp\ \gamma = supp\ \mu \times supp\ \nu$.

Theorem 5.5. *Suppose that $\rho_A, \rho_B \geq 0$, $\sqrt{\rho_A}, \sqrt{\rho_B} \in H^1(\mathbb{R}^3)$, and assume that $\gamma_{A\to A}$ belongs to the set $\mathcal{M}_+(\mathbb{R}^6)$ (see Section 3) and has equal marginals ρ_A.*

  (i) $\sqrt{\gamma_{A\to A}^{opt}} \in H^1(\mathbb{R}^3)$ *(smoothness);*

  (ii) $\gamma_{A\to A}^{opt} \geq \beta\rho_A \otimes \rho_A$ *for some constant $\beta > 0$ (strong positivity).*

Then the plan $\gamma_{B\to B} = P$ defined in (5.11) satisfies $\sqrt{P} \in H^1(\mathbb{R}^6)$.

Proof. Plugging formula (5.10) for $\gamma_{B\to A}$ into (5.11) and using that $\rho_A, \rho_B \geq 0$ and that $\int_{\mathbb{R}^3} f_A(y)\,dy = \int_{\mathbb{R}^3} f_B(y)\,dy$, we have

$$P(x,w) = \frac{f(x)}{\rho_A(x)}\int_{\mathbb{R}^3}\gamma_{A\to A}^{opt}(x,z))\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}\,dz + \frac{f_B(x)}{\int_{\mathbb{R}^3} f_B(y)\,dy}\int_{\mathbb{R}^3}\int_{R^3} f_A(y)\frac{\gamma_{A\to A}^{opt}(y,z)}{\rho_A(y)\rho_A(z)}\gamma_{A\to B}(z,w)\,dy\,dz.$$

Consequently,

$$\nabla\sqrt{P(x,w)} = \frac{1}{2\sqrt{P(x,w)}}$$

$$\left[\frac{\nabla f(x)}{\rho_A(x)}\int_{\mathbb{R}^3}\gamma_{A\to A}^{opt}(x,z))\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}\,dz - \frac{f(x)}{\rho_A(x)}\frac{\nabla\rho_A(x)}{\rho_A(x)}\int_{\mathbb{R}^3}\gamma_{A\to A}^{opt}(x,z))\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}\,dz\right.$$

$$\left.+\frac{f(x)}{\rho_A(x)}\int_{\mathbb{R}^3}\nabla_x\gamma_{A\to A}^{opt}(x,z))\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}\,dz + \frac{\nabla f_B(x)}{\int_{\mathbb{R}^3} f_B(y)\,dy}\int_{\mathbb{R}^3}\int_{R^3} f_A(y)\frac{\gamma_{A\to A}^{opt}(y,z)}{\rho_A(y)\rho_A(z)}\gamma_{A\to B}(z,w)\,dy\,dz\right]$$

$$= \ :\tilde{W}_1 + \tilde{W}_2 + \tilde{W}_3 + \tilde{W}_4,$$

with the obvious definitions for $\tilde{W}_1, \tilde{W}_2, \tilde{W}_3$ and $\tilde{W}_4$. We have to show that $\int_{\mathbb{R}^3} |\tilde{W}_i(\cdot, w)|^2\,dw \in L^1(\mathbb{R}^3)$ for $i = 1, \ldots, 4$. To estimate the first two terms, we use the following lower bound on P which neglects the contribution from $f_B(x)$ in $P(x,w)$.

$$P(x,w) \geq \frac{f(x)}{\rho_A(x)}\int_{\mathbb{R}^3}\gamma_{A\to A}^{opt}(x,z))\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}\,dz =: \frac{f(x)}{\rho_A(x)}g(x,w). \tag{5.16}$$

It follows that

$$|\tilde{W}_1(x,w)| \leq \frac{1}{2}\sqrt{\frac{\rho_A(x)}{f(x)g(x,w)}}\frac{\nabla f(x)}{\rho_A(x)}g(x,w) \quad\text{and}\quad |\tilde{W}_2(x,w)| \leq \frac{1}{2}\sqrt{\frac{\rho_A(x)}{f(x)g(x,w)}}\left(-\frac{f(x)}{\rho_A(x)}\frac{\nabla\rho_A(x)}{\rho_A(x)}\right)g(x,w)$$

32

and hence

$$|\tilde{W}_1(x,w)|^2 \leq \frac{1}{4}\frac{|\nabla f(x)|^2}{f(x)\rho_A(x)}g(x,w) \quad \text{and} \quad |\tilde{W}_2(x,w)|^2 \leq \frac{1}{4}\frac{|\nabla \rho_A(x)|^2 f(x)}{\rho_A(x)^3}g(x,w).$$

Next, due to $\int_{R^3} \gamma_{A\to B}(z,w)dw = \rho_A(z)$, we have that

$$\int_{\mathbb{R}^3} g(x,w)\, dw = \int_{\mathbb{R}^3}\frac{\gamma^{opt}_{A\to A}(x,z)}{\rho_A(z)}\int_{\mathbb{R}^3}\gamma_{A\to B}(z,w)\, dw = \int_{\mathbb{R}^3}\gamma^{opt}_{A\to A}(x,z)\, dz = \rho_A(x).$$

Consequently, using the fact that $|\nabla\sqrt{a}|^2 = \frac{1}{4}\frac{|\nabla a|^2}{a}$ for any function a, we have

$$\int_{\mathbb{R}^3}|\tilde{W}_1(x,w)|dw \leq \frac{1}{4}\frac{|\nabla f(x)|}{f(x)} = |\nabla\sqrt{f}|^2 \quad \text{and} \quad \int_{\mathbb{R}^3}|\tilde{W}_2(x,w)|dw \leq \frac{1}{4}\frac{f(x)}{\rho_A(x)}\frac{|\nabla \rho_A(x)|^2}{\rho_A(x)} \leq |\nabla\sqrt{\rho_A}|^2, \quad (5.17)$$

where in the last inequality we have used $f \leq \rho_A$. Since $\sqrt{f} = \min\{\sqrt{\rho_A},\sqrt{\rho_B}\}$, and $\sqrt{\rho_A},\sqrt{\rho_B} \in H^1(\mathbb{R}^3)$, by a standard fact concerning Sobolev functions we have $\sqrt{f} \in H^1(\mathbb{R}^3)$ and

$$\nabla\sqrt{f} = \chi_{\rho_A > \rho_B}\nabla\sqrt{\rho_B} + \chi_{\rho_A \leq \rho_B}\nabla\sqrt{\rho_A} \quad \text{a.e.}.$$

Consequently,

$$\int_{\mathbb{R}^3}|\tilde{W}_1(x,w)dw \leq |\nabla\sqrt{f}|^2 \leq |\nabla\sqrt{\rho_B}|^2 + |\nabla\sqrt{\rho_A}|^2. \quad (5.18)$$

Next we analyze $\tilde{W}_3$. To this end, we make use of the identity

$$\nabla_x\gamma^{opt}_{A\to A}(x,z) = \nabla_x\left(\sqrt{\gamma_{A\to A}(x,z)}\right)^2 = 2\sqrt{\gamma^{opt}_{A\to A}(x,z)}\,\nabla_x\sqrt{\gamma^{opt}_{A\to A}(x,z)}.$$

Together with (5.16) this yields

$$|\tilde{W}_3| \leq \sqrt{\frac{\rho_A(x)}{f(x)g(x,w)}}\frac{f(x)}{\rho_A(x)}\left|\int_{\mathbb{R}^3}\sqrt{\gamma^{opt}_{A\to A}(x,z)}\,\nabla_x\sqrt{\gamma^{opt}_{A\to A}(x,z)}\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}\,dz\right|.$$

To estimate the integral over z in the formula above, we write

$$\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)} = \sqrt{\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}}\sqrt{\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}},$$

group one of these factors with $\sqrt{\gamma^{opt}_{A\to A}(x,z)}$ and one with $\nabla_x\sqrt{\gamma^{opt}_{A\to A}(x,z)}$, and apply the Cauchy-Schwarz inequality. This yields

$$|\tilde{W}_3| \leq \sqrt{\frac{f(x)}{\rho_A(x)g(x,w)}}\sqrt{g(x,w)}\sqrt{\int_{\mathbb{R}^3}\left|\nabla_x\sqrt{\gamma^{opt}_{A\to A}(x,z)}\right|^2\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}\,dz}$$

and hence

$$|\tilde{W}_3|^2 \leq \frac{f(x)}{\rho_A(x)}\int_{\mathbb{R}^3}\left|\nabla_x\sqrt{\gamma^{opt}_{A\to A}(x,z)}\right|^2\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}\,dz.$$

Integrating over w and using that $\int_{\mathbb{R}^3}\frac{\gamma_{A\to B}(z,w)}{\rho_A(z)}\,dw = 1$ gives

$$\int_{\mathbb{R}^3}|\tilde{W}_3(x,w)|^2\,dw \leq \frac{f(x)}{\rho_A(x)}\int_{\mathbb{R}^3}\left|\nabla_x\sqrt{\gamma^{opt}_{A\to A}(x,z)}\right|^2\,dz \leq \int_{\mathbb{R}^3}\left|\nabla_x\sqrt{\gamma^{opt}_{A\to A}(x,z)}\right|^2\,dz. \quad (5.19)$$

Finally for $\tilde{W}_4$ is is natural to use a different lower bound for P than the one in (5.16), obtained by neglecting the first instead of the second term in $P(x,w)$.

$$P(x,w) \geq \frac{f_B(x)}{\int_{\mathbb{R}^3}f_B(y)\,dy}\int_{\mathbb{R}^3}\int_{R^3}f_A(y)\frac{\gamma^{opt}_{A\to A}(y,z)}{\rho_A(y)\rho_A(z)}\gamma_{A\to B}(z,w)\,dy\,dz =: \frac{f_B(x)}{\int_{\mathbb{R}^3}f_B(y)\,dy}\tilde{g}(w). \quad (5.20)$$

33

Substituting this estimate into the definition for $\tilde{W}_4$ immediately gives

$$|\tilde{W}_4| \leq \frac{1}{2} \frac{|\nabla f_B(x)|}{\sqrt{f_B(x)}} \sqrt{\frac{\tilde{g}(w)}{\int_{\mathbb{R}^3} f_B(y)\, dy}}$$

and, after squaring, integrating over w, and using $\int_{\mathbb{R}^3} \tilde{g}(w)\, dw = \int_{\mathbb{R}^3} f_B(y)\, dy$, we get

$$\int_{\mathbb{R}^3} |\tilde{W}_4(x,w)|\, dw \leq \frac{1}{4} \frac{|\nabla f_B(x)|^2}{f_B(x)} = |\nabla \sqrt{f_B}|^2. \qquad (5.21)$$

But unlike the analogous bounds on $\tilde{W}_1, \tilde{W}_2, \tilde{W}_3$, this estimate is insufficient to infer $\tilde{W}_4 \in H^1(\mathbb{R}^6)$ since $\sqrt{\rho_A}, \sqrt{\rho_B} \in H^1(\mathbb{R}^3)$ do **not** imply that the function

$$\sqrt{f_B} = \sqrt{\chi_{\rho_B > \rho_A}(\rho_B - \rho_A)}$$

belongs to H^1. In fact, even when $\sqrt{\rho_A}, \sqrt{\rho_B}$ are positive and belong to C^∞, $\sqrt{f}$ need not be in H^1_{loc}.

Example 5.6. Let $\rho_A(x) = (1 - x + x^2)e^{-x^2}$ and $\rho_B(x) = (1 + x + x^2)e^{-x^2}$. Because $1 \pm x + x^2 \geq \frac{1}{2}(1 + x^2)$ is bounded away from zero, we have $\sqrt{\rho_A}, \sqrt{\rho_B} \in H^1(\mathbb{R})$, but $\sqrt{\chi_{\rho_B > \rho_A}(\rho_B - \rho_A)} = \chi_{(0,\infty)}(x)\sqrt{2x}e^{-x^2/2} \notin H^1(\mathbb{R})$ since $|\nabla \sqrt{\chi_{\rho_B > \rho_A}(\rho_B - \rho_A)}|^2 = \chi_{(0,\infty)}(x) \left(\frac{1}{2x} - 2x + 2x^3\right) e^{-x^2} \notin L^1(\mathbb{R})$. $\qquad \square$

Note that this example captures the *generic* behaviour of f near a point where any two smooth functions ρ_A and ρ_B cross. This effect is the reason why the additional assumption (ii) was made in Theorem 5.5. This assumption, together with (5.16), yields the following alternative lower bound on P

$$P(x,w) \geq \frac{f(x)}{\rho_A(x)} \int_{\mathbb{R}^3} \beta \rho_A(x) \rho_A(z) \frac{\gamma_{A \to B}(z,w)}{\rho_A(z)}\, dz = \beta f(x) \rho_B(w). \qquad (5.22)$$

We fix a number $\delta \in (0,1)$ and we use the lower bounds (5.20) or (5.22), depending on whether $f_B(x) \geq \delta \rho_B(x)$ or $f_B(x) < \delta \rho_B(x)$.

Region 1: Assume $f_B(x) \geq \delta \rho_B(x)$. Via (5.20) and (5.21) we obtain

$$\begin{aligned}
\chi_{f_B \geq \delta \rho_B} \int_{\mathbb{R}^3} |\tilde{W}_4(\cdot, w)|^2 dw &\leq \frac{1}{4\delta} \frac{|\nabla f_B|^2}{\rho_B} \leq \frac{1}{2\delta} \chi_{\rho_B > \rho_A} \frac{|\nabla \rho_A|^2 + |\nabla \rho_B|^2}{\rho_B} \\
&\leq \frac{1}{2\delta} \chi_{\rho_B > \rho_A} \left(\frac{|\nabla \rho_A|^2}{\rho_A} + \frac{|\nabla \rho_B|^2}{\rho_B}\right) \leq \frac{2}{\delta} \left(|\nabla \sqrt{\rho_A}|^2 + |\nabla \sqrt{\rho_B}|^2\right). \quad (5.23)
\end{aligned}$$

Region 2: Assume $f_B(x) \leq \delta \rho_B(x)$. First of all, note that whenever $f_B(x) > 0$, i.e. $\rho_B(x) > \rho_A(x)$, we have the following equivalences

$$f_B(x) \leq \delta \rho_B(x) \Leftrightarrow \rho_B(x) - \rho_A(x) \leq \delta \rho_B(x) \Leftrightarrow \rho_B(x)(1 - \delta) \leq \rho_A(x) = \min\{\rho_A(x), \rho_B(x)\} = f(x).$$

Via (5.22) we have

$$|\tilde{W}_4| \leq \frac{1}{2\sqrt{\beta f(x) \rho_B(w)}} \frac{|\nabla f_B(x)|}{\int_{\mathbb{R}^3} f_B(y)\, dy}\, \tilde{g}(w).$$

We split the factor $\tilde{g}(w)$ into $\sqrt{\tilde{g}(w)}\sqrt{\tilde{g}(w)}$ and estimate one of the factors via the elementary inequality $f_A(y) \leq \rho_A(y)$, so as to eliminate the bad factor $\sqrt{\rho_B}$ from the denominator:

$$\begin{aligned}
\sqrt{\tilde{g}(w)} &= \left(\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{f_A(y)}{\rho_A(y)} \frac{\gamma^{opt}_{A \to A}(y,z)}{\rho_A(z)} \gamma_{A \to B}(z,w) dy dz\right)^{1/2} \leq \left(\int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \gamma^{opt}_{A \to A}(y,z) dy \frac{\gamma_{A \to B}(z,w)}{\rho_A(z)} dz\right)^{1/2} \\
&= \left(\int_{\mathbb{R}^3} \rho_A(z) \frac{\gamma_{A \to B}(z,w)}{\rho_A(z)} dz\right)^{1/2} = \sqrt{\rho_B(w)}.
\end{aligned}$$

34

Consequently,

$$|\tilde{W}_4| \leq \frac{1}{2\sqrt{\beta f(x)}} \frac{|\nabla f_B(x)|}{\int_{\mathbb{R}^3} f_B(y)dy} \sqrt{\tilde{g}(w)}.$$

Squaring, integrating over w and using $\int_{\mathbb{R}^3} \tilde{g}(w)dw = \int_{\mathbb{R}^3} f_B(y)dy$ yields

$$\int_{\mathbb{R}^3} |\tilde{W}_4(x,w)|^2 dw \leq \frac{1}{4\beta} \frac{|\nabla f_B(x)|^2}{f(x)} \frac{1}{\int_{\mathbb{R}^3} f_2(y)dy}. \tag{5.24}$$

But in the region $\{x \,|\, f_B(x) > 0\} = \{x \,|\, \rho_B(x) > \rho_A(x)\}$, as shown above we have $f = \rho_A \geq (1-\delta)\rho_B$, and consequently,

$$
\begin{aligned}
\chi_{f_B \leq \delta \rho_B} \int_{R^3} |\tilde{W}_4(\cdot,w)|^2 dw &\leq \frac{1}{2\beta} \frac{|\nabla \rho_A|^2 + |\nabla \rho_B|^2}{f} \frac{1}{\int_{\mathbb{R}^3} f_B(y)dy} \\
&\leq \frac{1}{2\beta} \left(\frac{|\nabla \rho_A|^2}{\rho_A} + \frac{1}{1-\delta} \frac{|\nabla \rho_B|^2}{\rho_B} \right) \frac{1}{\int_{\mathbb{R}^3} f_B(y)dy} \\
&= \frac{2}{\beta} \left(|\nabla \sqrt{\rho_A}|^2 + \frac{1}{1-\delta} |\nabla \sqrt{\rho_B}|^2 \right) \frac{1}{\int_{\mathbb{R}^3} f_B(y)dy}. \tag{5.25}
\end{aligned}
$$

Combining (5.17), (5.18), (5.19), (5.23) and (5.25) establishes the theorem. $\qquad\square$

Remark 5.7. *In region 1, the factor $\frac{1}{f(x)}$ appearing in (5.24) is uncontrollably bad. In region 2, the factor $\frac{1}{f_B(x)}$ appearing in (5.21) is uncontrollably bad.*

5.4 Smoothing

The third ingredient needed in the proof of Theorem 5.2 lies in the fact that the Coulomb cost functional $C[\gamma] = \int |x-y|^{-1} d\gamma(x,y)$ is well behaved under smoothing of γ, despite the fact that the cost function $|x-y|^{-1}$ is discontinuous and hence does not belong to the dual of the space of probability measures on $\mathbb{R}^6$.

Let $\rho \in \mathcal{R}$ (see (2.11)), and let $\gamma \in \Gamma(\rho,\rho)$ be a minimizer of $C[\gamma] = \int |x-y|^{-1} d\gamma(x,y)$ so that

$$C[\gamma] = E_{OT}[\rho].$$

We now introduce a standard mollification of γ, as follows. Let $\phi : \mathbb{R}^3 \to \mathbb{R}$ belong to the Schwartz space $\mathcal{S}(\mathbb{R}^3)$ of smooth, rapidly decaying functions, and assume that $\phi > 0$, $\int_{\mathbb{R}^3} \phi = 1$, ϕ radially symmetric. E.g., the choice $\phi(x) = \pi^{-3/2} e^{-|x|^2}$ will do. Let

$$\phi_\epsilon(x) = \frac{1}{\epsilon^3} \phi\left(\frac{x}{\epsilon}\right),$$

and let $\gamma_\epsilon = (\phi_\epsilon \otimes \phi_\epsilon) * \gamma$, that is to say

$$\gamma_\epsilon(x,y) = \int_{\mathbb{R}^6} \phi_\epsilon(x-x')\phi_\epsilon(y-y')d\gamma(x',y'). \tag{5.26}$$

Proposition 5.8. *The mollified pair density γ_ϵ introduced in (5.26) satisfies*

  (a) $\gamma_\epsilon \in C^\infty(\mathbb{R}^6)$, $\gamma_\epsilon > 0$

  (b) $\int \gamma_\epsilon(x,y)\, dx = \rho_\epsilon(y)$, $\int \gamma_\epsilon(x,y)\, dy = \rho_\epsilon(x)$, where ρ_ϵ is the mollified marginal $(\phi_\epsilon * \rho)(x) = \int_{\mathbb{R}^3} \phi_\epsilon(x - x')\rho(x')dx'$.

  (c) $C[\gamma_\epsilon] \leq C[\gamma]$.

Proof. (a): Smoothness is a standard fact concerning mollification of Radon measures, and positivity is obvious from the positivity of ϕ.

(b): This follows from the elementary calculation

$$
\begin{aligned}
\int \gamma_\epsilon(x,y)\,dx &= \int\int\int \phi_\epsilon(x-x')\phi_\epsilon(y-y')d\gamma(x',y')\,dx \\
&= \int\int \phi_\epsilon(y-y')\,d\gamma(x',y') = \int \phi_\epsilon(y-y')\rho(y')dy'.
\end{aligned}
$$

(c): First, we claim that the cost functional evaluated at the mollified transport plan, $C[\gamma_\epsilon]$, can be interpreted as a cost functional with modified cost function evaluated at the original transport plan. Indeed, by Fubini's theorem

$$
\begin{aligned}
C[\gamma_\epsilon] &= \int\int c(x,y)\Big[\int\int \phi_\epsilon(x-x')\phi_\epsilon(y-y')\,d\gamma(x',y')\Big]\,dx\,dy \\
&= \int\int \Big[\underbrace{\int\int \phi_\epsilon(x-x')\phi_\epsilon(y-y')c(x,y)\,dx\,dy}_{=:\tilde{c}(x',y')}\Big]\,d\gamma(x',y').
\end{aligned}
$$

The modified cost function $\tilde{c}(x',y')$ appearing here has an interesting physical meaning which we will exploit to establish (c), namely it is the electrostatic repulsion between the two charge distributions $\phi_\epsilon(\cdot - x')$ and $\phi_\epsilon(\cdot - y')$ (i.e., the charge distributions centered at x' respectively y' whose profile is given by the mollifier ϕ_ϵ). Now it is a standard fact going back to Newton that the electrostatic potential exerted by a radial charge distribution on a point outside it equals the potential exerted by the same amount of charge placed at the centre,

$$
\frac{1}{|S_r|}\int_{S_r}\frac{1}{|x-a|}dH^2(x) = \frac{1}{\max\{|a|,r\}},
$$

where S_r denotes the sphere of radius r around 0, H^2 is the Hausdorff measure (area element) on the sphere, and $|S_r|(= 4\pi r^2)$ is the total area of the sphere. This together with the radial symmetry of ϕ_ϵ (i.e., $\phi_\epsilon(x) = \tilde{\phi}_\epsilon(|x|)$ for some function $\tilde{\phi}_\epsilon$) gives

$$
\begin{aligned}
\int_{\mathbb{R}^3}\phi_\epsilon(x)\frac{1}{|x-a|}dx &= \int_{r=0}^\infty\int_{x\in S_r}\tilde{\phi}_\epsilon(r)\frac{1}{|x-a|}dH^2(x) = \int_0^\infty |S_r|\,\tilde{\phi}_\epsilon(r)\frac{1}{\max\{|a|,r\}}dr \\
&\le \Big(\int_0^\infty |S_r|\,\phi_\epsilon(r)\,dr\Big)\frac{1}{|a|} = \frac{1}{|a|}.
\end{aligned}
\tag{5.27}
$$

Hence by repeated application of (5.27)

$$
\begin{aligned}
\tilde{c}(x',y') &= \int\int \phi_\epsilon(x)\phi_\epsilon(y)\frac{1}{|x+x'-(y+y')|}dx\,dy \tag{5.28} \\
&\le \int \phi_\epsilon(y)\frac{1}{|x'-(y+y')|}dy = \int \phi_\epsilon(y)\frac{1}{|y-(x'-y')|}dy \tag{5.29} \\
&\le \frac{1}{|x'-y'|}. \tag{5.30}
\end{aligned}
$$

This establishes (c). $\qquad\square$

5.5 Passage to the limit

We are now in a position to give the

Proof of Theorem 5.2. Let ρ be any density in $\mathcal{R}$. Recall that $\rho \in \mathcal{R}$ implies that $\sqrt{\rho}$, $\nabla\sqrt{\rho} \in L^2(\mathbb{R}^3)$ and hence, by the Sobolev embedding theorem, $\sqrt{\rho} \in L^6(\mathbb{R}^3)$, whence $\rho \in L^1 \cap L^3(\mathbb{R}^3)$.

We have to show that $\lim_{\hbar\to 0} F_{HK}[\rho] = E_{OT}[\rho]$. We will do so via the following strategy:

36

- Start from an optimal transport plan γ with marginals ρ.

- Smooth it.

- Make it strongly positive (see Definition 5.4), by mixing in a small amount of the mean field (i.e., tensor product) plan.

- Re-instate the marginal constraint, via the technique introduced in Section 5.1.

- Infer from Theorem 5.5 that, unlike the original optimal transport plan γ, the so-obtained modified plan P is the pair density (2.6) of a wave function Ψ with square-integrable gradient.

- Pass to the semiclassical limit, by careful error estimates on the three modification steps listed above (smoothing, achieving strong positivity, re-instating the constraint).

We now implement this strategy in detail. Let γ be an optimal transport plan of the Coulomb cost functional C subject to equal marginals ρ. (Of course we know from Section 3 that γ is unique, but uniqueness is not needed here.) For $\epsilon > 0$, let γ_ϵ be its mollification (5.26). By Proposition 5.8, its right and left marginals are given by the mollification $\rho_\epsilon = \phi_\epsilon * \rho$ of the density ρ. Finally we introduce the "strong positivization"

$$\tilde{\gamma}_{\epsilon,\beta} := (1 - \beta)\gamma_\epsilon + \beta \rho_\epsilon \otimes \rho_\epsilon,$$

where $\beta \in (0, 1)$. Note that $\tilde{\gamma}_{\epsilon,\beta}$ has the same marginals as γ_ϵ, regardless of the value of β.

Observe now that the transportation plan $\tilde{\gamma}_{\epsilon,\beta}$ and the densities ρ_ϵ, ρ satisfy the assumptions of Theorem 5.5. Consequently, by Theorem 5.5 there exists a transportation plan $P_{\epsilon,\beta}$ with marginals ρ (i.e., with re-instated constraint) whose square root belongs to $H^1(\mathbb{R}^6)$.

Now comes the only step where we use the assumption $N = 2$. In this case we can achieve the (otherwise highly nontrivial) antisymmetry condition on Ψ appearing in (2.2) purely by means of an antisymmetric spin part. More precisely we define $\Psi : (\mathbb{R}^3 \times \mathbb{Z}_2)^2 \to \mathbb{C}$ by

$$\Psi(x, s, y, t) := \sqrt{P_{\epsilon,\beta}(x, y)} \frac{\alpha(s)\beta(t) - \beta(s)\alpha(t)}{\sqrt{2}},$$

where $\alpha, \beta : \mathbb{Z}_2 = \{\pm\frac{1}{2}\} \to \mathbb{C}$ are given by $\alpha(s) = \delta_{1/2}(s)$, $\beta(s) = \delta_{-1/2}(s)$. Then it is straightforward to check that Ψ belongs to the admissible set $\mathcal{A}$ defined in (2.2) and its pair density, density, and kinetic energy are

$$\rho_B^\Psi = P_{\epsilon,\beta}, \quad \rho^\Psi = \rho, \quad T_\hbar[\Psi] = \frac{\hbar^2}{2m} \int_{\mathbb{R}^6} |\nabla \sqrt{P_{\epsilon,\beta}}|^2.$$

It follows that

$$\lim_{\hbar \to 0} F_{HK}[\rho] \leq \lim_{\hbar \to 0} \left(T_\hbar(\Psi) + V_{ee}(\Psi) \right) = V_{ee}(\Psi) = C(P_{\epsilon,\beta}). \tag{5.31}$$

Next, (5.13) yields

$$C[P_{\epsilon,\beta}] \leq C[\tilde{\gamma}_{\epsilon,\beta}] + c_* \left(||\rho||_{L^1 \cap L^3} + ||\rho_\epsilon||_{L^1 \cap L^3} \right) ||\rho - \rho_\epsilon||_{L^1 \cap L^3}. \tag{5.32}$$

Next, we claim that

$$C[\tilde{\gamma}_{\epsilon,\beta}] = (1 - \beta)C[\gamma_\epsilon] + \beta C(\rho_\epsilon \otimes \rho_\epsilon) \leq C[\gamma_\epsilon] + c_0 \beta ||\rho_\epsilon||_{L^1} ||\rho_\epsilon||_{L^1 \cap L^3}. \tag{5.33}$$

This is immediate from the estimate

$$|C[f \otimes g]| \leq c_0 ||f||_{L^1} ||g||_{L^1 \cap L^3} \text{ for any } f, g \in L^1 \cap L^3,$$

which follows by applying (5.15), multiplying by f, and integrating over x.

Finally, we will need the following bound which was established in Proposition 5.8:

$$C[\gamma_\epsilon] \leq C[\gamma]. \tag{5.34}$$

37

Combining the estimates (5.31)–(5.34) yields

$$\lim_{\hbar \to 0} F_{HK}[\rho] \leq C[\gamma] + \frac{c_*}{3} \Big[\beta ||\rho_\epsilon||_{L^1} ||\rho_\epsilon||_{L^1 \cap L^3} + (||\rho||_{L^1 \cap L^3} + ||\rho_\epsilon||_{L^1 \cap L^3}) ||\rho - \rho_\epsilon||_{L^1 \cap L^3} \Big].$$

Letting β and ϵ tend to zero and using that ρ_ϵ, being the mollification $\phi_\epsilon * \rho$ of ρ, tends to ρ in $L^1 \cap L^3$ as $\epsilon \to 0$ yields

$$\lim_{\hbar \to 0} F_{HK}[\rho] \leq C[\gamma] = E_{OT}[\rho].$$

The reverse inequality is immediate from (5.4) and the positivity of $T_\hbar$. This completes the proof of Theorem 5.2.

A   Appendix

Lemma A.1 (Legendre transforms on the line). *Let $l : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous and convex. Define its dual function l° by (3.10). Then l° satisfies the same hypotheses as l, and*

(a) *$(\lambda, \xi) \in \partial.l$ if and only if $(\xi, \lambda) \in \partial.l^\circ$;*

(b) *the dual function of l° is l, that is $l = l^{\circ\circ}$;*

(c) *strict convexity of l implies l° differentiable, where it is subdifferentiable;*

(iv) *$l^\circ(\xi)$ is non-increasing if and only if $l(\lambda) = \infty$ for all $\lambda > 0$.*

Proof. (a)-(c) follow from the corresponding statements in Theorem A.1 in [GM96]. Assertion (d) is easily proved similarly to Theorem A.3 (iv) in [GM96]. To verify the *only if* implication, suppose that $l(\lambda)$ is finite at some $\lambda > 0$; we shall show that l° increases somewhere. Being convex, l must be subdifferentiable at λ (or some nearby point): $(\lambda, \xi) \in \partial.l$. Then (i) implies that l° is finite at ξ and increasing: $l^\circ(\xi + \epsilon) \geq l^\circ(\xi) + \lambda\epsilon$ for some $\varepsilon > 0$.

To prove the converse, suppose that l° increases somewhere. Then one has $(\xi, \lambda) \in \partial.l^\circ$ for some $\xi \in \mathbb{R}$ and $\lambda > 0$. Invoking once again (i) gives $(\lambda, \xi) \in \partial.l$, from which one concludes finiteness of $l(\lambda)$. □

For $x \in \mathbb{R}^d \setminus \{0\}$, denote by $\hat{x} := x/|x|$ the unit vector in direction of x.

Lemma A.2 (subdifferentiability of the cost). *Let $l : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be convex and non-increasing on $[0, \infty)$ and define $h(x) := l(|x|)$ on $\mathbb{R}^d$. Unless h is a constant: $(x, y) \in \partial.h$ if and only if $(|x|, -|y|) \in \partial.l$ with $y = |y|\hat{x}$ and $x \neq 0$.*

Proof. Fix $x \in \mathbb{R}^d \setminus \{0\}$ and suppose $l(\lambda)$ admits ξ as a subderivative at $|x|$: $(|x|, \xi) \in \partial.l$. Since l is convex and non-increasing, $\xi \leq 0$, while for $\epsilon \in \mathbb{R}$,

$$l(|x| + \epsilon) \geq l(|x|) + \epsilon\xi. \tag{A.1}$$

Let

$$\epsilon := |x + v| - |x| = \sqrt{|x|^2 + 2 <x, v> + |v|^2} - |x| \leq <\hat{x}, v> + \frac{v^2}{2|x|}, \tag{A.2}$$

which inequality follows from $\sqrt{1 + \lambda} \leq 1 + \frac{\lambda}{2}$. Now $h(x+v) = l(|x+v|) \geq l(|x|+\epsilon)$, with $\epsilon = <v, \hat{x}> + o(|v|)$, as seen from (A.2). It follows immediately from Definition 3.11(1) that h is subdifferentiable at x, with $(x, \xi\hat{x}) \in \partial.h$. On the other hand, h cannot be subdifferentiable at the origin as $h(0) = \infty$.

Now let $(x, y) \in \partial.h$, so $x \neq 0$ and for small $v \in \mathbb{R}^d$

$$h(x + v) \geq h(x) + <v, y> + o(|v|).$$

Spherical symmetry of h forces y to be parallel to x: otherwise, a slight rotation $x + v := x \cos\theta - \hat{z} \sin\theta$ of x in the direction $z := y - (< y, \hat{x} >)\hat{x}$ would contradict $h(x + v) = h(x)$ for θ sufficiently small. Moreover, taking $v := \epsilon\hat{x}$ yields (A.1) with $\xi := < \hat{x}, y > +o(1)$, which concludes the lemma: $|y| = \pm < \hat{x}, y >$ holds with a minus sign since l cannot increase. $\qquad\square$

Lemma A.3 (uniform subdifferentiability of the cost)**.** *Let l and h be defined as in the lemma above. Then h is subdifferentiable on $\mathbb{R}^d \setminus \{0\}$. Moreover, for $\delta > 0$, there is a real function $O_\delta(\lambda)$ tending to zero linearly with $|\lambda|$ such that $|x| > \delta, y \in \partial.h(x)$ and $v \in \mathbb{R}^d$ imply*

$$h(x + v) \geq h(x)+ < v, y > +O_\delta(v^2). \tag{A.3}$$

Proof. For $\lambda > 0$, the convex function l admits a subgradient $\xi \in \partial.l(\lambda)$: for example, take its right derivative $\xi = l'(\lambda^+)$. If $|x| = \lambda$, the lemma implies $(x, \xi\hat{x}) \in \partial.h$, so $h(x)$ is subdifferentiable at x.

Now suppose that $(x, y) \in \partial.h$. The opposite implication of the lemma yields $y = \xi\hat{x}$ with $(|x|, \xi) \in \partial.l$ so (3.24) holds. Morover, $\xi \leq 0$. If $v \in \mathbb{R}^d$, then $h(x + v) \geq l(|x| + \epsilon)$ where ϵ is as in (A.2). By convexity of l, its right derivative is a non-decreasing function of λ. Asssume $|x| > \delta$ so that $\xi \geq l'(\delta^+)$. Together with (A.1) and (A.2), this assumption gives

$$h(x + v) \geq h(x)+ < \xi\hat{x}, v > +v^2 l'(\delta^+)/2\delta.$$

$\qquad\square$

Lemma A.4. *Let l and h be defined as in the Lemma A.2. Define the dual function $h^* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ via (3.10). The for some $R \geq 0$,*

(i) $h^*(y)$ is continuously differentiable on $|y| > R$ while $h^* = +\infty$ on $|y| < R$;

(ii) $(y, x) \in \partial.h^*$ with $x \neq 0$ if and only if $(x, y) \in \partial.h$ with $y \neq 0$;

(iii) if $(y, x) \in \partial.h^*$, then $x = \nabla h^*(y)$.

Proof. The proof follows the same reasoning as the proof of Proposition A.6 (i)-(iii) from [GM96] and will be omitted. $\qquad\square$

References

[Be93] A. BECKE. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648 (1993)

[CFK11] C. COTAR, G. FRIESECKE, C. KLÜPPELBERG. In preparation.

[CY02] A. J. COLEMAN, V. I YUKALOV. Reduced density matrices. *Springer Lecture Notes in Chemistry* Vol. 72 (2002)

[Di30] P. A. M. DIRAC. Note on exchange phenomena in the Thomas atom. *Proc. Cambridge Philosophical Society* **26**, 376-385 (1930)

[FNM03] C. FIOLHAIS, F. NOQUEIRA, M. MARQUES (EDS). A Primer in Density Functional Theory. *Springer Lecture Notes in Physics* Vol. 620 (2003)

[Fr97] G. FRIESECKE. Pair correlations and exchange phenomena in the free electron gas, *Comm. Math. Phys.* **184**, 143-171 (1997)

[Fr03] G. FRIESECKE. The multiconfiguration equations for atoms and molecules: charge quantization and existence of solutions. *Arch. Rat. Mech. Analysis* **169**, 35-71 (2003)

[GO07] W. GANGBO, V. OLIKER. Existence of optimal maps in the reflector-type problems, preprint.

[GM95] W. GANGBO, R. MCCANN. Optimal maps in Monge's mass transport problem, *C.R. Acad. Sci. Paris. Ser. I. Math.* **325**, 1653-1658 (1995)

[GM96] W. GANGBO, R. MCCANN. The geometry of optimal transportation, *Acta Math.* **177**, 113-161 (1996)

[GD95] E. K. U. GROSS & R. M. DREIZLER (EDS.) *Density Functional Theory.* Plenum Press (1995)

[HK64] P. HOHENBERG, W. KOHN. Inhomogeneous electron gas, *Phys. Rev. B* **136**, 864-871 (1964)

[KM07] Y.-H. KIM, R. MCCANN. Continuity, curvature, and the general covariance of optimal transportation, *preprint*

[Ko95] W. KOHN. Overview of density functional theory. In [GD95], 3-10, 1995

[KS65] W. KOHN, L. J. SHAM. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A* **140**, 1133-1138 (1965)

[Le79] M. LEVY. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Natl. Acad. Sci. USA* **76**(12), 6062-6065 (1979)

[Li83] E.H. LIEB. Density functionals for Coulomb systems, *International Journal of Quantum Chemistry* **24**, 243-277 (1983)

[LYP88] C. LEE, W. YANG, R. G. PARR. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Phys. Rev. B* **37**, 785-789 (1988)

[PY95] R. G. PARR, W. YANG. *Density-Functional Theory of Atoms and Molecules.* Oxford University Press, Oxford (1995)

[Ra09] D. RAPPOPORT, N. R. M. CRAWFORD, F. FURCHE, K. BURKE. Which density functional should I choose? In: *Computational Inorganic and Bioinorganic Chemistry*, eds. E. I. Solomon, R. B. King, and R. A. Scott. Wiley (2009)

[Ro72] R. T. ROCKAFELLAR. *Convex Analysis.* Princeton University Press, Princeton, (1970)

[Ru96] L. RÜSCHENDORF On c-optimal random variables. *Stat. Prob. Letter* **27**, 267-270 (1996).

[SK92] C. SMITH, M. KNOTT. On Hoeffding-Fréchet bounds and cyclic monotone relations, *J. Mult. Anal.* **40**, 328-334 (1992).

[Vill09] C. VILLANI *Optimal Transport: Old and New.* Springer, Heidelberg (2009).