HAL
open science

A Numerical Method to solve Optimal Transport Problems with Coulomb Cost

Jean-David Benamou, Guillaume Carlier, Luca Nenna

To cite this version:

Jean-David Benamou, Guillaume Carlier, Luca Nenna. A Numerical Method to solve Optimal Transport Problems with Coulomb Cost. Splitting Methods in Communication, Imaging, Science, and Engineering, 2016. hal-01148954v2

HAL Id: hal-01148954

https://inria.hal.science/hal-01148954v2

Submitted on 8 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Numerical Method to solve Optimal Transport Problems with Coulomb Cost

Jean-David Benamou, Guillaume Carlier, and Luca Nenna

Abstract In this paper, we present a numerical method, based on iterative Bregman projections, to solve the optimal transport problem with Coulomb cost. This is related to the strong interaction limit of Density Functional Theory. The first idea is to introduce an entropic regularization of the Kantorovich formulation of the Optimal Transport problem. The regularized problem then corresponds to the projection of a vector on the intersection of the constraints with respect to the Kullback-Leibler distance. Iterative Bregman projections on each marginal constraint are explicit which enables us to approximate the optimal transport plan. We validate the numerical method against analytical test cases.

1 Introduction

1.1 On Density functional theory

Quantum mechanics for a molecule with N electrons boils down to the many-electron Schrödinger equation for a wave function $\psi \in L^2(\mathbb{R}^{3N};\mathbb{C})$ (in this paper, we neglect the spin variable). The limit of this approach is computational : in order to predict the chemical behaviour of H_2O (10 electrons) using a 10 gridpoints discretization of $\mathbb{R}$, we need to solve the Schrödinger equation on 10^{30} gridpoints. This is why Hohenberg, Kohn and Sham introduced, in [19] and [21], the Density Functional Theory (DFT) as an approximate computational method for solving the Schrödinger equation at a more reasonable cost.

Jean-David Benamou, Luca Nenna
INRIA, MOKAPLAN, Domaine de Voluceau Le Chesnay, FRANCE, e-mail: jean-david.benamou/luca.nenna@inria.fr

Guillaume Carlier
CEREMADE, Université Paris Dauphine e-mail: carlier@ceremade.dauphine.fr

The main idea of the DFT is to compute only the marginal density for one electron

$$\rho(x_1) = \int \gamma_N(x_1, x_2 \cdots, x_N) dx_2 \cdots dx_N,$$

where $\gamma_N = |\psi(x_1, \cdots, x_N)|^2$ is the joint probability density of electrons at positions $x_1, \cdots, x_N \in \mathbb{R}^3$, instead of the full wave function ψ. One scenario of interest for the DFT is when the repulsion between the electrons largely dominates over the kinetic energy. In this case, the problem can, at least formally, be reformulated as an Optimal Transport (OT) problem as emphasized in the pioneering works of Buttazzo, De Pascale and Gori-Giorgi [6] and Cotar, Friesecke and Klüppelberg [10].

1.2 Optimal Transport

Before discussing the link between DFT and OT, let us recall the standard optimal transport problem and its extension to the multi-marginal framework. Given two probability distributions μ and ν (on $\mathbb{R}^d$, say) and a transport cost c: $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, the optimal transport problem consists in finding the cheapest way to transport μ to ν for the cost c. A transport map between μ and ν is a Borel map T such that $T_\# \mu = \nu$ i.e. $\nu(A) = \mu(T^{-1}(A))$ for every Borel subset A of $\mathbb{R}^d$. The Monge problem (which dates back to 1781 when Monge [24] posed the problem of finding the optimal way to move a pile of dirt to a hole of the same volume) then reads

$$\min_{T_\# \mu = \nu} \int_{\mathbb{R}^d} c(x, T(x)) \mu(dx). \tag{1}$$

This is a delicate problem since the mass conservation constraint $T_\# \mu = \nu$ is highly nonlinear (and the feasible set may even be empty for instance if μ is a Dirac mass and ν is not). This is why, in 1942, Kantorovich [20] proposed a relaxed formulation of (1) which allows mass splitting

$$\min_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y) \gamma(dx, dy) \tag{2}$$

where $\gamma \in \Pi(\mu, \nu)$ consists of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ having μ and ν as marginals, that is:

$$\gamma(A \times \mathbb{R}) = \mu(A), \quad \forall A \text{ Borel subset of } \mathbb{R}^d, \tag{3}$$

$$\gamma(\mathbb{R} \times B) = \nu(B), \quad \forall B \text{ Borel subset of } \mathbb{R}^d. \tag{4}$$

Note that this is a linear programming problem and that there exists solutions under very mild assumptions (e.g. c continuous and μ and ν compactly supported). A minimizing γ in (2) is called an optimal transport plan and it gives the probability that a mass element in x be transported in y. Let us remark that if T is a transport map then it induces a transport plan $\gamma_T(x,y) := \mu(x)\delta(y - T(x))$ so if an optimal

plan of (2) has the form γ_T (which means that no splitting of mass occurs and γ is concentrated on the graph of T) then T is actually an optimal transport map i.e. a solution to (1). The linear problem (2) also has a convenient dual formulation

$$\max_{u,v|u(x)+v(y)\leq c(x,y)} \int_{\mathbb{R}^d} u(x)\mu(dx) + \int_{\mathbb{R}^d} v(y)\nu(dy) \tag{5}$$

where $u(x)$ and $v(y)$ are the so-called Kantorovich potentials. OT theory for two marginals has developed very rapidly in the 25 last years, there are well known conditions on c, μ and ν which guarantee that there is a unique optimal plan which is in fact induced by a map (e.g. $c = |x-y|^2$ and μ absolutely continuous, see Brenier [4]) and we refer to the textbooks of Villani [34, 35] for a detailed exposition.

Let us now consider the so-called multi-marginal problems i.e. OT problems involving N marginals $\mu_1, \cdots, \mu_N$ and a cost $c : \mathbb{R}^{dN} \to \mathbb{R}$, which leads to the following generalization of (2)

$$\min_{\gamma\in\Pi(\mu_1,\cdots,\mu_N)} \int_{\mathbb{R}^{\times\mathbb{N}}} c(x_1,\cdots,x_N)\gamma(dx_1,\cdots,dx_N) \tag{6}$$

where $\Pi(\mu_1,\cdots,\mu_N)$ is the set of probability measures on $(\mathbb{R}^d)^N$ having $\mu_1,\cdots,\mu_N$ as marginals. The corresponding Monge problem then becomes

$$\min_{T_{i\#}\mu_1=\mu_i,\, i=2,\cdots,N} \int_{\mathbb{R}^d} c(x_1, T_2(x_1),\cdots,T_N(x_1))\mu_1(dx_1). \tag{7}$$

Such multi-marginals problems first appeared in the work of Gangbo and Święch [16] who solved the quadratic cost case and proved the existence of Monge solutions. In recent years, there has been a lot of interest in such multi-marginal problems because they arise naturally in many different settings such as economics [7], [29], polar factorization of vector fields and theory of monotone maps [17] and, of course, DFT [6, 10, 8, 14, 23, 11], as is recalled below. Few results are known about the structure of optimal plans for (7) apart from the general results of Brendan Pass [28], in particular the case of *repulsive costs* such as the Coulomb's cost from DFT is an open problem.

The paper is structured as follows. In Section 2, we recall the link between Density Functional Theory and Optimal Transportation and we present some analytical solutions of the OT problem (e.g. optimal maps for radially symmetric marginals, for 2 electrons). In Section 3, we introduce a numerical method, based on iterative Bregman projections, and an algorithm which aims at refining the mesh where the transport plan is concentrated. In section 4 we present some numerical results. Section 5 concludes.

2 From Density Functional Theory to Optimal Transportation

2.1 Optimal Transportation with Coulomb cost

In Density Functional Theory [19] the ground state energy of a system (with N electrons) is obtained by minimizing the following functional w.r.t. the electron density $\rho(x)$:

$$E[\rho] = \min_{\rho \in \mathscr{R}} F_{HK}[\rho] + \int v_{ext}(x)\rho(x)dr \tag{8}$$

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

$v_{ext} := -\dfrac{Z}{|x-R|}$ is the electron-nuclei potential (Z and R are the charge and the position of the nucleus, respectively) and F_{HK} is the so-called Hohenberg-Kohn which is defined by minimizing over all wave functions ψ which yield ρ:

$$F_{HK}[\rho] = \min_{\psi \to \rho} \hbar^2 T[\psi] + V_{ee}[\psi] \tag{9}$$

where $\hbar^2$ is a semiclassical constant factor,

$$T[\psi] = \frac{1}{2} \int \cdots \int \sum_{i=1}^N |\nabla_{x_i} \psi|^2 dx_1 \cdots dx_N$$

is the kinetic energy and

$$V_{ee} = \int \cdots \int \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|x_i - x_j|} |\psi|^2 dx_1 \cdots dx_N$$

is the Coulomb repulsive energy operator.

Let us now consider the *Semiclassical* limit

$$\lim_{\hbar \to 0} \min_{\psi \to \rho} \hbar^2 T[\psi] + V_{ee}[\psi]$$

and assume that taking the minimum over ψ commutes with passing to the limit $\hbar \to 0$ (Cotar, Friesecke and Klüppelberg in [10] proved it for $N = 2$), we obtain the following functional

$$V_{ee}^{SCE}[\rho] = \min_{\psi \to \rho} \int \cdots \int \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|x_i - x_j|} |\psi|^2 dx_1 \cdots dx_N \tag{10}$$

where V_{ee}^{SCE} is the minimal Coulomb repulsive energy whose minimizer characterizes the state of *Strictly Correlated Electrons*(SCE).

Problem (10) gives rise to a multi-marginal optimal transport problem as (6) by considering that

- according to the indistinguishability of electrons, all the marginals are equal to ρ,

- the cost function is given the electron-electron Coulomb repulsion,

$$c(x_1,...,x_N) = \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{|x_i - x_j|}, \tag{11}$$

- we refer to $\gamma_N = |\psi(x_1,\cdots,x_N)|^2$ (which is the joint probability density of electrons at positions $x_1,\cdots,x_N \in \mathbb{R}^3$) as the transport plan.

The Coulomb cost function (11) is different from the costs usually considered in OT as it is not bounded at the origin and it decreases with distance. So it requires a generalized formal framework, but this is beyond the purpose of this work (see [6] and [10]). Finally (10) can be re-formulated as a Kantorovich problem

$$V_{ee}^{SCE}[\rho] = \min_{\pi_i(\gamma_N)=\rho, i=1,\cdots,N} \int_{\mathbb{R}^{3N}} c(x_1,\cdots,x_N)\gamma_N(x_1,\cdots,x_N)dx_1\cdots dx_N \tag{12}$$

where

$$\pi_i(\gamma_N) = \int_{\mathbb{R}^{3(N-1)}} \gamma_N(x_1,\cdots,x_i,\cdots,x_N)dx_1,\cdots,dx_{i-1},dx_{i+1},\cdots,dx_N$$

is the $i-$th marginal. As mentioned in section 1.2 if the optimal transport plan γ_N has the following form

$$\gamma_N(x_1,\cdots,x_N) = \rho(x_1)\delta(x_2 - f_2^\star(x_1))\cdots\delta(x_N - f_N^\star(x_1)) \tag{13}$$

then the functions $f_i^\star : \mathbb{R}^3 \to \mathbb{R}^3$ are the optimal transport maps (or *co-motion* functions) of the Monge problem

$$\begin{aligned} V_{ee}^{SCE}[\rho] &= \min_{\{f_i:\mathbb{R}^3\to\mathbb{R}^3\}_{i=1}^N} \int \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{|f_i(x) - f_j(x)|}\rho(x)dx \\ s.t. \quad & f_{i\#}\rho = \rho,\ i = 2,...,N, \quad f_1(x) = x. \end{aligned} \tag{14}$$

Remark 1. (Physical meaning of the co-motion function) $f_i(x)$ determine the position of the i-th electron in terms of x which is the position of the "1st"electron : V_{ee}^{SCE} defines a system with the maximum possible correlation between the relative electronic positions.

In full generality, problem (14) is delicate and proving the existence of the co-motion functions is difficult. However, the co-motion functions can be obtained via semianalytic formulations for spherically symmetric atoms and strictly 1D systems (see [10], [33], [22], [8]) and we will give some examples in the following section.

Problem (12) admits a useful dual formulation in which the so called Kantorovich potential u plays a central role

$$V_{ee}^{SCE} = \max_u \{N \int u(x)\rho(x)dx \quad s.t. \quad \sum_{i=1}^{N} u(x_i) \le c(x_1,...,x_N)\}. \tag{15}$$

Because c is invariant by permutation, there is a single dual Kantorovich potential for all all marginal constraints. Moreover, this potential $u(x)$ is related to the co-motion functions via the classical equilibrium equation (see [33])

$$\nabla u(x) = -\sum_{i=2}^{N} \frac{x - f_i(x)}{|x - f_i(x)|^3}. \tag{16}$$

Remark 2. (Physical meaning of (16)) The gradient of the Kantorovich potential equals the total net force exerted on the electron in x by electrons in $f_2(x), \cdots, f_N(x)$.

2.2 Analytical Examples

2.2.1 The case $N = 2$ and $d = 1$

In order to better understand the problem we have formulated in the previous section, we recall some analytical examples (see [6] for the details).

Let us consider 2 particles in one dimension and marginals

$$\rho_1(x) = \rho_2(x) = \begin{cases} a & if |x| \leq a/2 \\ 0 & otherwise. \end{cases} \tag{17}$$

After a few computations, we obtain the following associated co-motion function

$$f(x) = \begin{cases} x + \frac{a}{2} \\ x - \frac{a}{2} \end{cases} . \tag{18}$$

If we take

$$\rho_1(x) = \rho_2(x) = \frac{a - |x|}{a^2} \quad defined \quad in \quad [-a, a], \tag{19}$$

we get

$$f(x) = \frac{x}{|x|}(\sqrt{2a|x| - x^2} - a) \quad on \quad [-a, a] \tag{20}$$

Figure 1 shows the co-motion functions for (17) and (19).

2.2.2 The case $N > 2$ and $d = 1$

In [8], the authors proved the existence of optimal transport maps for problem (14) in dimension $d = 1$ and provided an explicit construction of the optimal maps. Let ρ be the normalized electron density and $-\infty = x_0 < x_1 < \cdots < x_N = +\infty$ be such that

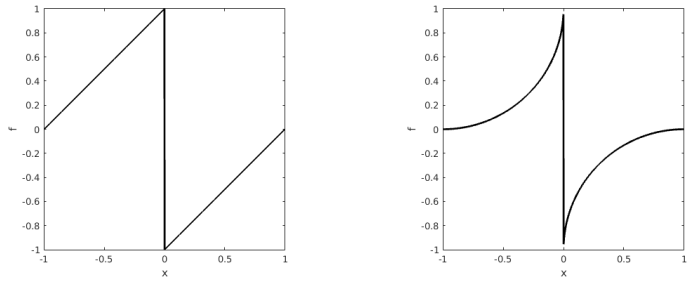

Fig. 1 *Right: Co-motion function for (17) with $a = 2$. Left: Co-motion function for (19) with $a = 1$.*

$$\int_{x_i}^{x_{i+1}} \rho(x)dx = 1/N \; \forall i = 0, \cdots, N-1.$$

Thus, there exists a unique increasing function $\tilde{f} : \mathbb{R} \to \mathbb{R}$ on each interval $[x_i, x_{i+1}]$ such that for every test-function φ one has

$$\int_{[x_i, x_{i+1}]} \varphi(\tilde{f}(x))\rho(x)dx = \int_{[x_{i+1}, x_{i+2}]} \varphi(x)\rho(x)dx \quad \forall i = 0, \cdots, N-2, \tag{21}$$

$$\int_{[x_{N-1}, x_N]} \varphi(\tilde{f}(x))\rho(x)dx = \int_{[x_0, x_1]} \varphi(x)\rho(x)dx, \tag{22}$$

The optimal maps are then given by

$$f_2(x) = \tilde{f}(x) \tag{23}$$

$$f_i(x) = f_2^{(i)}(x) \quad \forall i = 2, \cdots, N, \tag{24}$$

where $f_2^{(i)}$ stands for the $i-$th composition of f_2 with itself. Here, we present an example given in [6]. We consider the case where ρ is the Lebesgue measure on the unit interval $I = [0,1]$, the construction above gives the following optimal co-motion functions

$$\begin{aligned} f_2(x) &= \begin{cases} x + 1/3 & if \quad x \leq 2/3 \\ x - 2/3 & if \quad x > 2/3 \end{cases}, \\ f_3(x) = f_2(f_2(x)) &= \begin{cases} x + 2/3 & if \quad x \leq 1/3 \\ x - 1/3 & if \quad x > 1/3 \end{cases}. \end{aligned} \tag{25}$$

Furthermore, we know that the Kantorovich potential u satisfies the relation (here we take $N = 3$)

$$u'(x) = -\sum_{i=2}^{N} \frac{x - f_i(x)}{|x - f_i(x)|^3} \tag{26}$$

and by substituting the co-motion functions in (26) (and integrating it) we get

$$u(x) = \begin{cases} \frac{45}{4}x & 0 \le x \le 1/3 \\ \frac{15}{4} & 1/3 \le x \le 2/3 \\ -\frac{45}{4}x + \frac{45}{4} & 2/3 \le x \le 1 \end{cases} \tag{27}$$

Figure 2 illustrates this example.

When $N \geq 4$ similar arguments as above can be developed and we can similarly compute the co-motion functions and the Kantorovich potential.

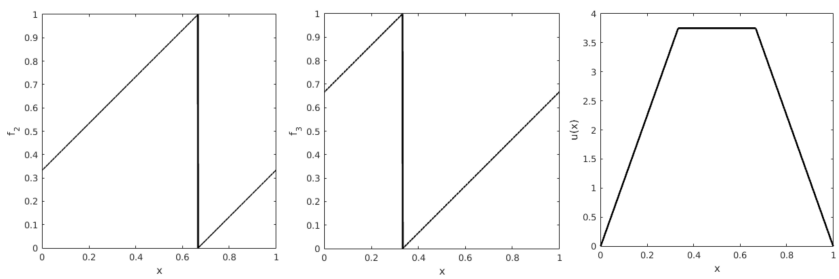

Fig. 2 *Right: co-motion function f_2 for (25). Center: co-motion function f_3 for (25). Left: Kantorovich Potential $u(x)$ (27).*

2.2.3 The radially symmetric marginal case for $N = 2$, $d \geq 2$

We discuss now the radial $d-$dimensional ($d \geq 2$) case for $N = 2$. We assume that the marginal ρ is radially symmetric, then we recall the following theorem from [10]:

Theorem 1. *[10] Suppose that $\rho(x) = \rho(|x|)$, then the optimal transport map is given by*

$$f^\star(x) = \frac{x}{|x|} g(|x|), \quad x \in \mathbb{R}^d, \tag{28}$$

with $g(r) = -F_2^{-1}(F_1(r))$, $F_1(t) := C(d)\int_0^t \rho(s)s^{d-1}ds$, $F_2(t) := C(d)\int_t^\infty \rho(s)s^{d-1}ds$ where $C(d)$ denotes the measure of S^{d-1}, the unit sphere in $\mathbb{R}^d$.

Example 1. (Spherical coordinates system) If ρ is radially symmetric $\rho(x) = \rho(|x|)$, it is convenient to work in spherical coordinates and then to set for every radius $r > 0$

$$\lambda(r) = C(d)r^{d-1}\rho(r) \tag{29}$$

so that for every test-function φ we have

$$\int_{\mathbb{R}^d} \varphi(x)\rho(|x|)dx = \int_0^{+\infty} \Big(\int_{S^{d-1}} \varphi(r,\omega) \frac{d\sigma(\omega)}{C_d} \Big) \lambda(r)dr$$

with $C(d)$ the measure of S^{d-1} and σ the $d-1$ measure on S^{d-1} which in particular implies that $\lambda := |.|_{\#}\rho$ i.e.

$$\int_{\mathbb{R}^d} \varphi(|x|)\rho(|x|)dx = \int_0^{+\infty} \varphi(r)\lambda(r)dr, \ \forall \varphi \in C_c(\mathbb{R}_+). \tag{30}$$

The radial part of the optimal co-motion function $a(r) = -g(r)$ can be computed by solving the ordinary differential equation

$$a'(r)\lambda(a(r)) = \lambda(r)$$

which gives

$$\int_0^{a(r)} \lambda(s)ds = 2 - \int_0^r \lambda(s)ds. \tag{31}$$

We define $R(r) = \int_0^r \lambda(s)ds$, since $r \mapsto R(r)$ is increasing, its inverse $R^{-1}(w)$ is well defined for $w \in [0,1)$. Thus, we see that $a(r)$ has the form

$$a(r) = R^{-1}(2 - R(r)). \tag{32}$$

2.2.4 Reducing the dimension under radial symmetry

In the case where the marginal $\rho(x) = \rho(|x|)$ is radially symmetric, the multi-marginal problem with Coulomb cost

$$\inf_{\gamma \in \Pi(\rho,\cdots,\rho)} \int_{\mathbb{R}^{dN}} c(x_1,\cdots,x_N)d\gamma(x_1,\cdots,x_N) \tag{33}$$

with c the Coulomb cost given by (11) involves plans on $\mathbb{R}^{dN}$ which is very costly to discretize. Fortunately, thanks to the symmetries of the problem, it can actually be solved by considering a multi-marginal problem only on $\mathbb{R}_+^N$. Let us indeed define for every $(r_1,\cdots,r_N) \in (0,+\infty)^N$:

$$\widetilde{c}(r_1,\cdots,r_N) := \inf\{c(x_1,\cdots,x_N) \ : \ |x_1| = r_1,\cdots,|x_N| = r_N\}. \tag{34}$$

Defining λ by (29) (or equivalently (30)) and defining $\Pi(\lambda,\cdots,\lambda)$ as the set of probability measures on $\mathbb{R}_+^N$ having each marginal equal to λ, consider

$$\inf_{\widetilde{\gamma} \in \Pi(\lambda,\cdots,\lambda)} \int_{\mathbb{R}_+^N} \widetilde{c}(r_1,\cdots,r_N)d\widetilde{\gamma}(r_1,\cdots,r_N). \tag{35}$$

We claim that $\inf(33) = \inf(35)$. The inequality $\inf(33) \geq \inf(35)$ is easy: take $\gamma \in \Pi(\rho,\cdots,\rho)$ and define its *radial component* $\widetilde{\gamma}$ by

$$\int_{\mathbb{R}_+^N} F(r_1,\cdots,r_N)d\widetilde{\gamma}(r_1,\cdots,r_N) := \int_{\mathbb{R}^{dN}} F(|x_1|,\cdots,|x_N|)d\gamma(x_1,\cdots,x_N), \ \forall F \in C_c(\mathbb{R}_+^N), \tag{36}$$

it is obvious that $\widetilde{\gamma} \in \Pi(\lambda, \cdots, \lambda)$ and since $c(x_1, \cdots, x_N) \geq \widetilde{c}(|x_1|, \cdots, |x_N|)$, the inequality $\inf(33) \geq \inf(35)$ easily follows. To show the converse inequality, we use duality. Indeed, by standard convex duality, we have

$$\inf(33) = \sup_u \Big\{ N \int_{\mathbb{R}^d} u(x)\rho(x)dx \ : \ \sum_{i=1}^N u(x_i) \leq c(x_1, \cdots, x_N) \Big\} \tag{37}$$

and similarly

$$\inf(35) = \sup_v \Big\{ N \int_{\mathbb{R}_+} v(r)\lambda(r)dr \ : \ \sum_{i=1}^N v(r_i) \leq \widetilde{c}(r_1, \cdots, r_N) \Big\}. \tag{38}$$

Now since ρ is radially symmetric and the constraint of (37) is invariant by changing u by $u \circ R$ with R a rotation (see (11)) , there is no loss of generality in restricting the maximization in (37) to potentials of the form $u(x_i) = w(r_i)$, but then the constraint of (37) implies that w satisfies the constraint of (38). Then we have $\inf(33) = \sup(37) \leq \sup(38) = \inf(35)$. Note then that $\gamma \in \Pi(\rho, \cdots, \rho)$ solves (33) if and only if its radial component $\widetilde{\gamma}$ solves (33) and $c(x_1, \cdots, x_N) = \widetilde{c}(|x_1|, \cdots, |x_N|)$ γ-a.e. Therefore (33) gives the optimal radial component, whereas the extra condition $c(x_1, \cdots, x_N) = \widetilde{c}(|x_1|, \cdots, |x_N|)$ γ-a.e. gives an information on the angular distribution of γ.

3 Iterative Bregman Projections

Numerics for multi-marginal problems have so far not been extensively developed. Discretizing the multi-marginal problem leads to the linear program (41) where the number of constraints grows exponentially in N, the number of marginals. In this section, we present a numerical method which is not based on linear programming techniques, but on an entropic regularization and the so-called alternate projection method. It has recently been applied to various optimal transport problems in [12] and [2].

The initial idea goes back to von Neumann [26], [25] who proved that the sequence obtained by projecting orthogonally iteratively onto two affine subspaces converges to the projection of the initial point onto the intersection of these affine subspaces. Since the seminal work of Bregman [3], it is by now well-known that one can extend this idea not only to several affine subspaces (the extension to convex sets is due to Dyskstra but we won't use it in the sequel) but also by replacing the euclidean distance by a general Bregman divergence associated to some suitable strictly and differentiable convex function f (possibly with a domain) where we recall that the Bregman divergence associated with f is given by

$$D_f(x,y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle. \tag{39}$$

In what follows, we shall only consider the Bregman divergence (also known as the Kullback-Leibler distance) associated to the Boltzmann/Shannon entropy $f(x) := \sum_i x_i(\log x_i - 1)$ for non-negative x_i. This Bregman divergence (restricted to probabilities i.e. imposing the normalization $\sum_i x_i = 1$) is the Kullback-Leibler distance or relative entropy:

$$D_f(x,y) = \sum_i x_i \log\left(\frac{x_i}{y_i}\right).$$

Bregman distances are used in many other applications most notably image processing, see [18] for instance.

3.1 The Discrete Problem and its Entropic Regularization

In this section we introduce the discrete problem solved using the iterative Bregman projections [3]. From now on, we consider the problem (12)

$$\min_{\gamma_N \in \mathscr{C}} \int_{(\mathbb{R}^d)^N} c(x_1, \cdots, x_N)\gamma_N(x_1, \cdots, x_N)dx_1 \cdots dx_N, \tag{40}$$

where N is the number of marginals (or electrons), $c(x_1, ..., x_N)$ is the Coulomb cost, γ_N the transport plan, is the probability distribution over $(\mathbb{R}^d)^N$ and $\mathscr{C} := \bigcap_{i=1}^N \mathscr{C}_i$ with $\mathscr{C}_i := \{\gamma_N \in Prob\{(\mathbb{R}^d)^N\} | \, \pi_i \gamma_N = \rho\}$ (we remind the reader that electrons are indistinguishable so the N marginals coincide with ρ).

In order to discretize (40), we use a discretisation with M_d points of the support of the kth electron density as $\{x_{j_k}\}_{j_k=1,\cdots,M_d}$. If the densities ρ are approximated by $\sum_{j_k} \rho_{j_k} \delta_{x_{j_k}}$, we get

$$\min_{\gamma \in \mathscr{C}} \sum_{j_1,\cdots j_N} c_{j_1,\cdots,j_N} \gamma_{j_1,\cdots,j_N}, \tag{41}$$

where $c_{j_1,\cdots,j_N} = c(x_{j_1}, \cdots, x_{j_N})$ and the transport plan support for each coordinate is restricted to the points $\{x_{j_k}\}_k = 1, \cdots, M_d$ thus becoming a $(M_d)^N$ matrix again denoted γ with elements $\gamma_{j_1,\cdots,j_N}$. The marginal constraints $\mathscr{C}_i$ becomes

$$\mathscr{C}_i := \{\gamma \in \mathbb{R}_+^{(M_d)^N} | \sum_{j_1,\ldots,j_{i-1},j_{i+1},\ldots,j_N} \gamma_{j_1,\ldots,j_N} = \rho_{j_i}, \forall j_i = 1, \cdots, M_d\}. \tag{42}$$

Recall that the electrons are indistinguishable, meaning that they have same densities : $\rho_{j_k} = \rho_{j_{k'}}, \forall j, \forall k \neq k'$.

The discrete optimal transport problem (41) is a linear program problem and is dual to the discretization of (15)

$$\begin{aligned} &\max_{u_j} \sum_{j=1}^{M} N u_j \rho_j \\ &s.t. \sum_{i=1}^{N} u_{j_i} \leq c_{j_1 \cdots j_N} \quad \forall j_i = 1, \cdots, M_d \end{aligned} \tag{43}$$

where $u_j = u_{j_i} = u(x_{j_i})$. Thus the primal (41) has $(M_d)^N$ unknown and $M_d \times N$ linear constraints and the dual (43) only M_d unknown but still $(M_d)^N$ constraints. They are computationally not solvable with standard linear programming methods even for small cases in the multi-marginal case.

A different approach consists in computing the problem (41) regularized by the entropy of the joint coupling. This regularization dates to E. Schrödinger [32] and it has been recently introduced in machine learning [12] and economics [15] (we refer the reader to [2] for an overview of the entropic regularization and the iterative Bregman projections in OT). Thus, we consider the following discrete regularized problem

$$\min_{\gamma \in \mathscr{C}} \sum_{j_1, \cdots j_N} c_{j_1, \cdots, j_N} \gamma_{j_1, \cdots, j_N} + \varepsilon E(\gamma), \tag{44}$$

where $E(\gamma)$ is defined as follows

$$E(\gamma) = \begin{cases} \sum_{j_1, \cdots j_N} \gamma_{j_1, \cdots, j_N} \log(\gamma_{j_1, \cdots, j_N}) \text{ if } \gamma \geq 0 \\ +\infty \text{ otherwise.} \end{cases} \tag{45}$$

After elementary computations, we can re-write the problem as

$$\min_{\gamma \in \mathscr{C}} KL(\gamma | \bar{\gamma}) \tag{46}$$

where $KL(\gamma|\bar{\gamma}) = \sum_{i_1, \ldots, i_N} \gamma_{i_1, \ldots, i_N} (\log(\frac{\gamma_{i_1, \ldots, i_N}}{\bar{\gamma}_{i_1, \ldots, iN}}))$ is the Kullback-Leibler distance and

$$\bar{\gamma}_{i_1, \ldots, i_N} = e^{-\frac{c_{j_1, \cdots, j_N}}{\varepsilon}}. \tag{47}$$

As explained in section 1.2, when the transport plan γ is concentrated on the graph of a transport map which solves the Monge problem, after discretisation of the densities, this property is lost along but we still expect the γ matrix to be sparse. The entropic regularization will spread the support and this helps to stabilize the computation: it defines a strongly convex program with a unique solution γ^ε which can be obtained through elementary operations (we detail this in section 3.3 for both the continuous and discrete framework). The regularized solutions γ^ε then converge to $\gamma^\star$, the solution of (41) with minimal entropy, as $\varepsilon \to 0$ (see [9] for a detailed asymptotic analysis and the proof of exponential convergence). Let us now apply the iterative Bregman projections to find the minimizer of (46).

3.2 Alternate Projections

The main idea of the iterative Bregman projections (we call it *Bregman* as the Kullback-Leibler distance is also called Bregman distance, see [3]) is to construct a sequence γ^n (which converges to the minimizer of (46)) by alternately projecting on each set $\mathscr{C}_i$ with respect to the Kullback-Leibler distance. Thus, the iterative KL (or Bregman) projections can be written

$$\begin{cases} \gamma^0 & = \bar{\gamma} \\ \gamma^n & = P^{KL}_{\mathscr{C}_n}(\gamma^{n-1}) \quad \forall n > 0 \end{cases} \tag{48}$$

where we have extended the indexing of the set by $N-$periodicity such that $\mathscr{C}_{n+N} = \mathscr{C}_n \quad \forall n \in \mathbb{N}$ and $P^{KL}_{\mathscr{C}_n}$ denotes the KL projection on $\mathscr{C}_n$.

The convergence of γ^n to the unique solution of (46) is well known, it actually holds for large classes of Bregman distances and in particular the Kullback-Leibler divergence as was proved by Bauschke and Lewis [1]

$$\gamma^n \to P^{KL}_{\mathscr{C}}(\bar{\gamma}) \text{ as } n \to \infty.$$

Remark 3. If the convex sets $\mathscr{C}_i$ are not affine sub-sets (that is not our case), γ^n converges toward a point of the intersection which is not the KL projection of $\bar{\gamma}$ anymore so that a correction term is needed as provided by Dykstra's algorithm (we refer the reader to [2]).

The KL projection on $\mathscr{C}_i \quad i = 1, ..., N$ can be computed explicitly as detailed in the following proposition

Proposition 1. *For $\bar{\gamma} \in (\mathbb{R}_+)^{M_d{}^N}$ the projection $P^{KL}_{\mathscr{C}_i}(\bar{\gamma})$ is given by*

$$P^{KL}_{\mathscr{C}_i}(\bar{\gamma})_{j_1,...,j_N} = \rho_{j_i} \frac{\bar{\gamma}_{j_1,...,j_N}}{\sum_{k_1,...,k_{i-1},k_{i+1},...,k_N} \bar{\gamma}_{k_1,...,k_N}} \quad \forall j_i = 1, ..., M_d. \tag{49}$$

Proof. Introducing Lagrange multipliers λ_{j_i} associated to the constraint $\mathscr{C}_i$

$$\sum_{j_1,...,j_{i-1},j_{i+1},...,j_N} \gamma_{j_1,...,j_N} = \rho_{j_i} \tag{50}$$

the KL projection is given by the optimality condition :

$$\log(\frac{\gamma_{j_1,...,j_N}}{\bar{\gamma}_{j_1,...,j_N}}) - \lambda_{j_i} = 0 \tag{51}$$

so that

$$\gamma_{j_1,...,j_N} = C_{j_i} \bar{\gamma}_{j_1,...,j_N}, \tag{52}$$

where $C_{j_i} = e^{\lambda_{j_i}}$. If we substitute (52) in (50), we get

$$C_{j_i} = \rho_{j_i} \frac{1}{\sum_{k_1,\dots,k_{i-1},k_{i+1},\dots,k_N} \bar{\gamma}_{k_1,\dots,k_N}} \tag{53}$$

which gives (49).

3.3 From the Alternate Projections to the Iterative Proportional Fitting Procedure

The alternate projection procedure (48) is performed on M_d^N matrices. Moreover each projection (49) involves computing partial sum of this matrix. The total operation cost of each Bregman iteration scales like $O(M_d^{2N-1})$.

In order to reduce the cost of the problem, we use an equivalent formulation of the Bregman algorithm known as the Iterative Proportional Fitting Procedure (IPFP). Let us consider the problem (46) in a continous measure setting and, for simplicity, 2-marginals framework

$$\min_{\{\gamma|\pi_1(\gamma)=\rho,\pi_2(\gamma)=\rho\}} \int \log(\frac{d\gamma}{d\bar{\gamma}})d\gamma, \tag{54}$$

where ρ, ρ and $\bar{\gamma}$ are nonnegative measures. The aim of the IPFP is to find the KL projection of $\bar{\gamma}$ on $\Pi(\rho,\rho)$ (see (47) for the definition of $\bar{\gamma}$ which depends on the cost function).

Under the assumption that the value of (54) is finite, Rüschendorf and Thomsen (see [31]) proved that a unique KL-projection γ^* exists and that it is of the form

$$\gamma^*(x,y) = a(x)b(y)\bar{\gamma}(x,y), \quad a(x) \geq 0, \quad b(y) \geq 0. \tag{55}$$

From now on, we consider (with a sligthly abuse of notation) Borel measures with densities γ, $\bar{\gamma}$, ρ and ρ w.r.t. the suitable Lebesgue measure. a and b can be uniquely determined by the marginal condition as follows

$$\begin{aligned} a(x) &= \frac{\rho(x)}{\int \bar{\gamma}(x,y)b(y)dy}, \\ b(y) &= \frac{\rho(y)}{\int \bar{\gamma}(x,y)a(x)dx}. \end{aligned} \tag{56}$$

Then, IPFP is defined by the following recursion

$$\begin{aligned} b_0 = 1, \quad a_0 &= \rho, \\ b_{n+1}(y) &= \frac{\rho(y)}{\int \bar{\gamma}(x,y)a_n(x)dx}, \\ a_{n+1}(x) &= \frac{\rho(x)}{\int \bar{\gamma}(x,y)b_{n+1}(y)dy}. \end{aligned} \tag{57}$$

Moreover, we can define the sequence of joint densities (and of the corresponding measures)

$$\gamma^{2n}(x,y) := a^n(x)b^n(y)\bar{\gamma}(x,y) \quad \gamma^{2n+1} := a^n(x)b^{n+1}(y)\bar{\gamma}(x,y), \quad n \geq 0. \tag{58}$$

Rüschendorf proved (see [30]) that γ^n converges to the KL-projection of $\bar{\gamma}$. We can, now, recast the IPFP in a discrete framework, which reads as

$$\gamma_{ij} = a_i b_j \bar{\gamma}_{ij}, b_j^0 = 1, \quad a_i^0 = \rho_i, \tag{59}$$

$$\begin{aligned} b_j^{n+1} &= \frac{\rho_j}{\sum_i \bar{\gamma}_{ij} a_i^n}, \\ a_i^{n+1} &= \frac{\rho_i}{\sum_j \bar{\gamma}_{ij} b_j^{n+1}}, \end{aligned} \tag{60}$$

$$\gamma_{ij}^{2n} = a_i^n \bar{\gamma}_{ij} b_j^n \quad \gamma_{ij}^{2n+1} = a_i^n \bar{\gamma}_{ij} b_j^{n+1}. \tag{61}$$

By definition of γ_{ij}^n, notice that

$$\bar{\gamma}_{ij} b_j^n = \frac{\gamma_{ij}^{2n-1}}{a_i^{n-1}} \text{ and } a_i^n \bar{\gamma}_{ij} = \frac{\gamma_{ij}^{2n}}{b_j^n}$$

and if (61) is re-written as follows

$$\begin{aligned} \gamma_{ij}^{2n} &= \rho_i \frac{\bar{\gamma}_{ij} b_j^n}{\sum_k \bar{\gamma}_{ik} b_k^n} \\ \gamma_{ij}^{2n+1} &= \rho_j \frac{\bar{\gamma}_{ij} a_i^n}{\sum_k \bar{\gamma}_{kj} a_k^n} \end{aligned} \tag{62}$$

then we obtain

$$\begin{aligned} \gamma_{ij}^{2n} &= \rho_i \frac{\gamma_{ij}^{2n-1}}{\sum_k \gamma_{ik}^{2n-1}} \\ \gamma_{ij}^{2n+1} &= \rho_j \frac{\gamma_{ij}^{2n}}{\sum_k \gamma_{kj}^{2n}}. \end{aligned} \tag{63}$$

Thus, we exactly recover the Bregman algorithm described in the previous section, for 2 marginals.

The extension to the multi-marginal framework is straightforward but cumbersone to write. It leads to a problem set on N M_d-dimensional vectors $a_{j,i_{(\cdot)}}, \quad j = 1, \cdots, N, \quad i_{(\cdot)} = 1, \cdots, M_d$. Each update takes the form

$$a_{j,i_j}^{n+1} = \frac{\rho_{i_j}}{\sum_{i_1,i_2,\ldots i_{j-1},i_{j+1},\ldots,i_N} \bar{\gamma}_{i_1,\ldots,i_N} a_{1,i_1}^{n+1} a_{2,i_2}^{n+1} \ldots a_{j-1,i_{j-1}}^{n+1} a_{j+1,i_{j+1}}^{n} \ldots a_{N,i_N}^{n}}, \tag{64}$$

Where each i_k takes values in $\{1, \cdots, M_d\}$.

Note that we still need a constant M_d^N cost matrix $\bar{\gamma}$. Thanks to the symmetry and separability properties of the cost function (see (11) and (47)) , it is possible to replace it by a $N(N-1)/2$ product of M_d^2 matrices. This is already a big improvement from the storage point of view. Further simplifications are under investigations but the brute force IPFP operational cost therefore scales like $O(NM_d^{N+1})$ which provides a small improvement over the Bregman iterates option.

3.4 A heuristic refinement mesh strategy

We will use a heuristic refinement mesh strategy allowing to obtain more accuracy without increasing the computational cost and memory requirements. This idea was introduced in [27] for the adaptative resolution of the pure Linear Programming formulation of the Optimal Transportation problem, i.e without the entropic regularisation.

If the optimal transport plan is supported by a lower dimensional set, we expect the entropic regularisation to be concentrated on a mollified version of this set. Its width should decrease with the entropic parameter ε if the discretisation is fine enough. Working with a fixed ε, the idea is to apply coarse to fine progressive resolution and work with a sparse matrix γ. At each level, values below a threshold are filtered out (set to 0), then new positive values are interpolated on a finer grid (next level) where γ is strictly positive.

To simplify the exposition, we describe the algorithm for $2-$marginals in $1D$ and take a $\sqrt{M}$ gridpoints discretization of $I = [a,b] \in \mathbb{R}$:

1. we start with a cartesian M gridpoints mesh on $I \times I$ to approximate transport plan γ^ε, obtained by running the IPFP on a coarse grid.
2. we take $m_c(j) = max_i \gamma_{ij}^\varepsilon$ and $m_r(i) = max_j \gamma_{ij}^\varepsilon$ which are the maximum values over the rows and over the columns respectively, and we define
$$m = \min[\min_j(m_c(j)), \min_i(m_r(i))].$$
We will refine the grid only inside the level curve $\gamma^\varepsilon = \xi m$ where we expect the finer solution is supported, see figure 3.
3. In order to keep approximately the same number of element in the sparse matrix γ at each level we refine the grid as follows : Let $\mathscr{T} := \{(i,j) | \gamma_{ij}^\varepsilon \geq \xi m\}$ and $M_{\mathscr{T}} := \sharp\mathscr{T}$ and $r := M_{\mathscr{T}}/M$, then the size of the grid at the next level is $M^{new} = M/r$.

4. We compute the interpolation $\gamma_{M^{new}}$ of the old transport plan γ_M on the finer grid.
5. Elements of $\gamma_{M^{new}}$ below the fixed threshold ξm are filtered out, i.e are fixed to 0 and are not used in the IPFP sum computations, see figure 3.

6. Finally, a new IPFP computation is performed and it can be initialised with an interpolation of the data at the previous level ($\bar{\gamma}$ can be easly re-computed on the gridpoints where $\gamma_{M^{new}}$ is strictly positive).

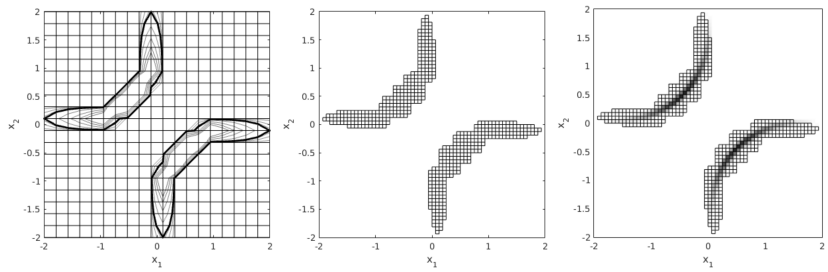

Fig. 3 *Left: $\mathscr{T}$ is the set of grid points inside the level curve $\gamma = \xi\, m$ ($\xi = 0.9$) (the bold line curve). Center: The new grid after the refinement. Right: The transport Plan after a new IPFP computation*

4 Numerical Results

4.1 $N = 2$ electrons: comparison between numerical and analytical results

In order to validate the numerical method, we now compare some numerical results for 2 electrons in dimension $d = 1, \cdots, 3$ with the analytical results from section 2.2. Let us first consider a uniform density (as (17) with $a = 2$) in $1D$. In table 1, we analyze the performance of the numerical method by varying the parameter ε. We notice that the error becomes smaller by decreasing the regularizing parameter, but the drawback is that the method needs more iterations to converge. Figure 4 shows the Kantorovich potential, the co-motion function which can be recovered from the potential by using (16) and the transport plan. The simulation is performed with a discretization of (17) with $a = 2$, $M = 1000$ (gridpoints) and $\varepsilon = 0.004$.

As explained in section 2.2.3, we can also compute the co-motion for a radially symmetric density. We have tested the method in $2D$ and $3D$, figure 5 and 6 respectively, by using the normalized uniform density on the unit ball. Moreover, in the radial case we have proved that the OT problem can be reduced to a $1-$dimensional problem by computing $\widetilde{c}$ which is trivial for the 2 electrons case: let us set the problem in $2D$ in polar coordinates (r_1, θ_1) and (r_2, θ_2), for the first and the second electron respectively (without loss of generality we can set $\theta_1 = 0$), then it is easy to verify that the minimum is achieved with $\theta_2 = \pi$. Figure 5 shows the Kantorovich

potential (the radial component $v(r)$ as defined in section 2.2.4), the co-motion and the transport plan for the $2-$dimensional case, the simulation is performed with $M = 1000$ and $\varepsilon = 0.002$. In figure 6 we present the result for th $3-$dimensional case, the simulation is performed with $M = 1000$ and $\varepsilon = 0.002$.

Remark 4. One can notice that, in the case of a uniform density, the transport plan presents a concentration of mass on the boundaries. This is a combined effect of the regularization and of the fact that the density has a compact support.

ε	Error ($\|u^\varepsilon - u\|_\infty / \|u\|_\infty$)	Iteration
0.256	0.1529	11
0.128	0.0984	16
0.064	0.0578	25
0.032	0.0313	38
0.016	0.0151	66
0.008	0.0049	114
0.004	0.0045	192

Table 1 *Numerical results for uniform density in 1D. u^ε is the numerical Kantorovich potential and u is the analytical one.*

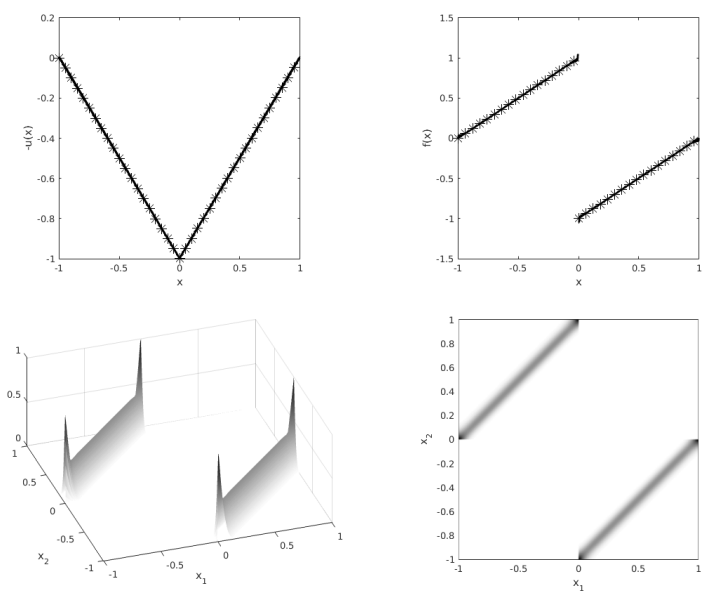

Fig. 4 *Top-Left: Kantorovich Potential $u(x)$. Top-Right: Numerical co-motion function (solid line) and analytical co-motion (star-solid line) . Bottom-Left: Transport plan $\tilde{\gamma}$. Bottom-Right: Support of $\tilde{\gamma}$.*

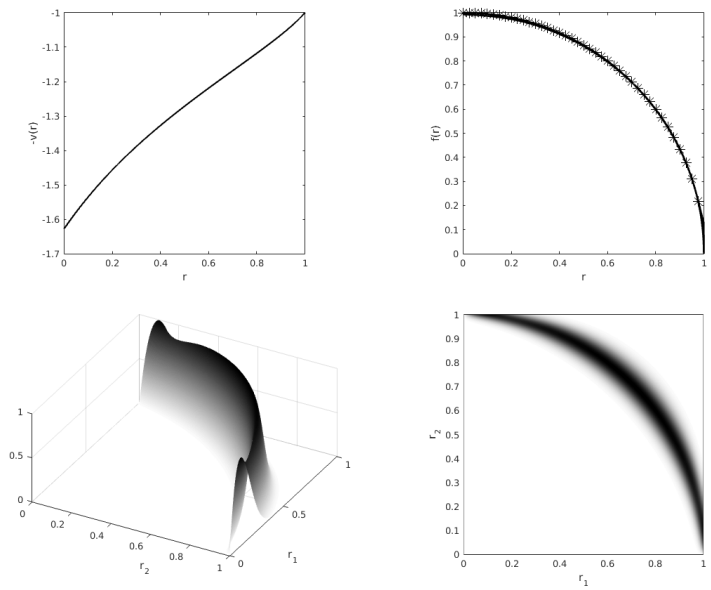

Fig. 5 *Top-Left: Kantorovich Potential $v(r)$. Top-Right: Numerical co-motion function (solid line) and analytical co-motion (star-solid line) . Bottom-Left: Transport plan $\tilde{\gamma}$. Bottom-Right: Support of $\tilde{\gamma}$.*

4.2 $N = 2$ *electrons in dimension* $d = 3$ *: Helium atom*

Once we have validated the method with some analytical examples, we solve the regularized problem for the Helium atom by using the electron density computed in [13]. In figure 7, we present the electron density, the Kantorovich potential and the transport plan. The simulation is performed with a discretization of $[0,4]$ with $M = 1000$ and $\varepsilon = 5\,10^{-3}$. We can notice the potential correctly fits the asymptotic behaviour from [33], namely $v(r) \sim \dfrac{N-1}{|r|}$ for $r \to \infty$, where N is the number of electrons.

4.3 $N = 3$ *electrons in dimension* $d = 1$

We present now some results for the $1-$dimensional multi-marginal problem with $N = 3$. They are validated against the analytical solutions given in section 2.2.2. We recall that splitting ρ into three tertiles ρ_i with equal mass, we will have $\rho_1 \to \rho_2$,

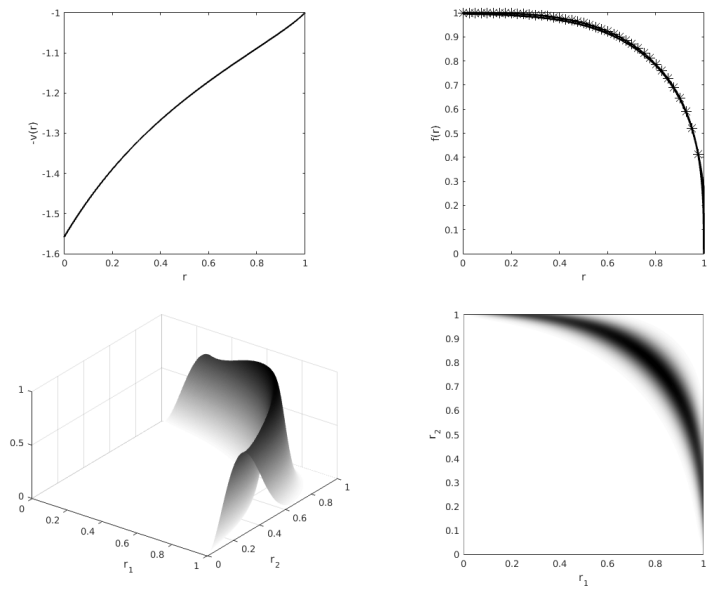

Fig. 6 *Top-Left: Kantorovich Potential $v(r)$. Top-Right: Numerical co-motion function (solid line) and analytical co-motion (star-solid line) . Bottom-Left: Transport plan $\tilde{\gamma}$. Bottom-Right: Support of $\tilde{\gamma}$.*

$\rho_2 \to \rho_3$ and $\rho_3 \to \rho_1$.

In table 2, we present the perfomance of the method for a uniform density on $[0,1]$ by varying ε and, as expected, we see the same behaviour as in the 2 marginals case. Figure 8 shows the Kantorovich potential and the projection of the transport plan onto two marginals (namely $\gamma^2 = \pi_{12}(\gamma^\varepsilon)$). The support gives the relative positions of two electrons.

The simulation is performed on a discretization of $[0,1]$ with a uniform density, $M = 1000$ and $\varepsilon = 0.02$. If we focus on the support of the projected transport plan we can notice that the numerical solution correctly reproduces the prescribed behavior The concentration of mass is again due to the compact support of the density, which is not the case of the gaussian as one can see in figure 9. In figure 9 we present the numerical results for $\rho(x) = e^{-x^2}/\sqrt{\pi}$. The simulation is performed on the discretization of $[-2.5, 2.5]$ with $M = 1000$ and $\varepsilon = 0.008$.

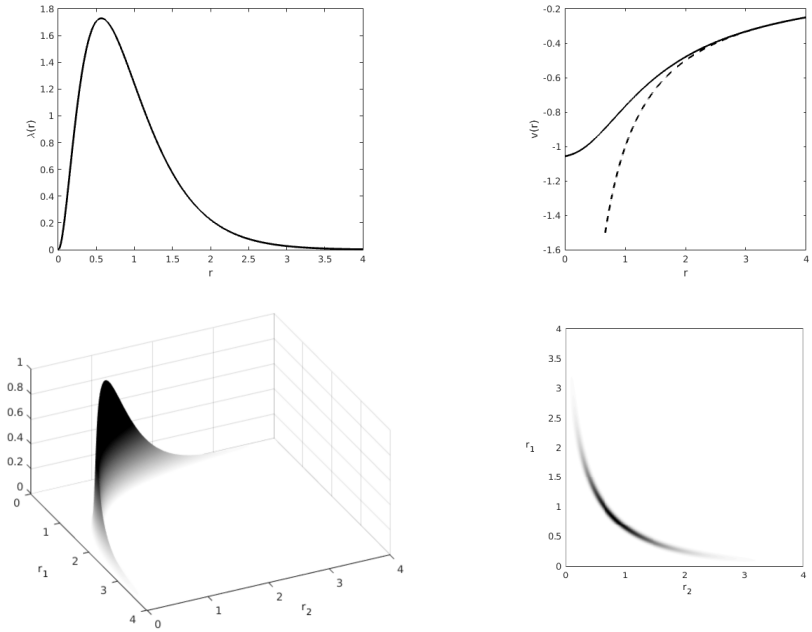

Fig. 7 *Top-Left: Helium density $\lambda(r) = 4\pi r^2 \rho(r)$. Top-Right: Kantorovich Potential $v(r)$ (blue) and asymptotic behaviour (red) $v(r) \sim \frac{1}{r} \quad r \to \infty$. Bottom-Left: Transport plan $\tilde{\gamma}$. Bottom-Right: Support of $\tilde{\gamma}$. All quantities are in Hartree atomic units.*

ε	Error ($\|u^\varepsilon - u\|_\infty / \|u\|_\infty$)	Iteration
0.32	0.0658	121
0.16	0.0373	230
0.08	0.0198	446
0.04	0.0097	878
0.02	0.0040	1714

Table 2 *Numerical results for uniform density in 1D and three electrons. u^ε is the numerical Kantorovich potential and u is the analytical one.*

4.4 $N = 3$ *electrons in dimension* $d = 3$ *radial case : Litium atom*

We finally perform some simulations for the radial 3−dimensional case for $N = 3$. As for the 3−dimensional case with 2 marginals we solve the reduced problem: let us consider the spherical coordinates (r_i, θ_i, ϕ_i) with $i = 1, \cdots, 3$ and we fix $\theta_1 = 0$ and $\phi_1 = \phi_2 = 0$ (the first electrons defines the z axis and the second one is on the xz plane). We then notice that $\phi_3 = 0$ as the electrons must be on the same plane of the nucleus to achieve compensation of forces (one can see it by computing the optimality conditions), so we have to minimize on θ_2 and θ_3 in order to obtain $\widetilde{c}$.

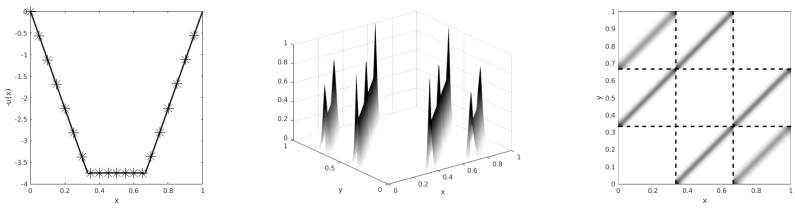

Fig. 8 *Left: Numerical Kantorovich potential $u(x)$ (solid line) and analytical potential (star-solid line). Center: Projection of the transport plan $\pi_{12}(\gamma(x,y,z))$. Rigth: Support of $\pi_{12}(\gamma(x,y,z))$ The dot-dashed lines delimit the intervals where ρ_i, with $i = 1,\cdots,3$, are defined.*

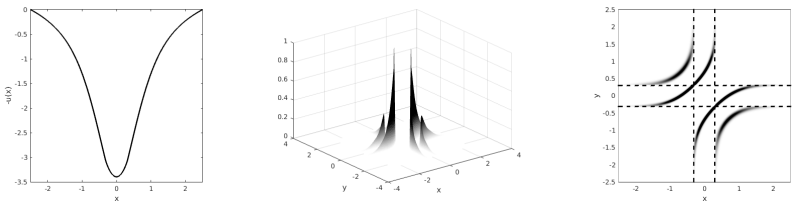

Fig. 9 *Left: Kantorovich potential $u(x)$. Center: Projection of the transport plan $\pi_{12}(\gamma(x,y,z))$. Rigth: Support of $\pi_{12}(\gamma(x,y,z))$. The dot-dashed lines delimit the intervals where ρ_i, with $i = 1,\cdots,3$, are defined.*

Figure 10 shows the electron density of the Litium (computed in [5]), the Kantorovich Potential (and the asymptotic behavior) and the projection of the transport plan onto two marginals $\tilde{\gamma}^2 = \pi_{12}(\tilde{\gamma}^\varepsilon)$. The support gives the relative positions of two electrons.

The simulation is performed on a discretization of $[0,8]$ with $M = 300$ and $\varepsilon = 0.007$. Our results show (taking into account the regularization effect) a concentrated transport plan for this kind of density and they match analogous result obtained in [33]. If we focus on the support of the transport plan we can notice that the optimal solution forces the electrons to occupy three different regions as conjectured in [33].

5 conclusion

We have presented a numerical scheme for solving multi-marginal OT problems arising from DFT. This is a challenging problems, not only because of the unusual features of the Coulomb cost which is singular and repulsive but also due to the high dimension of the space of plans.

Using an entropic regularization gives rise to a Kullback-Leibler projection problem onto the intersection of affine subsets given by the marginal constraints. Because

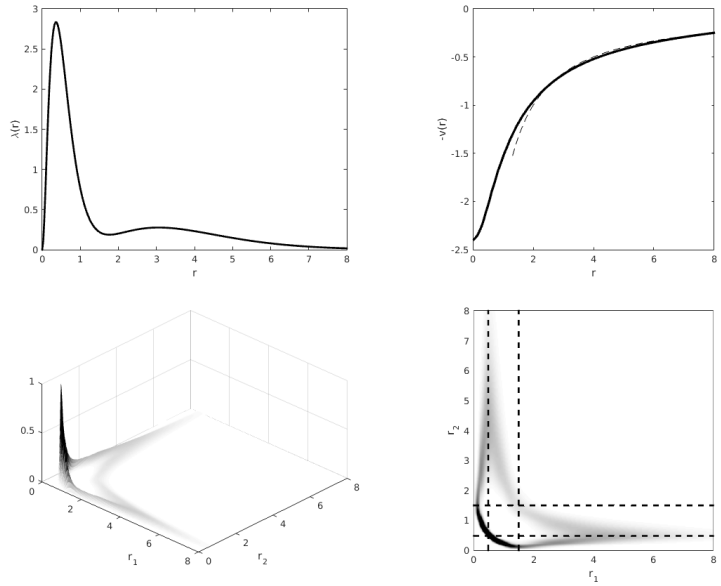

Fig. 10 *Top-Left: Litium density $\lambda(r) = 4\pi r^2 \rho(r)$. Top-Right: Kantorovich Potential $v(r)$ (blue) and asymptotic behaviour (red) $v(r) \sim \frac{2}{r} \quad r \to \infty$. Bottom-Left: Projection of the Transport plan $\tilde{\gamma}^2 = \pi_{12}(\tilde{\gamma}^\varepsilon)$. Bottom-Right: Support of the projected transport plan $\tilde{\gamma}^2$. The dot-dashed lines delimit the three regions that the electrons must occupy, we computed them numerically following the idea in [33].All quantities are in Hartree atomic units.*

each projection is explicit, one can use Bregman's iterative projection algorithm to approximate the solution.

The power of such an iterative projection approach was recently emphasized in [12, 2] for the entropic regularization of optimal transport problems, we showed that is also well suited to treat the multi-marginal OT problem with Coulomb cost and leads to the same benefits in terms of convexification of the problem and simplicity of implemention.

The method presented here is just a preliminary step which is simple to implement and therefore easy to use in practice. The cost of solving the general DFT problem in dimension 3 for a large number of electrons is still unfeasible and we need to use radial symmetry simplification and also a heuristic refinement mesh strategy.

A lot of questions are left for future research : can IPFP be used for sharper approximations for DFT? Can one justify rigorously and quantitatively the mesh refinement strategy? How should the regularization parameter ε be chosen in practice? Does the entropic regularization have a physical interpretation?

Acknowledgements

We would like to thank Adam Oberman and Brendan Pass for many helpful and stimulating discussions as well as Paola Gori-Giorgi for sharing numerical details concerning the Helium and Litium atom.

We gratefully acknowledge the support of the ANR, through the project ISO-TACE (ANR-12-MONU-0013) and INRIA through the “action exploratoire" MOKA-PLAN.

References

1. Bauschke, H.H., Lewis, A.S.: Dykstra's algorithm with Bregman projections: a convergence proof. Optimization **48**(4), 409–427 (2000)
2. Benamou, J.D., Carlier, G., Cuturi, M., Nenna, L., Peyré, G.: Iterative bregman projections for regularized transportation problems. arXiv preprint arXiv:1412.5154 (2014)
3. Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR computational mathematics and mathematical physics **7**(3), 200–217 (1967)
4. Brenier, Y.: Polar factorization and monotone rearrangement of vector-valued functions. Comm. Pure Appl. Math. **44**(4), 375–417 (1991). DOI 10.1002/cpa.3160440402. URL http://dx.doi.org/10.1002/cpa.3160440402
5. Bunge, C.: The full ci density of the li atom has been computed with a very large basis set with 8 s functions and up to k functions. private communication
6. Buttazzo, G., De Pascale, L., Gori-Giorgi, P.: Optimal-transport formulation of electronic density-functional theory. Phys. Rev. A **85**, 062,502 (2012). DOI 10.1103/PhysRevA.85.062502. URL http://link.aps.org/doi/10.1103/PhysRevA.85.062502
7. Carlier, G., Ekeland, I.: Matching for teams. Econom. Theory **42**(2), 397–418 (2010). DOI 10.1007/s00199-008-0415-z. URL http://dx.doi.org/10.1007/s00199-008-0415-z
8. Colombo, M., De Pascale, L., Di Marino, S.: Multimarginal optimal transport maps for one-dimensional repulsive costs. Canad. J. Math. **67**, 350–368 (2015)
9. Cominetti, R., Martin, J.S.: Asymptotic analysis of the exponential penalty trajectory in linear programming. Mathematical Programming **67**(1-3), 169–187 (1994)
10. Cotar, C., Friesecke, G., Klüppelberg, C.: Density functional theory and optimal transportation with Coulomb cost. Communications on Pure and Applied Mathematics **66**(4), 548–599 (2013). DOI 10.1002/cpa.21437. URL http://dx.doi.org/10.1002/cpa.21437
11. Cotar, C., Friesecke, G., Pass, B.: Infinite-body optimal transport with coulomb cost. Calculus of Variations and Partial Differential Equations pp. 1–26 (2013)
12. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: Advances in Neural Information Processing Systems (NIPS) 26, pp. 2292–2300 (2013)
13. Freund, D.E., Huxtable, B.D., Morgan, J.D.: Variational calculations on the helium isoelectronic sequence. Phys. Rev. A **29**, 980–982 (1984). DOI 10.1103/PhysRevA.29.980. URL http://link.aps.org/doi/10.1103/PhysRevA.29.980
14. Friesecke, G., Mendl, C.B., Pass, B., Cotar, C., Klüppelberg, C.: N-density representability and the optimal transport limit of the hohenberg-kohn functional. The Journal of chemical physics **139**(16), 164,109 (2013)
15. Galichon, A., Salanié, B.: Matching with trade-offs: Revealed preferences over competing characteristics. Tech. rep., Preprint SSRN-1487307 (2009)
16. Gangbo, W., Święch, A.: Optimal maps for the multidimensional Monge-Kantorovich problem. Comm. Pure Appl. Math. **51**(1), 23–45 (1998). DOI 10.1002/(SICI)1097-0312(199801)51:1<23::AID-CPA2>3.0.CO;2-H. URL http://dx.doi.org/10.1002/(SICI)1097-0312(199801)51:1<23::AID-CPA2>3.0.CO;2-H

17. Ghoussoub, N., Maurey, B.: Remarks on multi-marginal symmetric Monge-Kantorovich problems. Discrete Contin. Dyn. Syst. **34**(4), 1465–1480 (2014)
18. Goldstein, T., Bresson, X., Osher, S.: Geometric applications of the split bregman method: Segmentation and surface reconstruction. Journal of Scientific Computing **45**(1-3), 272–293 (2010). DOI 10.1007/s10915-009-9331-z. URL http://dx.doi.org/10.1007/s10915-009-9331-z
19. Hohenberg, P., Kohn, W.: Inhomogeneous electron gas. Phys. Rev. **136**, B864–B871 (1964). DOI 10.1103/PhysRev.136.B864. URL http://link.aps.org/doi/10.1103/PhysRev.136.B864
20. Kantorovich, L.: On the transfer of masses (in russian). Doklady Akademii Nauk **37**(2), 227–229 (1942)
21. Kohn, W., Sham, L.J.: Self-consistent equations including exchange and correlation effects. Phys. Rev. **140**, A1133–A1138 (1965). DOI 10.1103/PhysRev.140.A1133. URL http://link.aps.org/doi/10.1103/PhysRev.140.A1133
22. Malet, F., Gori-Giorgi, P.: Strong correlation in kohn-sham density functional theory. Phys. Rev. Lett. **109**, 246,402 (2012). DOI 10.1103/PhysRevLett.109.246402. URL http://link.aps.org/doi/10.1103/PhysRevLett.109.246402
23. Mendl, C.B., Lin, L.: Kantorovich dual solution for strictly correlated electrons in atoms and molecules. Physical Review B **87**(12), 125,106 (2013)
24. Monge, G.: Mémoire sur la théorie des déblais et des remblais. De l'Imprimerie Royale (1781)
25. von Neumann, J.: Functional Operators. Vol. 1: Measures and integrals. Vol 2: The geometry of orthogonal spaces. 21, 22 (1950–1951)
26. Neumann, J.V.: On rings of operators. reduction theory. Annals of Mathematics **50**(2), pp. 401–485 (1949). URL http://www.jstor.org/stable/1969463
27. Oberman, A.: private communication. paper in preparation
28. Pass, B.: Uniqueness and Monge solutions in the multimarginal optimal transportation problem. SIAM Journal on Mathematical Analysis **43**(6), 2758–2775 (2011). DOI 10.1137/100804917. URL http://link.aip.org/link/?SJM/43/2758/1
29. Pass, B.: Multi-marginal optimal transport and multi-agent matching problems: uniqueness and structure of solutions. Discrete Contin. Dyn. Syst. **34**(4), 1623–1639 (2014). DOI 10.3934/dcds.2014.34.1623. URL http://dx.doi.org/10.3934/dcds.2014.34.1623
30. Ruschendorf, L.: Convergence of the iterative proportional fitting procedure. The Annals of Statistics **23**(4), 1160–1174 (1995)
31. Ruschendorf, L., Thomsen, W.: Closedness of sum spaces and the generalized Schrodinger problem. Theory of Probability and its Applications **42**(3), 483–494 (1998)
32. Schrodinger, E.: Uber die umkehrung der naturgesetze. Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math. **144**, 144–153 (1931)
33. Seidl, M., Gori-Giorgi, P., Savin, A.: Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities. Phys. Rev. A **75**, 042,511 (2007). DOI 10.1103/PhysRevA.75.042511. URL http://link.aps.org/doi/10.1103/PhysRevA.75.042511
34. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics Series. American Mathematical Society (2003). URL http://books.google.fr/books?id=GqRXYFxe0l0C
35. Villani, C.: Optimal transport: old and new, vol. 338. Springer (2009)