SIAM J. SCI. COMPUT.
Vol. 41, No. 4, pp. B773–B795

CONVEX RELAXATION APPROACHES FOR STRICTLY CORRELATED DENSITY FUNCTIONAL THEORY*

YUEHAW KHOO† AND LEXING YING†

Abstract. In this paper, we introduce methods from convex optimization to solve the multimarginal transport type problems that arise in the context of density functional theory. Convex relaxations are used to provide outer approximation to the set of N-representable 2-marginals and 3-marginals, which in turn provide lower bounds to the energy. We further propose rounding schemes to obtain upper bound to the energy. Numerical experiments demonstrate a gap of the order of 10^{-3} to 10^{-2} between the upper and lower bounds. The Kantorovich potential of the multimarginal transport problem is also approximated with a similar accuracy.

Key words. convex relaxation, strictly correlated density functional theory, semidefinite programming

AMS subject classifications. 49M20, 90C22, 90C25

DOI. 10.1137/18M1207478

1. Introduction.

We propose a novel convex relaxation framework for solving multimarginal optimal transport type problems in the context of density functional theory for strictly correlated electrons. More precisely, we consider the type problems that take the form

$$\inf_{\lambda_1,\ldots,\lambda_N,\mu\in\Pi(\lambda_1,\ldots,\lambda_N)} \sum_{i=1}^{N} g_i(\lambda_i) + \int_{X_1\times\cdots\times X_N} f(x_1,\ldots,x_N)d\mu(x_1,\ldots,x_N), \tag{1}$$
$$\mathcal{A}_i(\lambda_i) = b_i,\ i = 1,\ldots,N,$$

where $g_i(\cdot), i = 1,\ldots,N$ are convex functionals, $\mathcal{A}_i, i = 1,\ldots,N$, are some linear operators, and $\Pi(\lambda_1,\ldots,\lambda_N)$ denotes the space of probability measures on $X_1\times\cdots\times X_N$ with marginals $\lambda_1,\ldots,\lambda_N$. In this paper, the domain of the cost $X_1\times\cdots\times X_N$ is discrete and the cost function f has the form

$$f(x_1,\ldots,x_N) = \sum_{i,j=1,i>j}^{N} C_{ij}(x_i,x_j). \tag{2}$$

A particular situation that we are interested in is when $f(x_1,\ldots,x_N)$ and $\mu(x_1,\ldots,x_N)$ are symmetric when any x_i and x_j are swapped, i.e., $g_i := g$, and $C_{ij} := C$ for $i,j = 1,\ldots,N$. In such a situation, the task is to solve

$$\inf_{\lambda,\mu\in\Pi_{N,\mathrm{sym}}(\lambda)} g(\lambda) + \int_{X^N} f(x_1,\ldots,x_N)d\mu(x_1,\ldots,x_N), \quad \mathcal{A}(\lambda) = b, \tag{3}$$

*Submitted to the journal's Computational Methods in Science and Engineering section August 14, 2018; accepted for publication (in revised form) May 20, 2019; published electronically July 25, 2019.

https://doi.org/10.1137/18M1207478

Funding: This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program and the National Science Foundation under award DMS-1818449.

†Department of Mathematics, Stanford University, Stanford, CA 94305 (ykhoo@stanford.edu, lexing@stanford.edu).

where $\Pi_{N,\mathrm{sym}}(\lambda)$ denotes the space of symmetric probability measures on X^N with the marginals being λ. Solving this problem is particularly useful in the context of density functional theory (DFT), where the density for many-electrons is indeed symmetric. A brief introduction to how such a problem can arise in DFT when the electrons are strictly correlated is given in section 1.1. Although problem (3) is a linear programming problem for discrete X, the domain of optimization is exponentially large for any practical computation.

1.1. Background on DFT for strictly correlated electrons. A key task in DFT is to determine the minimum of an energy functional $E(\rho)$ of the 1-marginal

$$\rho(x) = \int_{X^{N-1}} |\psi(x, \dots, x_N)|^2 dx_2 dx_3 \dots dx_N, \tag{4}$$

where $\psi(x_1, \dots, x_N)$ is a many-body wavefunction for N electrons (due to the properties of electrons $|\psi(x_1, \dots, x_N)|^2$ is symmetric). In this paper, we consider an energy functional introduced in [8]

$$E(\rho) = V_{\mathrm{ee}}^{\mathrm{SCE}}(\rho) + E_{\mathrm{kd}}(\rho) + \int_X v_{\mathrm{ext}}(x)\rho(x)dx, \tag{5}$$

which is suitable for studying strongly correlated electrons. The functional $E_{\mathrm{kd}}(\rho)$ corresponds to kinetic energy with some correction terms, v_{ext} is some external potential (for example, potential exerted by nuclei), and the central object of the study is the strictly correlated density functional $V_{\mathrm{ee}}^{\mathrm{SCE}}(\rho)$ defined as

$$V_{\mathrm{ee}}^{\mathrm{SCE}}(\rho) := \inf_{\lambda,\mu \in \Pi_{\mathrm{N,sym}}(\lambda)} \int_{X^N} \sum_{i,j=1,i>j}^{N} \frac{1}{\|x_i - x_j\|} d\mu(x_1, \dots, x_N), \quad \lambda = \rho. \tag{6}$$

This framework of DFT gives rise to the following two problems:

- Solving for the strictly correlated density functional $V_{\mathrm{ee}}^{\mathrm{SCE}}(\rho)$ via the optimization problem (6). This is, in fact, the well-known multimarginal optimal transport problem.

- Direct minimization of the total energy functional $E(\cdot)$ in (5), when the kinetic energy $E_{\mathrm{kd}}(\rho)$ is either convex or negligible (thus can be dropped). In this case, the minimization problem takes the form

$$\begin{aligned} \inf_{\rho} \quad & V_{\mathrm{ee}}^{\mathrm{SCE}}(\rho) + E_{\mathrm{kd}}(\rho) + \int_X v_{\mathrm{ext}}(x)\rho(x)dx \\ & \Leftrightarrow \inf_{\rho} \quad E_{\mathrm{kd}}(\rho) + \int_X v_{\mathrm{ext}}(x)\rho(x)dx \qquad (7) \\ & \quad + \inf_{\lambda,\mu \in \Pi_{\mathrm{N,sym}}(\lambda), \lambda=\rho} \int_{X^N} \sum_{i,j=1,i>j}^{N} \frac{1}{\|x_i - x_j\|} d\mu(x_1, \dots, x_N) \\ & \Leftrightarrow \inf_{\lambda,\mu \in \Pi_{\mathrm{N,sym}}(\lambda)} E_{\mathrm{kd}}(\lambda) + \int_X v_{\mathrm{ext}}(x)\lambda(x)dx \qquad (8) \\ & \quad + \int_{X^N} \frac{1}{\|x_i - x_j\|} d\mu(x_1, \dots, x_N). \end{aligned}$$

Notice that the first problem, i.e., (6), takes the form of (3) when $\mathcal{A}$ is the identity and $b = \rho$, while the second problem, presented in (9), takes the form of (3) when the constraint $\mathcal{A}(\lambda) = b$ is absent.

1.2. Our contributions. In this paper, we propose a scheme with practical running time to approximately solve (3) where the approximation is observed to be sufficiently tight. To this end, we work with an equivalent formulation of (3) in terms of the 2-marginals. Although this seems to break the aforementioned complexity barrier for solving (3), enforcing that the 2-marginals being the marginalization of a probability measure on X^N, is nontrivial. Leveraging the results of [7] concerning the extreme points of the N-representable 2-marginals, we propose a semidefinite programming (SDP) relaxation, SDP-Coulomb, to provide an outer approximation to the set of N-representable 2-marginals, therefore breaking the complexity barrier in optimizing the high-dimensional measure in (3). The property of the proposed SDP is discussed in light of the results in [7]. We further propose a tighter convex relaxation SDP-Coulomb2 based on a formulation of (3) in terms of the 3-marginals. As the proposed convex relaxations only provide lower bounds to the energy, we further propose rounding schemes to give upper bounds. Numerical simulations show that the proposed approaches give a relative gap between the upper and lower bounds of size 10^{-3} to 10^{-2}, which, in turn, sets an upper bound on the approximation accuracy. Before delving into the details, in Figure 1 we show an example where we solve the multimarginal transport problem (6) with $N = 8$, $\rho(x) \propto \exp(-x^2/\sqrt{\pi})$, and the discrete domain X has size $|X| = 1600$. The running time is 2560s. Such a problem size would be impossible to be solved by traditional methods such as linear programming since it requires the storage of a tensor with 10^{25} entries. Moreover, in this example, we obtain an estimate of $V_{\mathrm{ee}}^{\mathrm{SCE}}(\rho)$ with 3.6e-04 error.

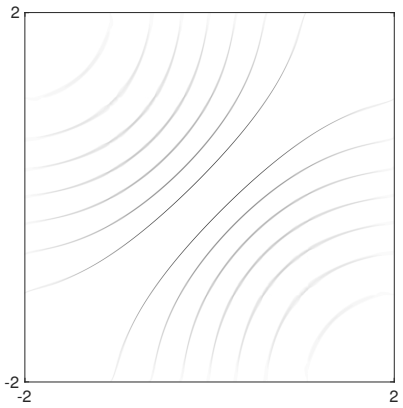


FIG. 1. *Approximating the* 2*-marginal of the solution to the multimarginal optimal transport problem* (6) *via solving SDP-Coulomb with* $\rho(x) \propto \exp(-x^2/\sqrt{\pi})$ *being the marginal. Here* $N = 8$, $|X| = 1600$, $d = 1$. *The error of the energy is* $3.6e-04$.

1.3. Prior works. The consideration of numerically solving an optimal transport problem with a Coulomb cost is a relatively new field. In [12], the dual problem to problem (3) is solved, via a parameterization of the dual function. In [5], linear programming is applied to solve the problem involving 2-electrons in three dimensions (3D) as part of a self-consistent DFT iterations. In [2], the Sinkhorn scaling algorithm is applied to an entropic regularized problem of (3). Although these methods have shown various levels of success in practice, the constraints or variables involved grow exponentially in the number of electrons.

1.4. Organization. In section 2, we detail the proposed SDP relaxation for problem (3) in terms of the 2-marginal. In section 3, we characterize the property of the SDP relaxation. In section 4, a further tightening of the SDP relaxation is proposed by formulating problem (3) in terms of the 3-marginal. In section 5, rounding schemes are provided to obtain an upper bound of the energy. In section 6, we demonstrate the effectiveness of the proposed method through numerical examples.

1.5. Notation. In what follows, I is used to denote the identity matrix as usual and we use A^T to denote the matrix transpose. For a p-dimensional tensor T, $T(j_1, j_2, \ldots, j_p)$ denotes its $(i_1, \ldots, i_p)$th entry. MATLAB notation ":" is used to extract a slice of a tensor. For example, for a matrix A, $A(:, i)$ gives the ith column of the matrix. $\mathbf{1}$ is used to denote an all-one vectors of appropriate size. For a matrix $A \in \mathbb{R}^{L\times L}$, the operator $\mathrm{diag}(A) \in \mathbb{R}^L$ extracts the diagonal of A and diag^* denotes the adjoint of diag. More precisely, for a vector $a \in \mathbb{R}^L$,

$$\mathrm{diag}^*(a) := \begin{bmatrix} a(1) & & \\ & \ddots & \\ & & a(L) \end{bmatrix}. \tag{9}$$

The notation $\odot$ and $\otimes$ are used to denote the Hadamard and tensor products respectively. For a p-dimensional tensor T, $\|T\|_F^2$ is defined as

$$\|T\|_F^2 := \sum_{i_1,\ldots,i_p} |T(i_1, \ldots, i_p)|^2. \tag{10}$$

2. Proposed method. In this section, we proposed an SDP relaxation to solve the equivalent problem of (3) in terms of the 2-marginals. In terms of the 2-marginals γ_{ij}, the cost of (3) is

$$g(\lambda) + \sum_{i,j=1,i<j}^{N} \int_{X^N} C_{ij}(x_i, x_j) d\gamma_{ij}(x_i, x_j) = g(\lambda) + \frac{N(N-1)}{2} \int_{X^2} C(x, y) d\gamma(x, y), \tag{11}$$

where $\gamma_{ij}(x_i, x_j) = \gamma(x_i, x_j)$ due to the symmetry of μ. The 2-marginal γ is called an N-representable measure, since it comes from the marginalization of a symmetric probability measure on X^N. A more general definition for k-marginal is given below.

Definition 1. *A k-marginal on X^N is called N-representable if it results from the marginalization of a symmetric probability distribution on X^N.*

As we consider a discrete state space X, problem (11) in terms of the discrete 2-marginals takes the form

$$\begin{aligned} \min_{\gamma\in\mathbb{R}^{L\times L}} \; & g(\gamma\mathbf{1}) + \frac{N(N-1)}{2}\mathrm{Tr}\big(C\gamma\big) \\ \text{s.t. } & \gamma \text{ is } N\text{-representable}, \\ & \mathrm{diag}(\gamma) = 0, \\ & \mathcal{A}(\gamma\mathbf{1}) = b. \end{aligned} \tag{12}$$

Here we added a problem-dependent constraint $\mathrm{diag}(\gamma) = 0$, due to the fact the Coulomb cost $C(\cdot, \cdot)$ is infinity when two arguments coincide. To derive an SDP relaxation to (12), one first needs a characterization of the N-representable 2-marginals. For this, we leverage the following result from [7], where $\mathrm{conv}(S)$ denotes the convex hull of a set S.

THEOREM 1. *The set of discrete N-representable 2-marginals is* $\operatorname{conv}(\Gamma_2)$ *where*

$$\Gamma_2 = \left\{ \frac{N}{N-1}\lambda\lambda^T - \frac{1}{N-1}\operatorname{diag}^*(\lambda) \mid \lambda \in \{0, 1/N, 2/N, \dots, 1\}^{|X|}, \quad \lambda^T \mathbf{1} = 1 \right\}. \tag{13}$$

Moreover, Γ_2 *is the set of extreme points of* $\operatorname{conv}(\Gamma_2)$.

Since we are interested in the 2-marginals γ where the diagonal element is zero, we characterize the subset $\tilde{\Gamma}_2 \subset \Gamma_2$ with this extra zero constraint in the following corollary. Let

$$\mathcal{B}_N(X) = \{\lambda \in \mathbb{R}^{|X|} \mid \lambda^T \mathbf{1} = 1,\ \lambda(i) \in \{0, 1/N\},\ i = 1, \dots, |X|\}, \tag{14}$$

which denotes the set of binarized probability vector on a discrete domain X.

COROLLARY 1. *Let*

$$\tilde{\Gamma}_2 = \left\{ \frac{N}{N-1}\lambda\lambda^T - \frac{1}{N-1}\operatorname{diag}^*(\lambda) \mid \lambda \in \mathcal{B}_N(X) \right\}, \tag{15}$$

then

$$\operatorname{conv}(\tilde{\Gamma}_2) = \{\gamma \in \mathbb{R}^{|X|\times|X|} \mid \gamma \text{ is } N\text{-representable},\ \operatorname{diag}(\gamma) = 0\}. \tag{16}$$

Moreover, $\tilde{\Gamma}_2$ *is the extreme points of* $\operatorname{conv}(\tilde{\Gamma}_2)$.

For completeness, a short proof of Corollary 1 is presented in section 3. With this characterization, an equivalent formulation of (12) is obtained as

$$\begin{aligned} \min_{\gamma \in \mathbb{R}^{|X|\times|X|}} \ & g(\gamma\mathbf{1}) + \frac{N(N-1)}{2}\operatorname{Tr}\big(C\gamma\big) \\ \text{s.t. } & \gamma \in \operatorname{conv}(\tilde{\Gamma}_2), \\ & \mathcal{A}(\gamma\mathbf{1}) = b. \end{aligned} \tag{17}$$

We claim that this is also equivalent to the following minimization problem:

$$\begin{aligned} \min_{\gamma,\lambda,a} \ & g(\gamma\mathbf{1}) + \frac{N(N-1)}{2}\operatorname{Tr}(C\gamma) \\ \text{s.t. } & \gamma = \frac{N}{N-1}\sum_{i=1}^{m} a(i)\lambda_i\lambda_i^T - \frac{1}{N-1}\operatorname{diag}^*\left(\sum_{i=1}^{m} a(i)\lambda_i\right), \\ & \sum_{i=1}^{m} a(i) = 1,\ a(i) \geq 0,\ i = 1, \dots, m, \\ & \lambda_i^T \mathbf{1} = 1,\ i = 1, \dots, m, \\ & \lambda_i \in \{0, 1/N\}^{|X|},\ i = 1, \dots, m, \\ & \mathcal{A}(\gamma\mathbf{1}) = b. \end{aligned} \tag{18}$$

Here, the first four constraints are equivalent to γ being an element in $\operatorname{conv}(\tilde{\Gamma}_2)$. The integer m specifies the number of elements in $\tilde{\Gamma}_2$ needed for representing γ, which depends on the number of linear constraints $\mathcal{A}(\gamma\mathbf{1}) = b$. For the purpose of this section it is not important to know what m is, and we can just treat it as an arbitrary integer. A detailed discussion on what m is for the problem considered is provided in section 5.2.

2.1. Convex relaxation. Problem (18) involves optimizating over the set $\mathcal{B}_N(X)$, which has a combinatorial complexity in the worst case. To cope with this issue, we propose the following convex relaxation to problems (12) and (18):

$$\min_{\gamma,\Lambda\in\mathbb{R}^{|X|\times|X|}} g(\gamma\mathbf{1}) + \frac{N(N-1)}{2}\mathrm{Tr}(C\gamma) \quad \text{(SDP-Coulomb)} \tag{19}$$
$$\text{s.t. } \gamma = \frac{N}{N-1}\Lambda - \frac{1}{N-1}\,\mathrm{diag}^*(\Lambda\mathbf{1}),$$
$$\mathcal{A}(\Lambda\mathbf{1}) = b,$$
$$\Lambda \succeq 0,$$
$$\Lambda \geq 0,$$
$$\mathbf{1}^T\Lambda\mathbf{1} = 1,$$
$$\mathrm{diag}(\Lambda) = \frac{1}{N}\Lambda\mathbf{1}.$$

The details of going from (18) to (19) are presented in the subsequent sections.

2.1.1. Changing the variables to Λ. We start to derive SDP-Coulomb from Problem (18). Instead of working with both sets of variables $\{\lambda_i\}_{i=1}^m$ and a as in Problem (18), we will only work with a single matrix variable Λ. First, let

$$\Lambda := \sum_{i=1}^m a(i)\lambda_i\lambda_i^T, \quad \lambda_i \in \mathcal{B}_N(X), \quad i = 1,\dots,m. \tag{20}$$

Since

$$\Lambda\mathbf{1} = \sum_{i=1}^m a(i)\lambda_i\lambda_i^T\mathbf{1} = \sum_{i=1}^m a(i)\lambda_i, \tag{21}$$

in terms of Λ the 2-marginal γ in (18) becomes

$$\gamma = \frac{N}{N-1}\Lambda - \frac{1}{N-1}\,\mathrm{diag}^*(\Lambda\mathbf{1}). \tag{22}$$

Notice that with such a change of variable,

$$\gamma\mathbf{1} = \Lambda\mathbf{1}. \tag{23}$$

2.1.2. Constraints on Λ. The variable Λ defined in (20) belongs to a nonconvex set as it is a quadratic form of the binarized vectors $\lambda_1,\dots,\lambda_m$. In order to obtain the convex program SDP-Coulomb, we only enforce certain necessary conditions of Λ having the form in (20). First,

$$\Lambda \succeq 0 \tag{24}$$

due to the fact that $a \geq 0$ in (20). Then

$$\Lambda \geq 0 \tag{25}$$

since $a, \lambda_1,\dots,\lambda_m \geq 0$ in (20). Since $\sum_{i=1}^m a(i) = 1, \lambda_i^T\mathbf{1} = 1, i = 1,\dots,m$,

$$\mathbf{1}^T\Lambda\mathbf{1} = 1. \tag{26}$$

As each $\lambda_i \in B_N(X)$, therefore

$$\lambda_i \odot \lambda_i = \lambda_i/N, \quad i = 1, \ldots, m, \tag{27}$$

implying

$$\operatorname{diag}(\Lambda) = \frac{1}{N}\Lambda\mathbf{1}. \tag{28}$$

Together, the constraints (24), (25), (26), and (28) give the last four constraints in SDP-Coulomb.

2.2. Duality and the Kantorovich potential. In [12], the (discrete) dual problem to (6),

$$\begin{aligned} V_{\mathrm{ee}}^{\mathrm{SCE}}(\rho) = \max_{v\in\mathbb{R}^{|X|}} \; & v^T\rho \\ \text{s.t. } & \sum_{\substack{k,l=1\\k<l}}^{N} C(i_k,i_l) - \sum_{k=1}^{N} v(i_k) \geq 0 \;\forall (i_1,\ldots,i_N), \end{aligned} \tag{29}$$

is used to solve for $V_{\mathrm{ee}}^{\mathrm{SCE}}(\rho)$. This is called the Kantorovich problem and the dual variable v is called the Kantorovich potential. Although the size of the optimization variable is reduced to $|X|$ when comparing to (6), the dual formulation has a number of constraints being exponential in N. We can also use SDP-Coulomb to provide an approximation to the Kantorovich potential. Let

$$\mathcal{A}(\Lambda\mathbf{1}) = b \to \Lambda\mathbf{1} = \rho \tag{30}$$

and $g = 0$ in the cost. We then have

$$\begin{aligned} \hat{V}_{\mathrm{ee}}^{\mathrm{SCE}}(\rho) := \min_{\Lambda\in\mathbb{R}^{|X|\times|X|}} \; & \frac{N(N-1)}{2}\operatorname{Tr}[(C - \operatorname{diag}^*(\operatorname{diag}(C)))\,\Lambda] \\ \text{s.t. } & w : \Lambda\mathbf{1} = \rho, \\ & Y : \Lambda \succeq 0, \\ & Z : \Lambda \geq 0, \\ & u : \operatorname{diag}(\Lambda) = \frac{1}{N}\Lambda\mathbf{1}, \end{aligned} \tag{31}$$

where the variables in front of the colon are the dual variables corresponding to the constraints. $\hat{V}_{\mathrm{ee}}^{\mathrm{SCE}}(\rho)$ can be seen as an approximation to $V_{\mathrm{ee}}^{\mathrm{SCE}}(\rho)$ in (29). The dual to (31) is then

$$\begin{aligned} \hat{V}_{\mathrm{ee}}^{\mathrm{SCE}}(\rho) = \max_{\substack{w\in\mathbb{R}^{|X|},\\ Y,Z\in\mathbb{R}^{|X|\times|X|}}} \; & w^T\rho \\ \text{s.t. } & \frac{N(N-1)}{2}[C - \operatorname{diag}^*(\operatorname{diag}(C))] - \frac{1}{2}(\mathbf{1}w^T + w\mathbf{1}^T) \\ & -\operatorname{diag}^*(u) + \frac{1}{2N}(\mathbf{1}u^T + u\mathbf{1}^T) = Y + Z, \\ & Y \succeq 0,\ Z \geq 0. \end{aligned} \tag{32}$$

The dual variable w can be seen as an approximation to the Kantorovich potential v in (29). As pointed out in the literatures of DFT [11, 12, 5], the Kantorovich

potential allows the functional derivative of $V_{\mathrm{ee}}^{\mathrm{SCE}}(\cdot)$ to be taken. From (32), we make the following identification:

$$\frac{dV_{\mathrm{ee}}^{\mathrm{SCE}}(\rho)}{d\rho} \approx \frac{d\hat{V}_{\mathrm{ee}}^{\mathrm{SCE}}(\rho)}{d\rho} = w^*, \tag{33}$$

where w^* is the optimizer of (32). The equality follows from the fact that for

$$g(x) = \sup_{\alpha\in\Omega} f_\alpha(x), \quad \alpha^* = \mathrm{argsup}_{\alpha\in\Omega} f_\alpha(x), \tag{34}$$

where $f_\alpha(x), \alpha \in \Omega$ are convex functions, a subgradient of $f_{\alpha^*}(x)$ is a subgradient of $g(x)$ [4]. Obtaining the approximate functional derivative of $\hat{V}_{\mathrm{ee}}^{\mathrm{SCE}}(\cdot)$ can provide a mean to optimize (8) via self-consistent field iterations (for example, in [5]), when the dependence of $E_{kd}(\cdot)$ on ρ is not analytically given.

3. Properties of SDP-Coulomb. The convex program SDP-Coulomb in section 2 intends to provide an outer approximation to the 2-marginals. In this section, we show that the extreme points of the N-representable 2-marginals are contained in the set of the extreme points of the domain of SDP-Coulomb. We first give the proof of Corollary 1.

Proof. It is clear in (16) that the left-hand side belongs to the right-hand side. Now if γ is N-representable, then

$$\gamma = \sum_{i=1}^{m} a(i)\left(\frac{N}{N-1}\lambda_i\lambda_i^T - \frac{1}{N-1}\,\mathrm{diag}^*(\lambda_i)\right),\ a \geq 0, \tag{35}$$
$$a^T\mathbf{1} = 1,\ \lambda_i \in \{0, 1/N, \ldots, N/N\}^{|X|}$$

for $a \in \mathbb{R}^m$. The constraint $\mathrm{diag}(\gamma) = 0$ gives

$$\sum_{i=1}^{m} a(i)(N\lambda_i \odot \lambda_i - \lambda_i) = 0, \tag{36}$$

where $\odot$ denotes pointwise product. Due to the domain of λ_i, $N\lambda_i \odot \lambda_i - \lambda_i \geq 0$. Then, together with $a(i) \geq 0$, (36) implies $a(i) = 0$ or $N\lambda_i \odot \lambda_i = \lambda_i$ for each i. This shows that $\lambda_i \in \{0, 1/N\}^{|X|}$, implying in (16) the right-hand side belongs to the left-hand side. Finally, it is clear that $\tilde{\Gamma}_2$ is the set of extreme points of $\mathrm{conv}(\tilde{\Gamma}_2)$, since $\tilde{\Gamma}_2$ is a subset of the extreme points $\mathrm{conv}(\Gamma_2)$ and $\mathrm{conv}(\tilde{\Gamma}_2) \subseteq \mathrm{conv}(\Gamma_2)$. ∎

In the following theorem, we show that $\tilde{\Gamma}_2$ also belongs to the set of the extreme points for the feasible set of γ used in problem SDP-Coulomb in (19), when the constraint $\mathcal{A}(\Lambda\mathbf{1}) = b$ is absent. This shows that our convex relaxation is rather tight.

THEOREM 2. *$\tilde{\Gamma}_2$ is a subset of the extreme points of the domain*

$$D = \left\{ \frac{N}{N-1}\Lambda - \frac{1}{N-1}\,\mathrm{diag}^*(\Lambda\mathbf{1}) \;\middle|\; \Lambda \succeq 0,\ \Lambda \geq 0,\ \mathbf{1}^T\Lambda\mathbf{1} = 1,\ \mathrm{diag}(\Lambda) = \frac{1}{N}\Lambda\mathbf{1} \right\}, \tag{37}$$

which is the feasible set of γ in (19) when the constraint $\mathcal{A}(\Lambda\mathbf{1}) = b$ is absent.

Proof. First, $\tilde{\Gamma}_2$ is a subset of D. We further need to show that each

$$\gamma_{\mathrm{ext}} = \frac{N}{N-1}\lambda_{\mathrm{ext}}\lambda_{\mathrm{ext}}^T - \frac{1}{N-1}\,\mathrm{diag}^*(\lambda_{\mathrm{ext}}), \quad \lambda_{\mathrm{ext}} \in \mathcal{B}_N(X) \tag{38}$$

in $\tilde{\Gamma}_2$ is also an extreme point in D. To this end, we simply show for every γ_{ext}, there exists some cost B such that the unique maximizer to

$$\max_{\gamma} \text{Tr}(B\gamma), \quad \text{s.t. } \gamma \in D \tag{39}$$

is γ_{ext}. If γ_{ext} is the unique maximizer to (39), then $\gamma_{\text{ext}} \neq \sum_i a(i)\gamma_i$, where $\forall i\ \gamma_i \in D, a(i) > 0$, and $\sum_i a(i) = 1$. Otherwise, $\text{Tr}(B\gamma_{\text{ext}}) = \sum_i a(i)\text{Tr}(B\gamma_i) < \sum_i a(i)\text{Tr}(B\gamma_{\text{ext}}) = \text{Tr}(B\gamma_{\text{ext}})$, where the inequality is due to the fact that γ_{ext} uniquely minimizes $\text{Tr}(B\gamma)$. Let

$$B := \lambda_{\text{ext}}\lambda_{\text{ext}}^T + \frac{1}{N-1}\mathbf{1}\,\text{diag}(\lambda_{\text{ext}}\lambda_{\text{ext}}^T)^T. \tag{40}$$

Then

$$\begin{aligned}\text{Tr}(B\gamma) &= \text{Tr}(\lambda_{\text{ext}}\lambda_{\text{ext}}^T\gamma) + \text{Tr}\left(\frac{1}{N-1}\mathbf{1}\,\text{diag}(\lambda_{\text{ext}}\lambda_{\text{ext}}^T)^T\gamma\right)\\ &= \text{Tr}(\lambda_{\text{ext}}\lambda_{\text{ext}}^T\gamma) + \text{Tr}\left(\frac{1}{N-1}\lambda_{\text{ext}}\lambda_{\text{ext}}^T\,\text{diag}^*(\gamma\mathbf{1})\right)\\ &= \text{Tr}\left(\lambda_{\text{ext}}\lambda_{\text{ext}}^T\left(\gamma + \frac{1}{N-1}\,\text{diag}^*(\gamma\mathbf{1})\right)\right). \end{aligned} \tag{41}$$

Plugging in $\gamma = \frac{N}{N-1}\Lambda - \frac{1}{N-1}\,\text{diag}^*(\Lambda\mathbf{1}) \in D$, (39) is therefore

$$\begin{aligned}&\min_{\gamma,\Lambda}\ \frac{N}{N-1}\text{Tr}(\lambda_{\text{ext}}\lambda_{\text{ext}}^T\Lambda)\\ &\text{s.t. } \gamma = \frac{N}{N-1}\Lambda - \frac{1}{N-1}\,\text{diag}^*(\Lambda\mathbf{1}),\\ &\qquad \Lambda \succeq 0,\ \Lambda \geq 0,\ \mathbf{1}^T\Lambda\mathbf{1} = 1,\ \text{diag}(\Lambda) = \frac{1}{N}\Lambda\mathbf{1}.\end{aligned} \tag{42}$$

To show that γ_{ext} in (38) is the unique minimizer of (42), it suffices to show that γ_{ext} is the unique minimizer for

$$\begin{aligned}&\min_{\gamma,\Lambda}\ \frac{N}{N-1}\text{Tr}(\lambda_{\text{ext}}\lambda_{\text{ext}}^T\Lambda)\\ &\text{s.t. } \gamma = \frac{N}{N-1}\Lambda - \frac{1}{N-1}\,\text{diag}^*(\Lambda\mathbf{1}),\ \Lambda \succeq 0,\ \text{Tr}(\Lambda) = 1/N,\end{aligned} \tag{43}$$

since the domain of (42) is contained within (43). It is clear that the unique minimizer to (43) is $\Lambda = \lambda_{\text{ext}}\lambda_{\text{ext}}^T$, implying that γ_{ext} is the unique minimizer. ∎

4. Tightening the convex relaxation. Though Theorem 2 shows that our convex relaxation with the 2-marginals contains $\tilde{\Gamma}_2$ as the extreme points (hence the relaxed domain also contains $\text{conv}(\tilde{\Gamma}_2)$), the relaxation may include points that are out of $\text{conv}(\tilde{\Gamma}_2)$. To further restrict the domain of optimization in SDP-Coulomb, one can consider applying convex relaxation to the k-marginals. In this section, we focus on the case of the 3-marginals. Let

$$\bar{C}(i,j,k) = C(i,j) + C(j,k) + C(k,i), \quad i,j,k = 1,\ldots,|X|. \tag{44}$$

Let the N-representable 3-marginal of μ be κ. In terms of $\bar{C}$ and κ, the cost of (3) becomes

$$g(\lambda) + \frac{N(N-1)(N-2)}{6}\sum_{i,j,k=1}^{|X|}\bar{C}(i,j,k)\kappa(i,j,k). \tag{45}$$

In the following sections, we work out the domain of κ in order to perform minimization. We follow the derivation in [7] in which the set Γ_2 is derived.

4.1. The extreme points of the symmetric discrete distribution on X^N. Let the set of symmetric discrete N-marginal be defined as

$$\Pi_{N,\mathrm{sym}} = \left\{ \mu \in (\mathbb{R}^{|X|})^N \mid \mu \text{ is symmetric}, \mu \geq 0, \sum_{i_1,\ldots,i_N=1}^{|X|} \mu(i_1,\ldots,i_N) = 1 \right\}. \tag{46}$$

Let $e_l \in \mathbb{R}^{|X|}$ be defined as $e_l(j) = \delta_{lj}$. For the set of probability measures on X^N, an extreme point is

$$e_{c_1} \otimes \ldots \otimes e_{c_N} \tag{47}$$

for some $c_1,\ldots,c_N \in \{1,\ldots,|X|\}$. Therefore, for the set of symmetric measure $\Pi_{N,\mathrm{sym}}$, an extreme point can be obtained from symmetrizing (47), giving rise to the set

$$\Gamma_N = \left\{ \frac{1}{N!} \sum_{\sigma\in S(N)} e_{c_{\sigma(1)}} \otimes \ldots \otimes e_{c_{\sigma(N)}} \mid c_1,\ldots,c_N \in \{1,\ldots,|X|\} \right\}, \tag{48}$$

where $S(N)$ is the symmetric group over N numbers. For physical measure of the electrons, we look at a restricted set

$$\tilde{\Pi}_{N,\mathrm{sym}} = \{\mu \in \Pi_{N,\mathrm{sym}} \mid \mu(i_1,\ldots,i_N) = 0 \text{ if } i_k = i_l \ \forall k,l = 1,\ldots,N\} \tag{49}$$

which ensures two electrons cannot be in the same state. A derivation similar to Corollary 1 reveals that

$$\mathrm{conv}(\tilde{\Gamma}_N) = \tilde{\Pi}_{N,\mathrm{sym}}, \tag{50}$$

where

$$\tilde{\Gamma}_N = \left\{ \frac{1}{N!} \sum_{\sigma\in S(N)} e_{c_{\sigma(1)}} \otimes \cdots \otimes e_{c_{\sigma(N)}} \mid c_1,\ldots,c_N \in \{1,\ldots,|X|\}, \right. \tag{51}$$
$$\left. c_i \neq c_j \ \forall\ i,j \in N,\ i \neq j \right\}.$$

4.2. Convex hull of the set of N-representable 3-marginals. To get a description to the set of N-representable 3-marginals in order to restrict κ in (45), we marginalize the measures in $\tilde{\Pi}_{N,\mathrm{sym}}$. Since $\tilde{\Pi}_{N,\mathrm{sym}} = \mathrm{conv}(\tilde{\Gamma}_N)$, it suffices to marginalize the elements in $\tilde{\Gamma}_N$. Picking an arbitrary element in $\tilde{\Gamma}_N$, then its 3-marginal is

$$\begin{aligned} &\frac{1}{N!} \sum_{\sigma\in S(N)} \sum_{l_4,\ldots,l_N=1}^{|X|} e_{c_{\sigma(1)}} \otimes \cdots \otimes e_{c_{\sigma(N-1)}}(l_{N-1}) \otimes e_{c_{\sigma(N)}}(l_N) \\ = &\frac{1}{N!} \sum_{\sigma\in S(N)} e_{c_{\sigma(1)}} \otimes e_{c_{\sigma(2)}} \otimes e_{c_{\sigma(3)}} \end{aligned}$$

$$
\begin{aligned}
&= \frac{(N-3)!}{N!} \sum_{\substack{i,j,k=1 \\ i>j>k}}^{N} e_{c_i} \otimes e_{c_j} \otimes e_{c_k} \\
&= \frac{1}{N(N-1)(N-2)} \Bigg(\sum_{i,j,k=1}^{N} e_{c_i} \otimes e_{c_j} \otimes e_{c_k} + 2 \sum_{k=1}^{N} e_{c_k} \otimes e_{c_k} \otimes e_{c_k} \\
&\quad - \sum_{i,j=1}^{N} e_{c_i} \otimes e_{c_j} \otimes e_{c_j} - \sum_{i,j=1}^{N} e_{c_j} \otimes e_{c_i} \otimes e_{c_j} - \sum_{i,j=1}^{N} e_{c_i} \otimes e_{c_i} \otimes e_{c_j} \Bigg). \qquad (52)
\end{aligned}
$$

The second equality follows from the fact that there are $(N-3)!$ $\sigma \in S(N)$ such that $e_{c_i} \otimes e_{c_j} \otimes e_{c_k} = e_{c_{\sigma(1)}} \otimes e_{c_{\sigma(2)}} \otimes e_{c_{\sigma(3)}}$ for a fixed $e_{c_i} \otimes e_{c_j} \otimes e_{c_k}$. Letting

$$
\lambda := \frac{1}{N} \sum_{i=1}^{N} e_{c_i}, \qquad (53)
$$

it follows that $\lambda \in \{0, 1/N\}^{|X|}$, and $\lambda^T \mathbf{1} = 1$, since each e_{c_i} has only an entry with value 1 and is 0 everywhere else, and $c_i \neq c_j$ for all $i \neq j$. Moreover,

$$
\frac{1}{N} \sum_{i=1}^{N} e_{c_i}(l) e_{c_i}(l) e_{c_i}(l) = \frac{1}{N} \sum_{i=1}^{N} e_{c_i}(l) e_{c_i}(l) = \frac{1}{N} \sum_{i=1}^{N} e_{c_i}(l) = \lambda(l), \quad l = 1, \dots, |X|, \qquad (54)
$$

and

$$
\frac{1}{N} \sum_{i=1}^{N} e_{c_i}(l) e_{c_i}(j) = 0, \quad \frac{1}{N} \sum_{i=1}^{N} e_{c_i}(l) e_{c_i}(j) e_{c_i}(k) = 0 \quad \text{if } l \neq j, \text{or } j \neq k, \text{ or } k \neq l. \qquad (55)
$$

Writing (52) in terms of λ using (54) and (55), one can marginalize $\tilde{\Gamma}_N$ to obtain

$$
\begin{aligned}
\tilde{\Gamma}_3 = \Bigg\{ & \frac{1}{N(N-1)(N-2)} \big(N^3 \lambda \otimes \lambda \otimes \lambda + 2N \sum_{l=1}^{|X|} \lambda(l) e_l \otimes e_l \otimes e_l \\
& - N^2 \sum_{l=1}^{|X|} \lambda(l) \lambda \otimes e_l \otimes e_l \\
& - N^2 \sum_{l=1}^{|X|} \lambda(l) e_l \otimes \lambda \otimes e_l - N^2 \sum_{l=1}^{|X|} \lambda(l) e_l \otimes e_l \otimes \lambda \big) \mid \lambda \in \mathcal{B}_N(X) \Bigg\}. \qquad (56)
\end{aligned}
$$

Since every physical N-representable 3-marginal comes from the marginalization of an element in $\tilde{\Pi}_{N,\mathrm{sym}} = \mathrm{conv}(\tilde{\Gamma}_N)$, the following statement holds.

PROPOSITION 1. *The set of N-representable 3-marginals coming from the marginalization of $\tilde{\Pi}_{N,sym}$ is* $\mathrm{conv}(\tilde{\Gamma}_3)$.

With this proposition, in order to minimize (45) one can solve

$$
\begin{aligned}
\min_{\substack{\kappa \in \mathbb{R}^{|X| \times |X| \times |X| \times |X|}, \\ \lambda \in \mathbb{R}^{|X|}}} & \; g(\lambda) + \frac{N(N-1)(N-2)}{6} \sum_{i,j,k=1}^{|X|} \bar{C}(i,j,k) \kappa(i,j,k) \qquad (57) \\
& \text{s.t. } \kappa \in \mathrm{conv}(\tilde{\Gamma}_3),
\end{aligned}
$$

$$\lambda(i) = \sum_{j,k=1}^{|X|} \kappa(i,j,k), \quad i = 1, \ldots, |X|,$$
$$\mathcal{A}(\lambda) = b.$$

4.3. Convex relaxation to the 3-marginal problem. The variable κ in (57) takes the form $\kappa = \sum_{i=1}^{m} a(i)\kappa_i, \kappa_i \in \tilde{\Gamma}_3$ with $a \geq 0$ and $a^T\mathbf{1} = 1$. Therefore, in order to derive a convex relaxation to (57), one seeks a convex set that contains all the elements in $\tilde{\Gamma}_3$. Such a set will certainly contain $\kappa = \sum_{i=1}^{m} a(i)\kappa_i$, which is a convex combination of $\kappa_i \in \tilde{\Gamma}_3, i = 1, \ldots, m$. For this purpose, let

$$\Theta := \lambda \otimes \lambda \otimes \lambda, \quad \lambda \in \mathcal{B}_N(X). \tag{58}$$

Since $\lambda^T\mathbf{1} = 1$,

$$\lambda\lambda^T = \sum_{k=1}^{|X|} \Theta(:,:,k), \quad \lambda = \sum_{j,k=1}^{|X|} \Theta(:,j,k). \tag{59}$$

Then in terms of Θ, an extreme point $\kappa \in \tilde{\Gamma}_3$ is

$$\begin{aligned}\kappa = \phi(\Theta) := \frac{1}{N(N-1)(N-2)} \Bigg(& N^3\Theta + 2N \sum_{l=1}^{|X|} \left(\sum_{j,k=1}^{|X|} \Theta(l,j,k) \right) e_l \otimes e_l \otimes e_l \\ & - N^2 \sum_{l=1}^{|X|} \left(\sum_{k=1}^{|X|} \Theta(l,:,k) \right) \otimes e_l \otimes e_l - N^2 \sum_{l=1}^{|X|} e_l \otimes \left(\sum_{k=1}^{|X|} \Theta(l,:,k) \right) \otimes e_l \\ & - N^2 \sum_{l=1}^{|X|} e_l \otimes e_l \otimes \left(\sum_{k=1}^{|X|} \Theta(l,:,k) \right) \Bigg).\end{aligned} \tag{60}$$

Next, we impose some necessary conditions on Θ in a convex manner so that Θ comes from the tensor product of the quantized marginals λ. Clearly, the symmetry property implies

$$\Theta(i,j,k) = \Theta(k,i,j) = \Theta(j,k,i) = \Theta(j,i,k) = \Theta(k,j,i) = \Theta(i,k,j). \tag{61}$$

Since $\lambda \in \{0, 1/N\}^{|X|}$,

$$\lambda(i)\lambda(i)\lambda(j) = \lambda(i)\lambda(j)/N \Rightarrow \Theta(i,i,j) = \frac{1}{N} \sum_{k=1}^{|X|} \Theta(i,j,k) \quad \forall i,j = 1, \ldots, |X|. \tag{62}$$

Then the constraint that $\lambda^T\mathbf{1} = 1$ gives

$$\sum_{i=1}^{|X|} \lambda(i) = 1 \Rightarrow \sum_{i,j,k=1}^{|X|} \Theta(i,j,k) = 1. \tag{63}$$

We also have the conic constraints

$$\Theta(:,:,i) = \lambda\lambda^T\lambda(i) \succeq 0 \quad \forall i = 1, \ldots, |X| \tag{64}$$

and

$$\Theta \geq 0. \tag{65}$$

Combining (61), (62), (63), (64), and (65) leads to the following optimization problem over Θ:

$$
\begin{aligned}
(66)\quad \min_{\Theta,\kappa\in\mathbb{R}^{|X|\times|X|\times|X|}} \quad & g\left(\sum_{j,k=1}^{|X|}\Theta(:,j,k)\right) \\
& + \frac{N(N-1)(N-2)}{6}\sum_{i,j,k=1}^{|X|}\bar{C}(i,j,k)\kappa(i,j,k) \quad \text{(SDP-Coulomb2)} \\
\text{s.t.} \quad & \kappa = \phi(\Theta) \\
& \Theta \text{ is symmetric} \\
& \Theta(i,i,j) = \frac{1}{N}\sum_{k=1}^{|X|}\Theta(i,j,k) \quad \forall i,j = 1,\dots,|X|, \\
& \sum_{i,j,k=1}^{|X|}\Theta(i,j,k) = 1, \\
& \Theta(:,:,i) \succeq 0 \ \forall i = 1,\dots,|X|, \Theta \geq 0, \\
& \mathcal{A}\left(\sum_{j,k=1}^{|X|}\Theta(:,j,k)\right) = b.
\end{aligned}
$$

4.4. A remark on Lassere's hierarchy. It is possible to use the Lassere hierarchy (or sum-of-squares hierarchy) [1, 3] to further tighten the convex relaxation. When applying this method to our problem, the task of determining some power of the quantized 1-marginal $\lambda \in \mathcal{B}_N(X)$ (for example, the problem of determining the 2- and 3-marginals), is reformulated as a moment determination problem. More precisely, instead of working with the monomials $\{\lambda^\alpha\}_\alpha$ where $\alpha \in \mathbb{N}^{|X|}$ is a multi-index and $\mathbb{N}$ is the set of natural numbers, one performs a change of variables according to

$$
[\lambda^\alpha\lambda^\beta]_{\alpha,\beta} \Rightarrow [\mathbb{E}(\lambda^\alpha\lambda^\beta)]_{\alpha,\beta}. \tag{67}
$$

The optimization variable, the matrix $[\mathbb{E}(\lambda^\alpha\lambda^\beta)]_{\alpha,\beta}$, has size $\binom{p+|X|}{p}$ for each dimension if we consider the monomials λ^α's and λ^β's up to degree p. Then, an equality constraint $h(\lambda) = 0$ (h is a polynomial) is changed according to

$$
h(\lambda) = 0 \Rightarrow \mathbb{E}(h(\lambda)\lambda^\alpha) = 0 \ \forall\alpha, \tag{68}
$$

and an inequality constraint $q(\lambda) \geq 0$ (q is a polynomial) is changed according to

$$
q(\lambda) \geq 0 \Rightarrow \mathbb{E}(q(\lambda)s(\lambda)^2) \geq 0 \ \forall s(\lambda), \tag{69}
$$

where s is some polynomial. The inequality constraints lead to a positive semidefinite constraint. For example the constraint $\lambda \geq 0$ simply gives

$$
v^T\left([\mathbb{E}(\lambda^\alpha\lambda^\beta)]_{\alpha,\beta}\right)v \geq 0 \quad \forall v \text{ with size } \binom{p+|X|}{p} \tag{70}
$$

if we consider the monomials λ^α's and λ^β's up to degree p. As can be seen, when choosing $p \geq 2$, we are already faced with $|X|^4$ variables. Therefore, we pursue a cheaper alternative.

5. Rounding. The previous sections describe several convex relaxation approaches for solving the multimarginal transport problem. The general philosophy is to enlarge the domain of optimization, therefore obtaining a lower bound for the global minimum. To obtain an upper bound for the global minimum, we need to project the solution back into the unrelaxed domain ($\text{conv}(\tilde{\Gamma}_2)$ or $\text{conv}(\tilde{\Gamma}_3)$). We consider two cases of practical importance:

1. When the linear constraint $\mathcal{A}(\lambda) = b$ is not present in (3).
2. When $\mathcal{A}(\lambda) = b \to \lambda = \rho$, for example when solving the multimarginal-optimal transport problem (6).

Section 5.1 addresses the first case. Here, we devise a scheme to round the solution from SDP-Coulomb to the set of extreme points $\tilde{\Gamma}_2$ for the set of N-representable 2-marginals. In section 5.2, we deal with the second case with the marginal constraint. For this case, it is difficult to work with SDP-Coulomb to obtain a rounded solution in $\tilde{\Gamma}_2$. Therefore, we discuss how we can use SDP-Coulomb2 for such a purpose.

5.1. Without the linear constraint $\mathcal{A}(\lambda) = b$. In the special case where the constraint $\mathcal{A}(\Lambda\mathbf{1}) = b$ is absent and $g(\cdot)$ is a linear functional, we simply minimize a linear functional of Λ in SDP-Coulomb. In principle, if the domain of SDP-Coulomb (without $\mathcal{A}(\Lambda\mathbf{1}) = b$) is close to the set of N-representable 2-marginals with zero diagonal ($\text{conv}(\tilde{\Gamma}_2)$ in Corollary 1), then SDP-Coulomb should return a solution $\Lambda^* \approx \lambda^* {\lambda^*}^T$ where $\lambda^* \in \mathcal{B}_N(X)$. This is because the extreme points of $\text{conv}(\tilde{\Gamma}_2)$ is $\tilde{\Gamma}_2$ (Corollary 1), and generically, the optimizer of a linear functional over a convex set is an extreme point of the set. We therefore propose a rounding procedure in Algorithm 1. If SDP-Coulomb returns a solution Λ^* where the entries on the diagonal of Λ^* are not exactly $1/N^2$ or 0, letting the index of the largest entry of $\text{diag}(\Lambda^*)$ be $i_{\max}$, we add a linear constraint $\text{diag}(\Lambda)(i_{\max}) = 1/N^2$ to SDP-Coulomb. This step is repeated until a rank-1 Λ^* is obtained. This is summarized in Algorithm 1.

Algorithm 1 Rounding in the absence of the linear constraint $\mathcal{A}(\lambda) = b$.

procedure ROUNDING
  $\Lambda^* \leftarrow$ Solution to SDP-Coulomb.
  $\mathcal{I} \leftarrow \{\emptyset\}$, $R \leftarrow I$
  while $\text{rank}(\Lambda^*) > 1$ **do**
    $i_{\max} \leftarrow$ index of the largest element in $R\,\text{diag}(\Lambda^*)$.
    $\mathcal{I} \leftarrow \mathcal{I} \cup i_{\max}$, $\mathcal{I}^c \leftarrow \{1, \ldots, |X|\} \setminus \mathcal{I}$.
    $R \leftarrow I(\mathcal{I}^c, :)$.
    $\Lambda^* \leftarrow$ Solution to SDP-Coulomb with the extra constraint $\text{diag}(\Lambda)_{\mathcal{I}} = 1/N^2$.
  end while
  return Λ^*.
end procedure

We remark that this procedure is crucial when there are degenerate solutions, giving a high rank solution in SDP-Coulomb.

5.2. With the marginal constraint $\lambda = \rho$. When having the constraint $\Lambda\mathbf{1} = \rho$ in SDP-Coulomb, we cannot pursue the same strategy as in section 5.1 to round the solution. When there exists a marginal constraint, we expect the solution to (12) to be a convex combination of the extreme points from $\tilde{\Gamma}_2$, implying SDP-Coulomb returns solution as $\Lambda^* \approx \sum_{i=1}^m a^*(i)\lambda_i^* {\lambda_i^*}^T, {a^*}^T\mathbf{1} = 1, a^* \geq 0$. However, in

order to round, one has to first disentangle each λ_i^* from such a convex combination. Since λ_i^*'s are not orthogonal to each other, it is not obvious how one can use matrix factorization techniques such as an eigendecomposition to obtain the λ_i^*'s from Λ^*. To this end, we resort to using SDP-Coulomb2 to obtain each λ_i^*. Since in SDP-Coulomb2 we expect to have the solution $\Theta^* \approx \sum_{i=1}^m a^*(i)\lambda_i^* \otimes \lambda_i^* \otimes \lambda_i^*, \lambda_i^* \in \mathcal{B}_N(X)$ (as we expect the solution to approximately lie in $\mathrm{conv}(\tilde{\Gamma}_3)$), we resort to using a CP-tensor decomposition [9] to obtain each individual λ_i^* approximately.

In order to use a CP-decomposition, one needs to have an idea of what m is. The following discussion demonstrates that $m = |X|$. We first look at the set of the physical symmetric probability measures on X^N that have the marginal being ρ:

$$\begin{aligned}\tilde{\Pi}_{N,\mathrm{sym}}(\rho) &= \left\{ \mu \in \tilde{\Pi}_{N,\mathrm{sym}} \;\middle|\; \sum_{i_2,\ldots,i_N=1}^{|X|} \mu(:, i_2, \ldots, i_N) = \rho \right\} \\ &= \mathrm{conv}(\tilde{\Gamma}_N) \cap \left\{ \mu \in (\mathbb{R}^{|X|})^N \mid \sum_{i_2,\ldots,i_N=1}^{|X|} \mu(i_1, i_2, \ldots, i_N) = \rho(i_1), \right. \\ &\qquad \left. i_1 = 1, \ldots, |X| - 1 \right\}. \end{aligned} \tag{71}$$

Notice that the marginal constraint in (71) is only enforced for $|X| - 1$ sites. This is because for $\mu \in \mathrm{conv}(\tilde{\Gamma}_N)$,

$$\sum_{i_2,\ldots,i_N=1}^{|X|} \mu(|X|, i_2, \ldots, i_N) \tag{72}$$

is completely determined by

$$\sum_{i_2,\ldots,i_N=1}^{|X|} \mu(i_1, i_2, \ldots, i_N), \quad i_1 = 1, \ldots, |X| - 1, \tag{73}$$

via

$$\sum_{i_2,\ldots,i_N=1}^{|X|} \mu(|X|, i_2 \ldots, i_N) = 1 - \sum_{i_1=1}^{|X|-1} \sum_{i_2,\ldots,i_N=1}^{|X|} \mu(i_1, i_2 \ldots, i_N). \tag{74}$$

We now appeal to the results in [6] to see what m is. The theorem in [6] implies that for a closed and bounded convex set $\mathcal{K}$, an extreme point of $\mathcal{K} \cap H_1 \cap \cdots H_n$, where $H_1, \ldots, H_n$ are n hyperplanes can be represented as $n+1$ convex combination of the extreme points of $\mathcal{K}$. Since $\tilde{\Pi}_{N,\mathrm{sym}}(\rho)$ in (71) is the intersection of $\mathrm{conv}(\tilde{\Gamma}_N)$ with $|X| - 1$ hyperplanes, it follows that for an extreme point $\mu \in \tilde{\Pi}_{N,\mathrm{sym}}(\rho)$, μ is the convex combination of $|X|$ elements in $\tilde{\Gamma}_N$. After a marginalization, it follows that a physical N-representable 3-marginal that satisfies the marginal constraint is a convex combination of $|X|$ elements of $\tilde{\Gamma}_3$, therefore $m = |X|$.

As $\Theta^* \approx \sum_{i=1}^{|X|} a^*(i)\lambda_i^* \otimes \lambda_i^* \otimes \lambda_i^*$, if the approximation $\approx$ holds with an $=$ sign, and if $\lambda_1^*, \ldots, \lambda_{|X|}^*$ are linearly independent, then Θ^* has a unique CP tensor decomposition, up to ordering and magnitude of λ_i^*'s. This can be seen in section 5.2.1 where Jenrich's algorithm provides an explicit construction of the λ_i^*'s. We note that

although the assumption of linearly independent $\lambda_1^*, \dots, \lambda_{|X|}^*$ is required for the success of Jenrich's algorithm, it is not a necessary condition to ensure the uniqueness of the CP-decomposition (see, for example, the theorem of Kruskal [10]). In the situation where the linearly independence assumption is violated, one may use a different algorithm such as the alternating least-squares (ALS) for recovering the tensor components. Therefore, our rounding algorithm has three phases. We first use Jenrich's algorithm to obtain an initialization for $\lambda_i^*, i = 1, \dots, |X|$. Then a procedure based on ALS is used to refine the solution from Jenrich's algorithm and also enlarge the set $\{\lambda_i^*\}_{i=1}^{|X|}$ to $\{\lambda_i^*\}_{i=1}^{p}$. Last, we solve a regression problem to determine the convex combination of $\{\lambda_i^*\}_{i=1}^{p}$ that approximate Θ^* while satisfying the marginal constraint. The algorithm is summarized in Algorithm 2.

Algorithm 2 Algorithm for rounding in the presence of the marginal constraint.

procedure ROUNDING2(δ, ρ)
  $\Theta^* \leftarrow$ Solution to SDP-Coulomb2.
  $\{\lambda_i^*\}_{i=1}^{|X|} \leftarrow$ JENRICH(Θ^*) (section 5.2.1).
  $\{\lambda_i^*\}_{i=1}^{p} \leftarrow$ ALS($\Theta^*, \{\lambda_i^*\}_{i=1}^{|X|}, \delta$) (section 5.2.2).
  $a^* \leftarrow \underset{a \in \mathbb{R}^p}{\operatorname{argmin}} \|\Theta^* - \sum_{i=1}^{p} a(i)\lambda_i^* \otimes \lambda_i^* \otimes \lambda_i^*\|_F^2$ s.t. $a \geq 0, a^T \mathbf{1} = 1, \sum_{i=1}^{p} a(i)\lambda_i^* = \rho$.
  $\Theta^* \leftarrow \sum_{i=1}^{p} a^*(i)\lambda_i^* \otimes \lambda_i^* \otimes \lambda_i^*$,
  return Θ^*
end procedure

Algorithm 3 Jenrich's algorithm.

procedure JENRICH(Θ)
  Get $w_1, w_2 \in \mathbb{R}^{|X|}$, $w_1(i), w_2(i) \sim \text{uniform}[0,1], i = 1, \dots, |X|$.
  $W_1 \leftarrow \sum_{k=1}^{|X|} w_1(k)\Theta(:,:,k)$, $W_2 \leftarrow \sum_{k=1}^{|X|} w_2(k)\Theta(:,:,k)$.
  Eigendecompose $W_1 W_2^\dagger = U \Sigma U^\dagger$, where Σ is a diagonal matrix.
  $\lambda_i \leftarrow U(:,i), i = 1, \dots, |X|$.
  $\lambda_i \leftarrow \frac{\lambda_i}{\sqrt{N}\|\lambda_i\|_2}, i = 1, \dots, |X|$.
  return $\{\lambda\}_{i=1}^{|X|}$.
end procedure

5.2.1. Jenrich's algorithm. In this section, we provide the details for Jenrich's algorithm in Algorithm 3 for the sake of completeness. The key idea of Algorithm 3 is that if $\Theta = \sum_{i=1}^{|X|} a(i)\lambda_i \otimes \lambda_i \otimes \lambda_i$, then

$$W_1 = \sum_{i=1}^{|X|} (a(i) w_1^T \lambda_i)\lambda_i \lambda_i^T, \quad W_2 = \sum_{i=1}^{|X|} (a(i) w_2^T \lambda_i)\lambda_i \lambda_i^T. \tag{75}$$

Thus

$$W_1 W_2^\dagger = U S U^\dagger, \quad U = [\lambda_1 \cdots \lambda_{|X|}], \quad \Sigma = \text{diag}^*\left(\left[\frac{a(1)w_1^T\lambda_1}{a(1)w_2^T\lambda_1}, \dots, \frac{a(|X|)w_1^T\lambda_{|X|}}{a(|X|)w_2^T\lambda_{|X|}}\right]\right). \tag{76}$$

So the eigenvectors of $W_1 W_2^\dagger$ give $\lambda_1, \dots, \lambda_{|X|}$. The last step in Algorithm 3 is a normalization step to ensure $\|\lambda_i\| = 1/\sqrt{N}$ for all i, since in principle $\lambda_i \in \mathcal{B}_N(X)$.

As we see, if in (76) $\lambda_1, \ldots, \lambda_{|X|}$ are linearly independent, Jenrich's algorithm gives a unique decomposition since $\mathrm{diag}(\Sigma)$ is nondegenerate generically (except for the entries correspond to $a(i) = 0$).

5.2.2. Alternating least-squares. To further refine the solution from Jenrich's algorithm to approximate a given tensor Θ, we propose using a variant of the ALS that is similar to a projected gradient descent. Ideally, if $\Theta = \sum_{i=1}^{|X|} a(i)\lambda_i \otimes \lambda_i \otimes \lambda_i$, one can try to solve

$$\begin{aligned} \min_{\substack{a\in\mathbb{R}^{|X|},\\ P,Q,R\in\mathbb{R}^{|X|\times|X|}}} & \left\| \sum_{i=1}^{|X|} P(:,i)\otimes Q(:,i)\otimes R(:,i) - \Theta \right\|_F^2 \\ \text{s.t. } & Q = R, P = R\,\mathrm{diag}^*(a), \\ & a \geq 0,\ a^T\mathbf{1} = 1, \\ & R(:,i) \in \mathcal{B}_N(X), \end{aligned} \tag{77}$$

using a local optimization algorithm and identify the λ_i's with the $R(:,i)$'s, provided Jenrich's algorithm gives a good initialization. There is, however, a caveat. Although $\sum_{i=1}^{|X|} P(:,i)\otimes Q(:,i)\otimes R(:,i)$ provides an approximation to the 3-marginal Θ, $\sum_{k,j=1}^{|X|}\sum_{i=1}^{|X|} P(:,i)\otimes Q(k,i)\otimes R(j,i) \neq \rho$ in general, hence the marginal constraint can be violated. To deal with such an issue, we want to identify a set of λ_i's in $\mathcal{B}_N(X)$, $\{\lambda_i\}_{i=1}^p$, where $p > |X|$. With a more generous selection of the λ_i's, some convex combination of $\{\lambda_i\}_{i=1}^p$ should give the correct marginal while approximating Θ from SDP-Coulomb2 (66).

To this end, the following problem with a less stringent constraint is solved instead:

$$\begin{aligned} \min_{P,Q,R\in\mathbb{R}^{|X|\times|X|}} & \left\| \sum_{i=1}^{|X|} P(:,i)\otimes Q(:,i)\otimes R(:,i) - \Theta \right\|_F^2 \\ \text{s.t. } & \|Q(:,i)\|_2 = 1/\sqrt{N}, \\ & N \text{ entries of } |R(:,i)| \text{ are } 1/N\ i = 1,\ldots,|X|. \end{aligned} \tag{78}$$

Notice that each of the $R(:,i)$'s is not required to have only N nonzero entries, unlike in (77) where $R(:,i)$'s belong to $\mathcal{B}_N(X)$. To solve (78), we use an ALS procedure detailed in Algorithm 4. The outer-loop of this procedure controls the number of the entries of $R(:,i)$ that have magnitude $1/N$. At every step of Algorithm 4, each column of Q is normalized to $1/\sqrt{N}$ after solving the least-squares concerning Q. To enforce the constraint on $R(:,i)$ in (78), after solving the least-squares concerning R, for each $R(:,i)$, k entries with the largest magnitude are picked out and have their magnitude being set to $1/N$. When the iteration converges, we then enforce $k+1$ entries of each $R(:,i), i = 1,\ldots,|X|$ to have magnitude $1/N$ in the ALS. These steps are repeated until $k = N$. We expect each $R(:,i), i = 1,\ldots,|X|$ to have N or slightly greater than N entries that are large in magnitude. Using the large magnitude entries in each column of R, we exhaustively enumerate the candidate $\{\lambda_i\}_{i=1}^p$ where $\lambda_i \in \mathcal{B}_N(X)$. The number p is controlled via the parameter δ.

6. Numerical simulations. In this section, we demonstrate the effectiveness of our approach using a few numerical examples. The energy is computed using

$$E(\gamma) = \sum_{i,j=1}^{|X|} \mathrm{Tr}(C(i,j)\gamma(i,j)), \tag{79}$$

Algorithm 4 Modified ALS.

procedure ALS($\Theta, \{\lambda_i\}_{i=1}^{|X|}, \delta$)
  Initialize $Q = [\lambda_1, \ldots, \lambda_{|X|}]$, $R = [\lambda_1, \ldots, \lambda_{|X|}]$.
  for k from 1 to N **do**
    while not converge **do**
      $P \leftarrow \arg\min_{\tilde{P} \in \mathbb{R}^{|X|\times|X|\times|X|}} \| \sum_{i=1}^{|X|} \tilde{P}(:,i) \otimes Q(:,i) \otimes R(:,i) - \Theta \|_F^2$.
      $Q \leftarrow \arg\min_{\tilde{Q} \in \mathbb{R}^{|X|\times|X|\times|X|}} \| \sum_{i=1}^{|X|} P(:,i) \otimes \tilde{Q}(:,i) \otimes R(:,i) - \Theta \|_F^2$.
      $Q(:,i) \leftarrow \frac{Q(:,i)}{\sqrt{N}\|Q(:,i)\|_2}$, $i = 1, \ldots, N$.
      $R \leftarrow \arg\min_{\tilde{R} \in \mathbb{R}^{|X|\times|X|\times|X|}} \| \sum_{i=1}^{|X|} P(:,i) \otimes Q(:,i) \otimes \tilde{R}(:,i) - \Theta \|_F^2$.
      Set k entries of $R(:,i), i = 1, \ldots, |X|$ with the largest magnitude to have magnitude $1/N$.
    end while
  end for
  for i from 1 to $|X|$ **do**
    $\mathcal{I}_i \leftarrow \{j \mid |C(j,i)| > \delta/N\}$.
    Form $\xi_l^{(i)} \in \mathcal{B}_N(X), l = 1, \ldots, \binom{|\mathcal{I}_i|}{N}$. The nonzero entries of $\xi_l^{(i)}$ for each l are indexed by
    a subset of $\mathcal{I}_i$ with N elements.
    $p_i \leftarrow \binom{|\mathcal{I}_i|}{N}$.
  end for
  $\{\lambda_i\}_{i=1}^{p} \leftarrow \cup_{i=1}^{|X|} \{\xi_l^{(i)}\}_{l=1}^{p_i}$
  return $\{\lambda_i\}_{i=1}^{p}$.
end procedure

where γ is the 2-marginal, obtained either via SDP-Coulomb or SDP-Coulomb2 (or their rounded versions). We denote the solution to SDP-Coulomb and SDP-Coulomb2 γ_1^-, γ_2^-, and their rounded solutions γ_1^+, γ_2^+. The superscripts are used to indicate whether we are using the solutions for the purpose of obtaining a lower bound or an upper bound for the energy. We always choose C such that $C(x,y) = \frac{1}{\|x-y\|_2}, x, y \in X$,. In all cases, we choose a box $[-2,2]^d$ where d is the dimension of the space where the electrons reside. A uniform discretization is then applied to $[-2,2]^d$ to get the discrete domain X. We use

$$E_{\mathrm{gap}_i} = \frac{E(\gamma_i^+) - E(\gamma_i^-)}{E(\gamma_i^-)}, \quad i = 1, 2, \tag{80}$$

to provide an idea on how close we are to the true energy. SDP-Coulomb and SDP-Coulomb2 are implemented using the large scale SDP solver SDPNAL+[15].

6.1. Optimizing a linear functional over the 2-marginal. In this section, we let $g(\lambda)$ in (3) be an arbitrary linear functional $c^T\lambda$. This can be seen as an external potential v_{ext} in (9). Then SDP-Coulomb is solved to obtain the 2-marginals. Since one can already devise a rounding scheme (section 5.1) based on the solution of SDP-Coulomb, we only present the energy gap derived from γ_1^- and γ_1^+. Unlike SDP-Coulomb2, SDP-Coulomb only involves a matrix with size $|X| \times |X|$, therefore we can apply it to grids with larger size. The model for the vector c considered is

$$c = \sigma\Big(\min_{i,j} C(i,j)\Big)\mathcal{N}(0,1). \tag{81}$$

In Tables 1 and 2, we present E_{gap_1} for $d = 2, 3$, with $N = 5, 9, 13$. When $d = 2$, we use a grid with size $|X| = 20^2$. When $d = 3$, we let $|X| = 9^3$.

TABLE 1

E_{gap_1} for electrons in two dimensions (2D). Here $|X| = 20^d$, $d = 2$. The energy gap is averaged over 12 realizations of c. The negative gap between the upper and lower bounds when $\sigma = 0, N = 13$ is due to the accuracy limitation of the optimization package.

	$\sigma = 0$	$\sigma = 0.25$	$\sigma = 0.5$
$n = 5$	3.3e-03	7.6e-03	1.3e-02
$n = 9$	3.8e-03	3.0e-03	3.6e-03
$n = 13$	-2.0e-05	3.1e-03	3.4e-03

TABLE 2

E_{gap_1} for electrons in 3D. Here $|X| = 9^d$, $d = 3$. The energy gap is averaged over 12 realizations of c.

	$\sigma = 0$	$\sigma = 0.25$	$\sigma = 0.5$
$n = 5$	3.7e-02	8.1e-03	5e-03
$n = 9$	7.9e-03	5.1e-03	3.5e-03
$n = 13$	3.2e-03	2.8e-03	3.1e-03

6.2. Multimarginal optimal transport. In this section, we present numerical results for different instances of problem (6). Both SDP-Coulomb and SDP-Coulomb2 are tested. Due to the size of the variable in SDP-Coulomb2, we can only afford a smaller grid size. The point of the simulation is to demonstrate how an upper bound of the energy can be extracted using SDP-Coulomb2, through the method presented in section 5.2.

In the case of one dimension, we use three different marginals:

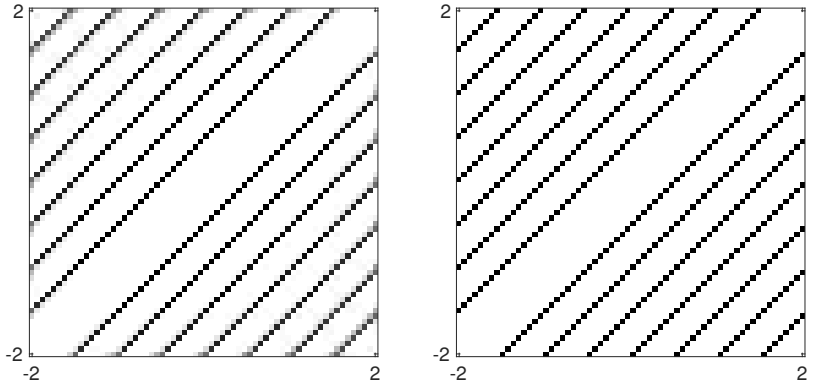
$$\rho_1(x) \propto 1, \quad \rho_2(x) \propto \exp(-x^2/\sqrt{\pi}), \quad \rho_3(x) \propto \sin(4x) + 1.5, \tag{82}$$

where ρ_1, ρ_2, ρ_3 are appropriately normalized. Using the combination of SDP-Coulomb2 and Algorithm 2, an upper-bound can be obtained. We present the results with $|X| = 64$ and $N = 8$ in Figures 2, 3, and 4. In all examples, we obtained an energy gap from the order of 1e-04 to 1e-02. The running times for SDP-Coulomb and SDP-Coulomb2 are about 7s and 249s on average. In general, we observe a fuzzier 2-marginal in SDP-Coulomb, especially when the marginal is ρ_3. We note that the marginals chosen are bounded away from 0. This is because if there are sites where the marginal is close to zero, due to the approximation error of SDP-Coulomb2, Θ^* may be inaccurate on these sites, making rounding difficult. Qualitatively, we observe that when fixing a slice (a horizontal index) of the 2-marginal, the 2-marginals concentrates at $N - 1$ locations. This behavior indicates the existence of Monge solution for the multimarginal optimal transport problem in the one-dimensional (1D) problem, a fact that is shown analytically in [14]. For the two-dimensional (2D) case, we tested it on a Gaussian distribution

$$\rho_4(x, y) \propto \exp(-(x^2 + y^2)/\sqrt{12\pi}) \tag{83}$$

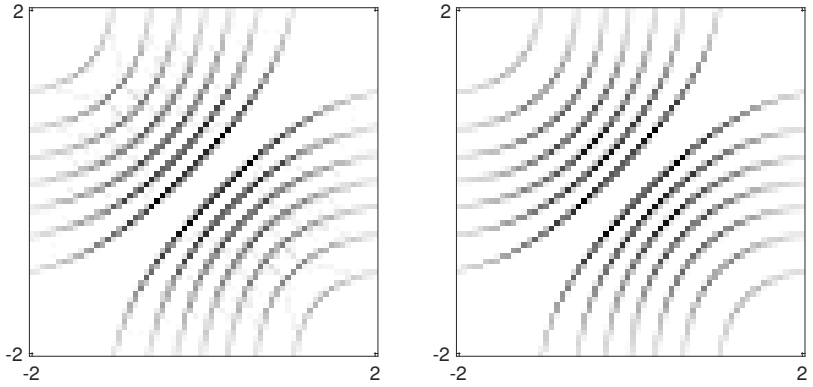
with $|X| = 10^2$ and $N = 6$. The results are presented in Figure 5 and the running times for SDP-Coulomb and SDP-Coulomb2 are 4.7s and 731s, respectively. Again,

the difference between the quality of the solutions from SDP-Coulomb and SDP-Coulomb2 is rather small. We use a rather small grid size in this case since we need to deal with a $100 \times 100 \times 100$ 3-tensor when solving for SDP-Coulomb2. In order to further evaluate the solution qualitatively, SDP-Coulomb is solved using $|X| = 55^2$ in Figure 6. For a slice of the 2-marginal, instead of seeing the support of the 2-marginals concentrated at $N-1$ site, we see a distribution of electrons. This suggests that Monge solution may not exist for 2D or three-dimensional multimarginal optimal transport problems with Coulomb cost.



(a) SDP-Coulomb. (b) SDP-Coulomb2.

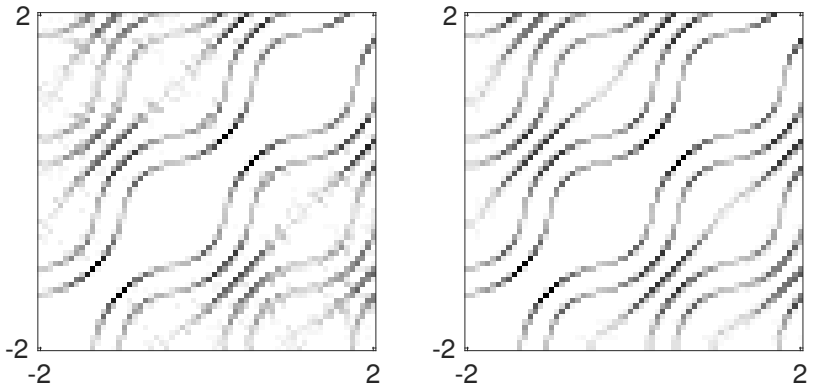
FIG. 2. 2-*marginal from solving the multimarginal transport problem with the marginal* $\rho_1(x)$*, where* $N = 8$*,* $|X| = 64$*,* $d = 1$*. (a): Solution from SDP-Coulomb.* $E_{gap_1} = 4.9e-04$*. (b): Solution from SDP-Coulomb*2*.* $E_{gap_2} = -1.0e-06$*. The negative sign for the energy gap is due to the limitation of numerical accuracy.*



(a) SDP-Coulomb. (b) SDP-Coulomb2.

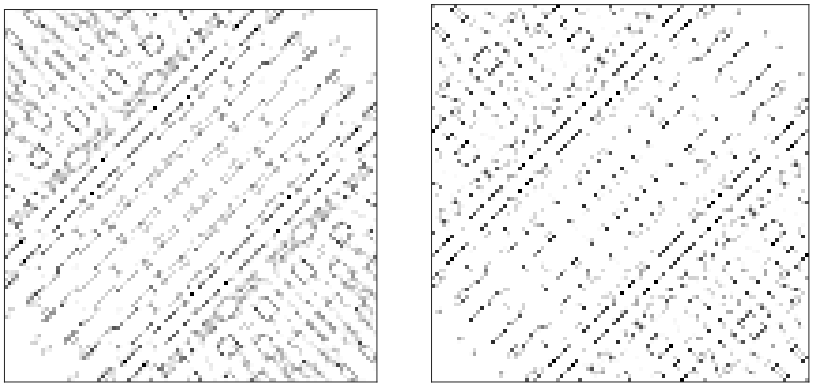
FIG. 3. 2-*marginal from solving the multimarginal transport problem with the marginal* $\rho_2(x)$ *where* $N = 8$*,* $|X| = 64$*,* $d = 1$*. (a): Solution from SDP-Coulomb.* $E_{gap_1} = 1.8e-03$*. (b): Solution from SDP-Coulomb*2*.* $E_{gap_2} = 1.5e-03$*.*

6.3. Approximating the Kantorovich potential. As mentioned previously, the dual problem (32) can also be used to approximate the Kantorovich problem (29). The 1D cases admit semianalytic solutions for the dual potential [13]. First, the



(a) SDP-Coulomb. (b) SDP-Coulomb2.

FIG. 4. 2-*marginal from solving the multimarginal optimal transport problem with the marginal* $\rho_3(x)$ *where* $N = 8$, $|X| = 64$, $d = 1$. (a): *Solution from SDP-Coulomb.* $E_{gap_1} = 4.2e-02$. (b): *Solution from SDP-Coulomb*2. $E_{gap_2} = 3.9e-02$.



(a) SDP-Coulomb. (b) SDP-Coulomb2.

FIG. 5. *Solution to the multimarginal optimal transport problem with the marginal* $\rho_4(x, y)$ *where* $N = 6$, $|X| = 10^2$, $d = 2$. *The* 2*D domain* X *is vectorized in order to present the* 2*-marginal.* (a): *Solution from SDP-Coulomb.* $E_{gap_1} = 3.8e-02$. (b): *Solution from SDP-Coulomb*2. $E_{gap_2} = 3.5e-02$.

comotion function is defined as

$$f_i(x) = \begin{cases} N_e^{-1}(N_e(x) + i - 1), & x \le N_e^{-1}(N + 1 - i), \\ N_e^{-1}(N_e(x) + i - 1 - N), & x > N_e^{-1}(N + 1 - i) \end{cases} \tag{84}$$

for $i = 1, \dots, N$, where

$$N_e := N \int_{-\infty}^{x} \rho(x)dx. \tag{85}$$

Then the Kantorovich potential $v^*(x)$ is defined via

$$\nabla v^*(x) = -\sum_{i \neq 1}^{N} \frac{x - f_i(x)}{\|x - f_i(x)\|_2^3}. \tag{86}$$

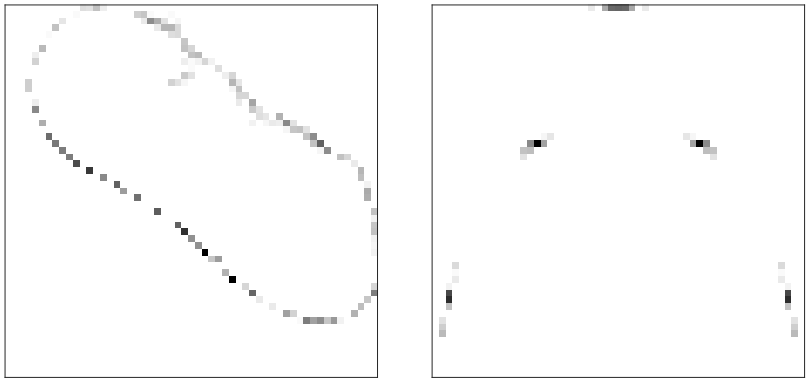
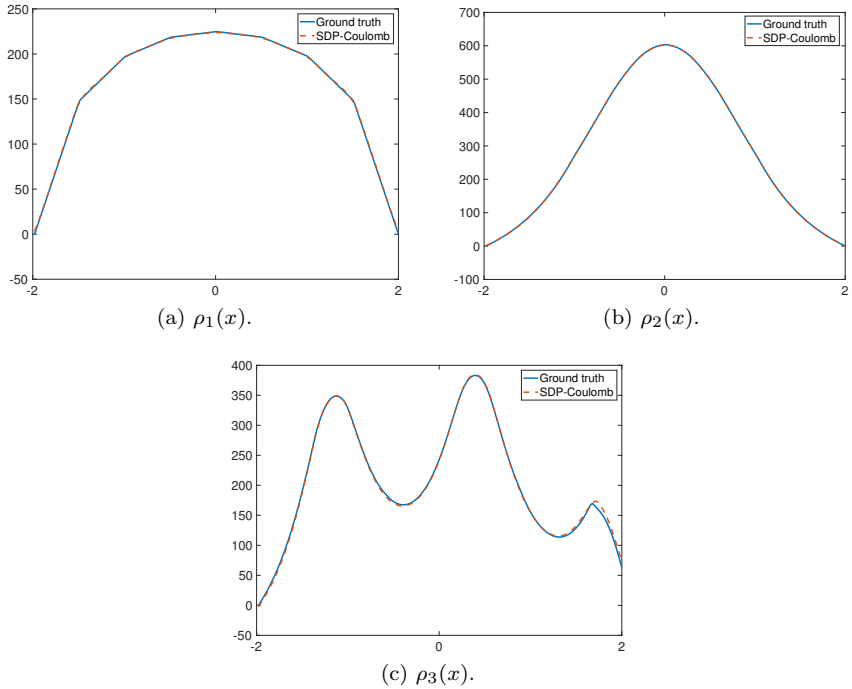


FIG. 6. *Solution to the multimarginal transport problem with the marginal* $\rho_4(x)$ *where* $N = 6$, $|X| = 55^2$, $d = 2$. *Two slices of the* 2*-marginals solved by SDP-Coulomb are presented.*



(a) $\rho_1(x)$. (b) $\rho_2(x)$.

(c) $\rho_3(x)$.

FIG. 7. *Solution to the dual problem* (29), *where* $N = 8$, $|X| = 200$, $d = 1$. *The ground truth is given by* (86), *and the approximation is given by the solution to the dual problem of SDP-Coulomb* (32). (a): *With marginal* $\rho_1(x)$. *Error*$_v = 4.5e-03$. (b): *With marginal* $\rho_2(x)$. *Error*$_v = 1.4e-03$. (c): *With marginal* $\rho_3(x)$. *Error*$_v = 1.2e-02$.

We compare the dual potential w^* obtained from solving (32) to the ground truth Kantorovich potential (86). We let $|X| = 200$ and the marginals being $\rho_1(x), \rho_2(x)$, and $\rho_3(x)$. The error is reported using the metric

$$\text{Error}_v = \frac{\|v^* - w^*\|_2}{\|v^*\|_2}. \tag{87}$$

In these cases, we obtain errors of the order of 10^{-3} to 10^{-2}. The results are presented in Figure 7.

7. Conclusion. We propose methods based on convex relaxation for solving the multimarginal transport type problems in the context of DFT. By convexly relaxing the domain of 2- and 3-marginals, the resulting convex optimization problems have computational complexities independent of the number of electrons. For the numerical simulations presented here, directly applying linear programming or Sinkhorn scaling based algorithm [2] to problem (3) would have led to a tensor with number of entries between 10^{14} to 10^{25}, for the choice of N and $|X|$ used here.

Furthermore, a key feature of the proposed methods is that they provide both upper and lower bounds on the energy. From an algorithmic point of view, it is crucial to develop faster customized optimizer in order to address large-scale applications in the future. From a theoretical point of view, it is important to study theoretically how well SDP-Coulomb and SDP-Coulomb2 approximate problem (3).

Acknowledgments. The authors thank Prof. Lin Lin for introducing the problem. The first author thanks Prof. Emmanuel Candès for the partial support from a Math+X postdoctoral fellowship.

REFERENCES

[1] M. F. Anjos and J. B. Lasserre, *Introduction to semidefinite, conic and polynomial optimization*, in Handbook on Semidefinite, Conic and Polynomial Optimization, Springer, New York, 2012, pp. 1–22.
[2] J.-D. Benamou, G. Carlier, and L. Nenna, *A numerical method to solve multi-marginal optimal transport problems with Coulomb cost*, in Splitting Methods in Communication, Imaging, Science, and Engineering, Springer, Cham, 2016, pp. 577–601.
[3] G. Blekherman, P. A. Parrilo, and R. R. Thomas, *Semidefinite Optimization and Convex Algebraic Geometry*, SIAM, Philadelphia, 2012.
[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
[5] H. Chen, G. Friesecke, and C. B. Mendl, *Numerical methods for a Kohn-Sham density functional model based on optimal transport*, J. Chem. Theory Comput., 10 (2014), pp. 4360–4368.
[6] L. E. Dubins, *On extreme points of convex sets*, J. Math. Anal. Appl., 5 (1962), pp. 237–244.
[7] G. Friesecke and D. Vögler, *Breaking the curse of dimension in multi-marginal Kantorovich optimal transport on finite state spaces*, SIAM J. Math. Anal., 50 (2018), pp. 3996–4019, https://doi.org/10.1137/17M1150025.
[8] P. Gori-Giorgi, M. Seidl, and G. Vignale, *Density-functional theory for strongly interacting electrons*, Phys. Rev. Lett., 103 (2009), 166402.
[9] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500, https://doi.org/10.1137/07070111X.
[10] J. B. Kruskal, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra and Appl., 18 (1977), pp. 95–138.
[11] F. Malet and P. Gori-Giorgi, *Strong correlation in Kohn-Sham density functional theory*, Phys. Rev. Lett., 109 (2012), 246402.
[12] C. B. Mendl and L. Lin, *Kantorovich dual solution for strictly correlated electrons in atoms and molecules*, Phys. Rev. B, 87 (2013), 125106.
[13] M. Seidl, *Strong-interaction limit of density-functional theory*, Phys. Rev. A, 60 (1999), 4387.
[14] M. Seidl, P. Gori-Giorgi, and A. Savin, *Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities*, Phys. Rev. A, 75 (2007), 042511.
[15] L. Yang, D. Sun, and K.-C. Toh, *SDPNAL+: A majorized semismooth Newton-CG augmented lagrangian method for semidefinite programming with nonnegative constraints*, Math. Program. Comput., 7 (2015), pp. 331–366.