Journal of Computational Physics 510 (2024) 113041

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Scalable semidefinite programming approach to variational embedding for quantum many-body problems

Yuehaw Khoo [a], Michael Lindsey [b,c,*]

[a] Department of Statistics, University of Chicago, Chicago, IL 60637, USA
[b] Department of Mathematics, University of California, Berkeley, CA 94720, USA
[c] Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

ARTICLE INFO

Keywords:
Quantum many-body problem
Semidefinite programming
Augmented Lagrangian
Quantum embedding

ABSTRACT

In quantum embedding theories, a quantum many-body system is divided into localized clusters of sites which are treated with an accurate 'high-level' theory and glued together self-consistently by a less accurate 'low-level' theory at the global scale. The recently introduced variational embedding approach for quantum many-body problems combines the insights of semidefinite relaxation and quantum embedding theory to provide a lower bound on the ground-state energy that improves as the cluster size is increased. The variational embedding method is formulated as a semidefinite program (SDP), which can suffer from poor computational scaling when treated with black-box solvers. We exploit the interpretation of this SDP as an embedding method to develop an algorithm which alternates parallelizable local updates of the high-level quantities with updates that enforce the low-level global constraints. Moreover, we show how translation invariance in lattice systems can be exploited to reduce the complexity of projecting a key matrix to the positive semidefinite cone.

1. Introduction

The problem of determining the ground state of a quantum many-body system has wide-ranging applications in physics, chemistry, and materials science. This problem can be viewed as the problem of determining the lowest eigenvalue of a Hermitian operator on a Hilbert space whose dimension grows exponentially with the size of the system or the number of particles, such as electrons in the case of electronic structure. Here we highlight two relevant categories of approaches to taming this curse of dimensionality.

The first category is that of semidefinite relaxations, which rephrase the aforementioned energy minimization problem as an optimization problem in terms of a reduced set of physical observables, almost always a semidefinite program (SDP). In principle these observables satisfy representability constraints, i.e., constraints that ensure that they can be recovered from a bona fide quantum many-body state. However, only a subset of representability constraints can be efficiently enforced, yielding tractable optimization problems that provide lower bounds on the ground-state energy. Such approaches include the 2-RDM theories [28,26,4,27,42,24,1,29,7], as well as methods that may be classified as quantum marginal relaxations such as [23,30,9,16].

Meanwhile, quantum embedding theories take the perspective of dividing a system into local clusters, small enough to be treated with a highly accurate or exact method referred to as the 'high-level' method. Local problems are then stitched together via a reduced set of global quantities or a less-accurate 'low-level' method that operates on the global scale, and the local and global

* Corresponding author.
E-mail address: lindsey@berkeley.edu (M. Lindsey).

https://doi.org/10.1016/j.jcp.2024.113041
Received 5 June 2021; Received in revised form 1 February 2024; Accepted 19 April 2024

Available online 3 May 2024

perspectives are constrained to be compatible via some self-consistency condition. Such approaches include dynamical mean-field theory (DMFT) [13,22] and density matrix embedding theory (DMET) [20,21], as well as variants such as the energy-weighted DMET (EwDMET) [10,11] which in a certain sense interpolates between DMFT and DMET [35].

Recently, variational embedding [25] was introduced as a semidefinite relaxation which is also a quantum embedding method. Like other relaxations such as [23,30,9,16], the key optimization variables are quantum marginals for local clusters, but variational embedding additionally includes *global* constraints which tighten the relaxation and accommodate the treatment of, e.g., long-range interactions.

1.1. Contribution

As an SDP, variational embedding can be solved with black-box methods, as is done in [25], but scalability calls for a solver that is specially adapted to the problem. In this work, we introduce a scalable solver for this SDP which takes advantage of the embedding interpretation of the approach. The aforementioned global constraint is dualized to reduce the problem to a simpler relaxation (similar to those of [23,30,9,16]) in which the global constraints have been exchanged for effective contributions to relevant Hamiltonian operators at the local scale. This problem can then be solved in a fashion in which effective problems for the key variables (the two-cluster marginals) are completely decoupled and can be solved in parallel, with dual variable optimization enforcing the self-consistency of these problems. Furthermore, translation invariance of a lattice system can be used to significantly speed up the running time.

We now briefly mention some further details of our method for solving the simplified problem yielded by partial dualization. Our approach is similar to that of augmented Lagrangian methods (ALM) [17,31,32,37], in that it is derived from an augmented Lagrangian function. However, the primal update for the ALM suggested by our problem is too expensive to solve exactly in each iteration. Inexact ALMs [33] provide one option for circumventing this difficulty, but due to the separable structure of our objective (after partial dualization of the global constraint), it is natural for us to pursue an ADMM-type approach for this subproblem. The alternating direction method of multipliers (ADMM) was introduced by [15,12], and see [40] for recent developments. This approach allows us to solve several decoupled, physically local effective sub-problems completely in parallel, preserving the flavor of a quantum embedding theory.

We note that [41,36] have developed variants of ALM and ADMM, respectively, to solve large-scale SDPs. One key difference of our work is that we dualize a partial set of semidefinite constraints, then we apply ADMM to solve the partially dualized problem, which has convenient revealed structure. On the other hand, [41] and [36] directly work with the full primal and dual problems respectively.

1.2. Outline

In Section 2 we provide relevant background on the ground-state eigenvalue problem, examples of interest, and the two-marginal relaxation for variational embedding introduced in [25]. (In Appendix A, additional background is provided for the context of fermionic systems.) In Section 3, we describe our optimization approach to this problem, which is an SDP. The section begins with an idealized scheme of projected gradient ascent on the dual variable to the aforementioned global constraint. In order to implement such a scheme, it is necessary to solve an effective problem in terms of the primal variables. In Section 3.1, we introduce an ADMM-type approach to this problem, and in Section 3.2 we integrate this approach with dual ascent to define our practical scheme. In Section 3.3 we explain how translation-invariance can be exploited, and in Section 3.4 we include a detailed discussion of the computational scaling. Finally in Section 4 we present numerical experiments on several model systems of quantum spins and fermions.

2. Preliminaries

In this section we review the formulation of variational embedding for quantum spins, following [25]. In Appendix A, we review the case of fermions (also following [25]), which requires a bit more care but nonetheless yields a semidefinite program of identical form after suitable manipulations.

2.1. The ground-state eigenvalue problem

We consider a model with M sites, indexed $i = 1, \dots, M$, each endowed with a classical local state space X_i (which shall be discrete). These in turn yields local quantum state spaces $Q_i = \mathbb{C}^{|X_i|}$. The global quantum state space (i.e., the space of wavefunctions) is then given by

$$\mathcal{Q} := \bigotimes_{i=1}^{M} Q_i \simeq \mathbb{C}^{|\mathcal{X}|},$$

where $\mathcal{X} := \prod_{i=1}^{M} X_i$ is the global classical state space, so quantum states (wavefunctions) correspond to complex-valued functions on the classical state space. Let H_i (resp. H_{ij}) denote Hermitian operators $Q_i \to Q_i$ (resp. $Q_{ij} \to Q_{ij}$), and let $\hat{H}_i$ (resp. $\hat{H}_{ij}$) denote

the corresponding operators $\mathcal{Q} \to \mathcal{Q}$ obtained by tensoring H_i (resp. H_{ij}) by the identity operator on all sites $k \neq i$ (resp. $k \neq i, j$). We consider *pairwise* Hamiltonian $\hat{H} : \mathcal{Q} \to \mathcal{Q}$ of the form

$$\hat{H} = \sum_i \hat{H}_i + \sum_{i<j} \hat{H}_{ij},$$

and our interest is in determining the ground-state energy, i.e., the smallest eigenvalue, of $\hat{H}$. We denote this eigenvalue by E_0, which is defined variationally by

$$E_0 = \inf \left\{ \Phi^* \hat{H} \Phi \, : \, \Phi \in \mathcal{Q}, \, \Phi^* \Phi = 1 \right\}. \tag{2.1}$$

Now we review some examples of interest. First consider the case of quantum spin-$\frac{1}{2}$ systems, i.e., the case $X_i = \{-1, 1\}$. To construct operators on $\mathcal{Q}$, one first starts with the Pauli matrices

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

which (together with the identity I_2) form a basis for the real vector space of Hermitian operators on $\mathbb{C}^2$. We let $\sigma_i^{x/y/z}$ denote the operator $\mathcal{Q} \to \mathcal{Q}$ obtained by tensoring $\sigma^{x/y/z}$ on the i-th site with I_2 on all other sites. Then in terms of these operators we can define the transverse-field Ising (TFI) Hamiltonian and the anti-ferromagnetic Heisenberg (AFH) Hamiltonian by

$$\hat{H}_{\text{TFI}} = -h \sum_i \sigma_i^x - \sum_{i \sim j} \sigma_i^z \sigma_j^z \tag{2.2}$$

$$\hat{H}_{\text{AFH}} = \sum_{i \sim j} \left[\sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y + \sigma_i^z \sigma_j^z \right], \tag{2.3}$$

h is a scalar parameter and summation over $i \sim j$ indicates summation over pairs of indices that are adjacent within some graph defined on the index set $\{1, \dots, M\}$, often a rectangular lattice in some dimension. These problems have been considered as prototypical quantum many-body problems, e.g., in [5], as well as models for the study of quantum phase transitions, as in [34].

2.2. The two-marginal relaxation

In [25], the optimization problem (2.1) is reformulated as an optimization over the density operator ρ, which (for nondegenerate ground states) corresponds at optimality to $\Phi_0 \Phi_0^*$ where Φ_0 is the ground state eigenvector, i.e., the optimizer of (2.1). This optimization is in turn relaxed as a computationally tractable optimization over the quantum two-marginals

$$\rho_{ij} = \text{Tr}_{\{i,j\}^c}[\rho], \quad i < j,$$

which are defined as partial traces of ρ, analogous to classical marginals. Indeed, recall [25] that for any subset $S \subset \{1, \dots, M\}$, the partial trace $\rho_S = \text{Tr}_{S^c}[\rho]$ may be defined as the unique operator on $\bigotimes_{i \in S} Q_i$ such that $\text{Tr}[A \rho_S] = \text{Tr}[\hat{A} \rho]$ for all operators A on $\bigotimes_{i \in S} Q_i$ (lifted to operators $\hat{A}$ on $\mathcal{Q}$ by tensoring with the identity on S^c). In particular, ρ_{ij} is an operator on $Q_i \otimes Q_j$, and moreover it is positive semidefinite with unit trace (following from the same properties for ρ).

Then the two-marginal relaxation of [25] reads in terms of the two marginals (and the analogously-defined one-marginals, which can be obtained from the two-marginals by further partial trace) as the following semidefinite program, whose optimal value we denote by $E_0^{(2)}$:

$$\underset{\{\rho_i\}, \{\rho_{ij}\}_{i<j}}{\text{minimize}} \ \sum_i \text{Tr}\left[H_i \rho_i\right] + \sum_{i<j} \text{Tr}\left[H_{ij} \rho_{ij}\right] \tag{2.4}$$

$$\text{subject to } \rho_{ij} \succeq 0, \quad 1 \leq i < j \leq M, \tag{2.5}$$

$$\rho_i = \text{Tr}_{\{2\}}[\rho_{ij}], \ \rho_j = \text{Tr}_{\{1\}}[\rho_{ij}], \quad 1 \leq i < j \leq M, \tag{2.6}$$

$$\text{Tr}[\rho_i] = 1, \quad i = 1, \dots, M, \tag{2.7}$$

$$G[\{\rho_i\}, \{\rho_{ij}\}_{i \leq j}] \succeq 0. \tag{2.8}$$

Here $G = G[\{\rho_i\}, \{\rho_{ij}\}_{i \leq j}]$ is an operator defined linearly in terms of the one- and two-marginals, subordinate to the specification of an arbitrary collection $\{O_{i,\alpha} \, : \, \alpha = 1, \dots, n_i\}$ of linear operators $Q_i \to Q_i$ at each site $i = 1, \dots, M$. In specific, G is specified blockwise, with blocks G_{ij} for $1 \leq i, j, \leq M$ of size $n_i \times n_j$ defined by

$$\left(G_{ij}\right)_{\alpha\beta} = \begin{cases} \text{Tr}\left[O_{i,\alpha}^\dagger O_{i,\beta} \, \rho_i\right] & i = j \\ \text{Tr}\left[\left(O_{i,\alpha}^\dagger \otimes O_{j,\beta}\right) \rho_{ij}\right] & i \neq j. \end{cases}$$

The choice of operators only matters up to $\mathrm{span}\{O_{i,\alpha} : \alpha = 1, \dots, n_i\}$, and in our numerical experiments we shall consider the complete operator collection spanning all linear maps $Q_i \to Q_i$. The last constraint (2.8) is called the *global semidefinite constraint*.

2.2.1. Classical marginal relaxation

To motivate the relaxation (2.4) further, we examine the problem of finding the lowest energy state of a classical energy function of a pairwise form. (This can be viewed as a special case of the more general quantum ground-state problem by taking the H_{ij} to be diagonal operators.) More concretely, for $x_i \in X_i$, $i = 1, \dots, M$, define an energy function

$$E(x_1, \dots, x_M) = \sum_{i<j} E_{ij}(x_i, x_j), \tag{2.9}$$

and observe that the minimizer of E can be determined via the linear program

$$\underset{\mu \in \Pi(\mathcal{X})}{\arg\min} \sum_{(x_1,\dots,x_M)\in\mathcal{X}} E(x_1, \dots, x_M)\, \mu(x_1, \dots, x_M), \tag{2.10}$$

where $\Pi(\mathcal{X})$ is the space of probability measures on $\mathcal{X}$. Indeed, the optimizer is a δ-function supported on the minimizer of E (provided that it is unique). Exploiting the pairwise structure of E, we have

$$\underset{\{\mu_{ij}\}_{i<j}\ \mathrm{repr.}}{\arg\min} \sum_{i<j} \sum_{x_i, x_j \in \mathcal{X}} E_{ij}(x_i, x_j)\, \mu_{ij}(x_i, x_j) \tag{2.11}$$

where the two-marginal variables $\{\mu_{ij}\}_{i<j}$ are constrained to be jointly representable, i.e., to be derivable as the two-marginals of a high-dimensional measure μ. Enforcing this constraint demands exponential complexity, so various convex relaxation approaches have been proposed, where only certain necessary conditions for the $\{\mu_{ij}\}_{i<j}$ are kept; see, for instance, [39] for a review. In particular, the analogous convex relaxation to (2.4) is

$$\underset{\{\mu_i\}, \{\mu_{ij}\}_{i<j}}{\text{minimize}} \sum_{i<j} \mathrm{Tr}\left[E_{ij}\mu_{ij}\right] \tag{2.12}$$

$$\text{subject to } \mu_{ij} \geq 0, \quad 1 \leq i < j \leq M, \tag{2.13}$$

$$\mu_i = \mu_{ij}\mathbf{1},\ \mu_j = \mu_{ij}^\top \mathbf{1}, \quad 1 \leq i < j \leq M, \tag{2.14}$$

$$\mu_i^\top \mathbf{1} = 1, \quad i = 1, \dots, M, \tag{2.15}$$

$$\begin{bmatrix} \mathrm{diag}(\mu_1) & \mu_{12} & \cdots & \mu_{1M} \\ \mu_{21} & & & \\ \vdots & & \ddots & \vdots \\ \mu_{M1} & \cdots & & \mathrm{diag}(\mu_M) \end{bmatrix} \succeq 0, \tag{2.16}$$

where $\mathbf{1}$ is an all-one vector of appropriate size. Here (2.13) and (2.15) are standard constraints for any discrete probability distributions, and (2.14) are constraints that enforces 'local consistency' [39] among the $\{\mu_{ij}\}_{i<j}$. The global semidefinite constraint (2.16) is discussed in [19] and [6] in the contexts of multi-marginal optimal transport and energy minimization, respectively.

2.3. Cluster relaxation

Given a quantum spin model as above and a decomposition of the sites $\{1, \dots, M\}$ as a disjoint union $\bigcup_{i'=1}^{M'} C_{i'}$ of clusters $C_{i'}$, we may define $X'_{i'} := \prod_{i\in C_{i'}} C_{i'}$ to be the classical state space for the i'-th cluster. One see that any Hamiltonian that is pairwise with respect to sites is pairwise with respect to clusters, so by viewing our clusters as sites and applying the above formalism, we obtain a tighter relaxation [25].

2.4. Partial duality

In [25] it was shown that (2.4) admits the minimax formalization (obtained via dualization of the global semidefinite constraint (2.8))

$$E_0^{(2)} = \sup_{X \succeq 0} \mathcal{F}[X], \tag{2.17}$$

where

$$\mathcal{F}[X] := \inf \left\{ \sum_i \mathrm{Tr}\left(H_i[X_{ii}]\rho_i\right) + \sum_{i<j} \mathrm{Tr}\left(H_{ij}[X_{ij}]\rho_{ij}\right) \ : \ \{\rho_i\}, \{\rho_{ij}\}_{i<j} \text{ satisfy (2.5)-(2.7)} \right\}. \tag{2.18}$$

Here X_{ij} denote the blocks of X, and the 'effective' Hamiltonian terms $H_i[X_{ii}]$ and $H_{ij}[X_{ij}]$ are defined linearly in terms of X via

$$H_i[X_{ii}] := H_i - \sum_{\alpha,\beta=1}^{n_i} (\overline{X}_{ii})_{\alpha\beta} O_{i,\alpha}^\dagger O_{i,\beta}, \quad H_{ij}[X_{ij}] := H_i - \left[\sum_{\alpha=1}^{n_i}\sum_{\beta=1}^{n_j} (\overline{X}_{ij})_{\alpha\beta}\left(O_{i,\alpha}^\dagger \otimes O_{i,\beta}\right) + \text{h.c.}\right], \tag{2.19}$$

where 'h.c.' denotes the Hermitian conjugate term.

This partial dual formulation can be obtained from (2.4) by exchanging the global semidefinite constraint (2.8) for an extra term

$$-\mathrm{Tr}[X\, G[\{\rho_i\}, \{\rho_{ij}\}_{i\le j}]]$$

in the Lagrangian, where $X \succeq 0$ is a dual variable with respect to which the Lagrangian is to be maximized. Then (2.19) is recovered by breaking this additional term into a blockwise sum, collecting terms, and minimizing over the primal variables, subject to the remaining constraints.

Since X and G are dual variables [25] we have that

$$\nabla_X \mathcal{F}[X] = -G[\{\rho_i\}, \{\rho_{ij}\}_{i<j}], \tag{2.20}$$

where $\{\rho_i\}, \{\rho_{ij}\}_{i<j}$ are the minimizers of the infimum in (2.18).

3. Optimization approach

In order to solve the two-marginal relaxation (2.4), our point of departure is the partial dual formulation (2.17). For simplicity we consider the case in which the Hamiltonian and all variables are purely real, and we let $\Pi_{\succeq 0}$ be defined by

$$\Pi_{\succeq 0}(C) := \min_{S\succeq 0} \|S - C\|_F^2, \tag{3.1}$$

for a symmetric matrix A, i.e. the Euclidean projection (in Frobenius norm) of a symmetric matrix onto the set of real symmetric positive semidefinite matrices (equivalent to setting all negative eigenvalues of the argument to zero). As an idealized scheme, we can imagine performing projected gradient ascent on (2.17), which is implemented by Algorithm 1. In practice, the step size $\varepsilon > 0$ in Algorithm 1 should be chosen sufficiently small to guarantee convergence, but not much smaller.

Algorithm 1 Exact projected gradient ascent.

Require: $\varepsilon > 0$, $X \succeq 0$
1: **while** not converged **do**
2: Set $\left(\{\rho_i\}, \{\rho_{ij}\}_{i<j}\right)$ to be the minimizer in (2.18), holding X fixed
3: $X \leftarrow \Pi_{\succeq 0}\left(X + \varepsilon G[\{\rho_i\}, \{\rho_{ij}\}_{i<j}]\right)$
4: **end while**

The focus of this section is in the development of algorithm for step 2 for a general Hamiltonian, and we also study the case in the presence of translational invariance. In practice, we will not fully converge a solution to step 2 of Algorithm 1, resulting in an inexact projected gradient ascent scheme. However, in order to motivate our practical scheme, we will first discuss how to solve step 2 exactly for fixed X.

3.1. Details for step 2 in Algorithm 1

We rephrase step 2 as the following optimization problem:

$$\begin{aligned} \underset{\{\rho_i\},\{\rho_{ij}\}_{i<j}}{\text{minimize}} \ & \sum_i \mathrm{Tr}\left[H_i' \rho_i\right] + \sum_{i<j} \mathrm{Tr}\left[H_{ij}' \rho_{ij}\right] \\ \text{subject to} \ & \rho_{ij} \succeq 0, \quad 1 \le i < j \le M, \\ & \rho_i = A_1[\rho_{ij}], \ \rho_j = A_2[\rho_{ij}], \quad 1 \le i < j \le M, \\ & \mathrm{Tr}[\rho_i] = 1, \quad i = 1, \dots, M, \end{aligned} \tag{3.2}$$

where $H_i' := H_i[X_{ii}]$, $H_{ij}' := H_{ij}[X_{ij}]$, and X is fixed for the duration of this subsection. Moreover, for simplicity we have assumed that $m := |X_i|$ is constant, and A_1, A_2 are defined to be the linear operators $\mathrm{Tr}_{\{2\}}$ and $\mathrm{Tr}_{\{1\}}$, respectively. Hence A_1, A_2 can be realized as sparse matrices of size $m^2 \times m^4$ We then formulate an equivalent optimization problem via the introduction of dummy variables $\tilde{\rho}_{ij}$ and the inclusion of augmented Lagrangian terms in the objective:

$$\begin{aligned} \underset{\{\rho_i\},\{\rho_{ij},\tilde{\rho}_{ij}\}_{i<j}}{\text{minimize}} \ & \sum_i \mathrm{Tr}\left[H_i' \rho_i\right] + \sum_{i<j} \mathrm{Tr}\left[H_{ij}' \rho_{ij}\right] \\ & + \sum_{i<j}\left(\frac{\mu}{2}\|\rho_{ij} - \tilde{\rho}_{ij}\|_{\mathrm{F}}^2 + \frac{\nu}{2}\|\rho_i - A_1[\rho_{ij}]\|_{\mathrm{F}}^2 + \frac{\nu}{2}\|\rho_j - A_2[\rho_{ij}]\|_{\mathrm{F}}^2\right) \end{aligned} \tag{3.3}$$

$$\text{subject to} \ \tilde{\rho}_{ij} \succeq 0, \quad 1 \le i < j \le M,$$

$$\Lambda_{ij} \ : \ \rho_{ij} = \tilde{\rho}_{ij}, \quad 1 \le i < j \le M, \tag{3.4}$$

$$\Lambda_{ij}^{(1)}:\ \rho_i = A_1[\rho_{ij}],\ \Lambda_{ij}^{(2)}:\ \rho_j = A_2[\rho_{ij}],\quad 1 \le i < j \le M,$$
$$\mathrm{Tr}[\rho_i] = 1,\quad i = 1,\dots,M, \tag{3.5}$$

where $\mu, \nu > 0$ are constant parameters, and $\Lambda_{ij}, \Lambda_{ij}^{(1)}, \Lambda_{ij}^{(2)} \in \mathrm{End}(Q_i \otimes Q_j)$ are the dual variables for the associated constraints in (3.4) and (3.5). Let $f[\{\rho_i\}, \{\rho_{ij}, \tilde{\rho}_{ij}\}_{i<j}]$ denote the objective function (3.3), yielding the Lagrangian

$$\begin{aligned}&\mathcal{L}\left(\{\rho_i\}, \{\rho_{ij}, \tilde{\rho}_{ij}\}_{i<j}; \{\Lambda_{ij}, \Lambda_{ij}^{(1)}, \Lambda_{ij}^{(2)}\}_{i<j}\right)\\ &:= f[\{\rho_i\}, \{\rho_{ij}, \tilde{\rho}_{ij}\}_{i<j}] + \sum_{i<j}\left(\left\langle \Lambda_{ij}, \tilde{\rho}_{ij} - \rho_{ij}\right\rangle_{\mathrm{F}} + \left\langle \Lambda_{ij}^{(1)}, A_1[\rho_{ij}] - \rho_i\right\rangle_{\mathrm{F}} + \left\langle \Lambda_{ij}^{(2)}, A_2[\rho_{ij}] - \rho_j\right\rangle_{\mathrm{F}}\right),\end{aligned} \tag{3.6}$$

with domain specified by the (undualized) primal constraints $\tilde{\rho}_{ij} \succeq 0$ and $\mathrm{Tr}[\rho_i] = 1$. Here $\langle \cdot, \cdot \rangle_{\mathrm{F}}$ indicates the Frobenius inner product. Then the Augmented Lagrangian method [2] for (3.2) is implemented by Algorithm 2.

Algorithm 2 Augmented Lagrangian method for (3.2).

Require: $\mu, \nu > 0$, $\{H_i'\}$, $\{H_{ij}', \Lambda_{ij}, \Lambda_{ij}^{(1)}, \Lambda_{ij}^{(2)}\}_{i<j}$
1: **while** not converged **do**
2: $\{\rho_i\}, \{\rho_{ij}, \tilde{\rho}_{ij}\}_{i<j} \leftarrow \underset{\{\rho_i\}, \{\rho_{ij}, \tilde{\rho}_{ij}\}_{i<j}}{\arg\min}\ \mathcal{L}\left(\{\rho_i\}, \{\rho_{ij}, \tilde{\rho}_{ij}\}_{i<j}; \{\Lambda_{ij}, \Lambda_{ij}^{(1)}, \Lambda_{ij}^{(2)}\}_{i<j}\right)$
3: **for** each pair $i < j$ **do**
4: $\Lambda_{ij} \leftarrow \Lambda_{ij} + \mu\left(\tilde{\rho}_{ij} - \rho_{ij}\right)$
5: $\Lambda_{ij}^{(1)} \leftarrow \Lambda_{ij}^{(1)} + \nu(A_1[\rho_{ij}] - \rho_i)$
6: $\Lambda_{ij}^{(2)} \leftarrow \Lambda_{ij}^{(2)} + \nu(A_2[\rho_{ij}] - \rho_j)$
7: **end for**
8: **end while**

In practice, it is difficult to solve step 2 of Algorithm 2 exactly. Therefore, instead of optimizing $\{\rho_i\}, \{\rho_{ij}, \tilde{\rho}_{ij}\}_{i<j}$ jointly, we consider an ADMM-type [3] substitute, namely Algorithm 3. Notice that in step 2 of Algorithm 3, the ρ_{ij} are all determined *independently*

Algorithm 3 Pseudo-code for ADMM-type method for (3.2).

Require: $\mu, \nu > 0$, $\{H_i', \rho_i\}$, $\{H_{ij}', \tilde{\rho}_{ij}, \Lambda_{ij}, \Lambda_{ij}^{(1)}, \Lambda_{ij}^{(2)}\}_{i<j}$
1: **while** not converged **do**
2: $\{\rho_{ij}\}_{i<j} \leftarrow \underset{\{\rho_{ij}\}_{i<j}}{\arg\min}\ \mathcal{L}\left(\{\rho_i\}, \{\rho_{ij}, \tilde{\rho}_{ij}\}_{i<j}; \{\Lambda_{ij}, \Lambda_{ij}^{(1)}, \Lambda_{ij}^{(2)}\}_{i<j}\right)$
3: $\{\rho_i\}, \{\tilde{\rho}_{ij}\}_{i<j} \leftarrow \underset{\{\rho_i\}, \{\tilde{\rho}_{ij}\}_{i<j}}{\arg\min}\ \mathcal{L}\left(\{\rho_i\}, \{\rho_{ij}, \tilde{\rho}_{ij}\}_{i<j}; \{\Lambda_{ij}, \Lambda_{ij}^{(1)}, \Lambda_{ij}^{(2)}\}_{i<j}\right)$
4: **for** each pair $i < j$ **do**
5: $\Lambda_{ij} \leftarrow \Lambda_{ij} + \mu\left(\tilde{\rho}_{ij} - \rho_{ij}\right)$
6: $\Lambda_{ij}^{(1)} \leftarrow \Lambda_{ij}^{(1)} + \nu(A_1[\rho_{ij}] - \rho_i)$
7: $\Lambda_{ij}^{(2)} \leftarrow \Lambda_{ij}^{(2)} + \nu(A_2[\rho_{ij}] - \rho_j)$
8: **end for**
9: **end while**

as the solutions of decoupled optimization problems

$$\begin{aligned}\rho_{ij} \leftarrow \underset{\rho_{ij}}{\mathrm{argmin}}\Bigg\{&\left\langle H_{ij}', \rho_{ij}\right\rangle_{\mathrm{F}} + \frac{\mu}{2}\|\rho_{ij} - \tilde{\rho}_{ij}\|_{\mathrm{F}}^2 + \frac{\nu}{2}\|A_1[\rho_{ij}] - \rho_i\|_{\mathrm{F}}^2 + \frac{\nu}{2}\|A_2[\rho_{ij}] - \rho_j\|_{\mathrm{F}}^2\\ &- \left\langle \Lambda_{ij}, \rho_{ij}\right\rangle_{\mathrm{F}} + \left\langle \Lambda_{ij}^{(1)}, A_1[\rho_{ij}]\right\rangle_{\mathrm{F}} + \left\langle \Lambda_{ij}^{(2)}, A_2[\rho_{ij}]\right\rangle_{\mathrm{F}}\Bigg\}.\end{aligned}$$

After suitable manipulation of the objective (neglecting constant terms), we obtain

$$\frac{1}{2}\left\langle \rho_{ij}, \left(\mu I + \nu A_1^* A_1 + \nu A_2^* A_2\right)\rho_{ij}\right\rangle_{\mathrm{F}} - \left\langle \mu\tilde{\rho}_{ij} + A_1^*\left[\nu\rho_i - \Lambda_{ij}^{(1)}\right] + A_2^*\left[\nu\rho_j - \Lambda_{ij}^{(2)}\right] + \Lambda_{ij} - H_{ij}', \rho_{ij}\right\rangle_{\mathrm{F}},$$

which can be exactly optimized via the update

$$\rho_{ij} \leftarrow \left(\mu I + \nu A_1^* A_1 + \nu A_2^* A_2\right)^{-1}\left(\mu\tilde{\rho}_{ij} + A_1^*\left[\nu\rho_i - \Lambda_{ij}^{(1)}\right] + A_2^*\left[\nu\rho_j - \Lambda_{ij}^{(2)}\right] + \Lambda_{ij} - H_{ij}'\right). \tag{3.7}$$

Meanwhile, in step 3, the ρ_i and the $\tilde{\rho}_{ij}$ can all be updated via decoupled optimization problems. In particular, we find that

$$\tilde{\rho}_{ij} \leftarrow \underset{\tilde{\rho}_{ij} \succeq 0}{\mathrm{argmin}}\left\{\|\tilde{\rho}_{ij} - (\rho_{ij} - \mu^{-1}\Lambda_{ij})\|^2\right\} = \Pi_{\succeq 0}\left(\rho_{ij} - \mu^{-1}\Lambda_{ij}\right). \tag{3.8}$$

Finally we turn to the ρ_i update. Collecting the relevant terms we have that

$$\rho_i \leftarrow \underset{\rho_i\,:\,\mathrm{Tr}[\rho_i]=1}{\mathrm{argmin}} \left\{ \left\langle H_i', \rho_i \right\rangle_{\mathrm{F}} + \sum_{j>i} \left(\frac{\nu}{2} \|\rho_i - A_1[\rho_{ij}]\|_{\mathrm{F}}^2 - \left\langle \Lambda_{ij}^{(1)}, \rho_i \right\rangle_{\mathrm{F}} \right) + \sum_{j<i} \left(\frac{\nu}{2} \|\rho_i - A_2[\rho_{ji}]\|_{\mathrm{F}}^2 - \left\langle \Lambda_{ji}^{(2)}, \rho_i \right\rangle_{\mathrm{F}} \right) \right\}.$$

Observe that the objective may be rewritten as

$$\frac{(M-1)\nu}{2} \|\rho_i\|_{\mathrm{F}}^2 - \left\langle \sum_{j>i} \left(\nu A_1[\rho_{ij}] + \Lambda_{ij}^{(1)} \right) + \sum_{j<i} \left(\nu A_2[\rho_{ji}] + \Lambda_{ji}^{(2)} \right) - H_i', \rho_i \right\rangle_{\mathrm{F}},$$

which we must minimize subject to $\mathrm{Tr}\left[\rho_i\right] = 1$. This is simply a constrained least squares problem, the solution of which yields the update

$$\rho_i \leftarrow \frac{1}{\nu(M-1)} \left(\sum_{j>i} \left(\nu A_1[\rho_{ij}] + \Lambda_{ij}^{(1)} \right) + \sum_{j<i} \left(\nu A_2[\rho_{ji}] + \Lambda_{ji}^{(2)} \right) - H_i' \right) + z I_m, \tag{3.9}$$

where z is a Lagrange multiplier chosen to satisfy the constraint.

Then via (3.7), (3.8), and (3.9), we can rewrite Algorithm 3 concretely as the equivalent Algorithm 4. Observe that all of the for-loops in Algorithm 4 can be run in parallel.

Algorithm 4 Details of the ADMM-type method for (3.2).

Require: $\mu, \nu > 0$, $\{H_i', \rho_i\}$, $\{H_{ij}', \tilde{\rho}_{ij}, \Lambda_{ij}, \Lambda_{ij}^{(1)}, \Lambda_{ij}^{(2)}\}_{i<j}$

while not converged **do**
  for each pair $i < j$ **do**
    $\rho_{ij} \leftarrow \left(\mu I + \nu A_1^* A_1 + \nu A_2^* A_2\right)^{-1} \left(\mu \tilde{\rho}_{ij} + A_1^* \left[\nu \rho_i - \Lambda_{ij}^{(1)} \right] + A_2^* \left[\nu \rho_j - \Lambda_{ij}^{(2)} \right] + \Lambda_{ij} - H_{ij}' \right)$
  end for
  for each pair $i < j$ **do**
    $\tilde{\rho}_{ij} \leftarrow \Pi_{\succeq 0} \left(\rho_{ij} - \mu^{-1} \Lambda_{ij} \right)$
  end for
  for each i **do**
    $\rho_i' \leftarrow \frac{1}{\nu(M-1)} \left(\sum_{j>i} \left(\nu A_1[\rho_{ij}] + \Lambda_{ij}^{(1)} \right) + \sum_{j<i} \left(\nu A_2[\rho_{ji}] + \Lambda_{ji}^{(2)} \right) - H_i' \right)$
    $z \leftarrow m^{-1}(1 - \mathrm{Tr}[\rho_i'])$
    $\rho_i \leftarrow \rho_i' + z I_m$
  end for
  for each pair $i < j$ **do**
    $\Lambda_{ij} \leftarrow \Lambda_{ij} + \mu \left(\tilde{\rho}_{ij} - \rho_{ij} \right)$
    $\Lambda_{ij}^{(1)} \leftarrow \Lambda_{ij}^{(1)} + \nu(A_1[\rho_{ij}] - \rho_i)$
    $\Lambda_{ij}^{(2)} \leftarrow \Lambda_{ij}^{(2)} + \nu(A_2[\rho_{ij}] - \rho_j)$
  end for
end while

3.2. Practical scheme

Now we return to the full Algorithm 1 where we optimize X via an idealized gradient ascent. Instead of exactly implementing step 2 of Algorithm 1, we replace it with a single iteration of Algorithm 4, yielding our practical approach Algorithm 5 for solving the two-marginal relaxation (2.4).

Note that within Algorithm 5 it is possible to include an additional loop allowing for several repetitions of the block of primal updates (steps 2-18) before the block of dual updates (steps 19-24). If this additional loop were iterated to convergence, we would recover an exact augmented Lagrangian method. However, in practice we find that a single iteration of such a loop is the most effective choice (noting that the cost of the primal updates dominates that of the dual updates), and there is no practical motivation for further complicating Algorithm 5 as written. A practical comparison is provided in Appendix B.

Furthermore, inspired by [14], it is natural to consider a dual step size $\tau \in (1, (1+\sqrt{5})/2)$ appearing as a prefactor in front of the dual variable increments of steps 19-24. This possibility is explored in Appendix C, but due to a lack of demonstrated advantage is left out of our main numerical experiments for simplicity.

We pause now to provide some perspectives on the convergence of Algorithm 5. First, it is worth noting by inspection that our iterative algorithm formally retains the correct fixed point at optimality. However, a proof of convergence is beyond the scope of this paper. Indeed, certain features of the current algorithm differ importantly from the conventional augmented Lagrangian method for SDP. In our outer loop, we perform a dual update via gradient ascent on the dual variable for the global PSD constraint. The inner loop provides a gradient for the outer loop, though it is in practice solved only approximately. For the outer loop, we cannot readily leverage the nice properties of proximal point methods as in the case of ALM [33] to establish convergence. Furthermore, as the inner loop is only roughly solved, the resulting gradient is inexact, which introduces further difficulty.

3.3. Exploiting translation-invariance

One of the most expensive step in Algorithm 5 is step 24, where a projection to the positive semidefinite cone is required. We now discuss how translation-invariance of a lattice system can be exploited algorithmically in the solution of the two-marginal relaxation

Algorithm 5 Practical ADMM / projected dual gradient ascent method for (2.4).

Require: $\varepsilon, \mu, \nu > 0,\ X \succeq 0,\ \{\rho_i\},\ \{\tilde{\rho}_{ij}, \Lambda_{ij}, \Lambda^{(1)}_{ij}, \Lambda^{(2)}_{ij}\}_{i<j}$

1: **while** not converged **do**
2: **for** each i **do**
3: $H'_i \leftarrow H_i[X_{ii}]$
4: **end for**
5: **for** each pair $i < j$ **do**
6: $H'_{ij} \leftarrow H_{ij}[X_{ij}]$
7: **end for**
8: **for** each pair $i < j$ **do**
9: $\rho_{ij} \leftarrow \left(\mu I + \nu A_1^* A_1 + \nu A_2^* A_2\right)^{-1} \left(\mu\tilde{\rho}_{ij} + A_1^*\left[\nu\rho_i - \Lambda^{(1)}_{ij}\right] + A_2^*\left[\nu\rho_j - \Lambda^{(2)}_{ij}\right] + \Lambda_{ij} - H'_{ij}\right)$
10: **end for**
11: **for** each pair $i < j$ **do**
12: $\tilde{\rho}_{ij} \leftarrow \Pi_{\succeq 0}\left(\rho_{ij} - \mu^{-1}\Lambda_{ij}\right)$
13: **end for**
14: **for** each i **do**
15: $\rho'_i \leftarrow \frac{1}{\nu(M-1)}\left(\sum_{j>i}\left(\nu A_1[\rho_{ij}] + \Lambda^{(1)}_{ij}\right) + \sum_{j<i}\left(\nu A_2[\rho_{ji}] + \Lambda^{(2)}_{ji}\right) - H'_i\right)$
16: $z \leftarrow m^{-1}(1 - \mathrm{Tr}[\rho'_i])$
17: $\rho_i \leftarrow \rho'_i + zI_m$
18: **end for**
19: **for** each pair $i < j$ **do**
20: $\Lambda_{ij} \leftarrow \Lambda_{ij} + \mu\left(\tilde{\rho}_{ij} - \rho_{ij}\right)$
21: $\Lambda^{(1)}_{ij} \leftarrow \Lambda^{(1)}_{ij} + \nu(A_1[\rho_{ij}] - \rho_i)$
22: $\Lambda^{(2)}_{ij} \leftarrow \Lambda^{(2)}_{ij} + \nu(A_2[\rho_{ij}] - \rho_j)$
23: **end for**
24: $X \leftarrow \Pi_{\succeq 0}\left(X + \varepsilon G[\{\rho_i\}, \{\rho_{ij}\}_{i<j}]\right)$
25: **end while**

(2.4). It is convenient in this section to adopt zero-indexing for the site, i.e., to index the sites by the multi-index $\mathbf{i} = (i_1, \ldots, i_d)$ where d is the lattice dimension and $i_p = 0, \ldots, M-1$. Then we assume translation-invariance in that $\hat{H}_{\mathbf{i}} = \hat{H}_{\mathbf{0}}$ for all i and $\hat{H}_{\mathbf{ij}} = \hat{H}_{0,\mathbf{j-i}}$ for all $i < j$. The symmetries $\rho_{\mathbf{i}} = \rho_{\mathbf{0}}$ and $\rho_{\mathbf{ij}} = \rho_{0,\mathbf{j-i}}$ are in turn guaranteed to be satisfied by some optimizer of (2.4) [25].

Then we can implement Algorithm 5 (whose iterations preserve this symmetry) without any reference to variables besides $\rho_{\mathbf{0}}$ and the $\rho_{\mathbf{0,j}}$. The main challenge is the implementation of step 24 of Algorithm 5. Given $\rho_{\mathbf{0}}$ and the $\rho_{\mathbf{0,j}}$, we can only compute the top row $X'_{\mathbf{0,j}}$ of $X' := X + \varepsilon G[\{\rho_i\}, \{\rho_{ij}\}_{i<j}]$. However, by translation invariance, the rest of X' is determined by the property that $X'_{\mathbf{i,j}} = X'_{\mathbf{0,j-i}}$. (In the case $d = 1$, X' is a *block-circulant* matrix, though a more general term is lacking for the case of arbitrary d.) Via translation invariance, X' is block-diagonalized by the block-discrete Fourier transform. More precisely, one can write X' as $X' = (\mathcal{F} \otimes I)\hat{X}'(\mathcal{F}^* \otimes I)$, where $\mathcal{F}$ indicates the appropriate d-dimensional discrete Fourier transform matrix, I the identity matrix of the appropriate block size, $\otimes$ the Kronecker product, and $\hat{X}'$ a block-diagonal matrix. Hence to compute the projection $\Pi_{\succeq 0}[X']$ we first compute the diagonal block $\hat{X}'_{\mathbf{k}}$ of $\hat{X}'$ as

$$\hat{X}'_{\mathbf{k}} = \frac{1}{\sqrt{M^d}} \sum_{j_1,\ldots,j_d=0}^{M-1} \exp\left(-\imath\frac{2\pi\mathbf{j}\cdot\mathbf{k}}{M}\right) X'_{0,\mathbf{j}}.$$

Note that the blocks $\hat{X}'_{\mathbf{k}}$ (concatenated into a block row) can be viewed as the entrywise discrete Fourier transform of the first block row of X', hence can be computed simultaneously via FFT. Then project $Y_{\mathbf{k}} := \Pi_{\succeq 0}\left[\hat{X}'_{\mathbf{k}}\right]$ for all $\mathbf{k}$, and set

$$X_{\mathbf{0,j}} \leftarrow \frac{1}{\sqrt{M^d}} \sum_{j_1,\ldots,j_d=0}^{M-1} \exp\left(\imath\frac{2\pi\mathbf{j}\cdot\mathbf{k}}{M}\right) Y_{\mathbf{k}}.$$

The final pseudocode for the translation-invariant setting is given in Algorithm 6.

3.4. Discussion of scaling

In Algorithm 5, observe that the for-loops run over $M(M-1)$ pairs $i < j$, and the scaling bottleneck among these loops is the projection $\tilde{\rho}_{ij} \leftarrow \Pi_{\succeq 0}\left(\rho_{ij} - \mu^{-1}\Lambda_{ij}\right)$ occurring in step 12. Since this step requires full diagonalization of a matrix of size $m^2 \times m^2$, for which the cost is $O(m^6)$. Meanwhile, suppose for simplicity that $n_i = m^2$, corresponding to the complete choice of operator collection $\left\{O_{\alpha,i} : \alpha = n_i\right\}$ for each site. Then the size of G and X is $Mm^2 \times Mm^2$. Hence step 24, which involves a complete diagonalization of a matrix of this size, costs $O(m^6M^3)$, dominating the $O(m^6M^2)$ cost of the for-loops. If our sites are in fact supersites, each formed from clusters of L sites in an underlying spin-$\frac{1}{2}$ model, then $m = 2^L$. Therefore the scaling is $O(2^{6L}M^3)$ per iteration, where L is the cluster size and M is the number of clusters.

Meanwhile, in the translation-invariant setting of Algorithm 6, the for-loops run only over $O(M)$ sites, so—neglecting the update for X—the asymptotic cost per iteration is $O(m^6M)$. Meanwhile, the construction of the $\hat{X}'_{\mathbf{k}}$ in terms of the $X_{0,\mathbf{j}}$ can be achieved in time $O(m^4M\log M)$ via FFT. (Note that we simply treat the lattice dimension d as constant.) The cost of each projection of step

Algorithm 6 Translation-invariant ADMM / projected dual gradient ascent method for (2.4).

Require: $\varepsilon, \mu, \nu > 0$, $(X_{\mathbf{0},\mathbf{j}})$, $\rho_{\mathbf{0}}$, $\{\tilde{\rho}_{\mathbf{0},\mathbf{j}}, \Lambda_{\mathbf{0},\mathbf{j}}, \Lambda^{(1)}_{\mathbf{0},\mathbf{j}}, \Lambda^{(2)}_{\mathbf{0},\mathbf{j}}\}_{\mathbf{j}\neq\mathbf{0}}$

1: **while** not converged **do**
2: $H'_{\mathbf{0}} \leftarrow H_{\mathbf{0}}[X_{\mathbf{0},\mathbf{0}}]$
3: **for** each $\mathbf{j} \neq \mathbf{0}$ **do**
4: $H'_{\mathbf{0},\mathbf{j}} \leftarrow H_{\mathbf{0},\mathbf{j}}[X_{\mathbf{0},\mathbf{j}}]$
5: **end for**
6: **for** each $\mathbf{j} \neq \mathbf{0}$ **do**
7: $\rho_{\mathbf{0},\mathbf{j}} \leftarrow \left(\mu I + \nu A_1^* A_1 + \nu A_2^* A_2\right)^{-1} \left(\mu\tilde{\rho}_{\mathbf{0},\mathbf{j}} + A_1^*\left[\nu\rho_{\mathbf{0}} - \Lambda^{(1)}_{\mathbf{0},\mathbf{j}}\right] + A_2^*\left[\nu\rho_{\mathbf{0}} - \Lambda^{(2)}_{\mathbf{0},\mathbf{j}}\right] + \Lambda_{\mathbf{0},\mathbf{j}} - H'_{\mathbf{0},\mathbf{j}}\right)$
8: **end for**
9: **for** each $\mathbf{j} \neq \mathbf{0}$ **do**
10: $\tilde{\rho}_{\mathbf{0},\mathbf{j}} \leftarrow \Pi_{\succeq 0}\left(\rho_{\mathbf{0},\mathbf{j}} - \mu^{-1}\Lambda_{\mathbf{0},\mathbf{j}}\right)$
11: **end for**
12: $\rho'_{\mathbf{0}} \leftarrow \frac{1}{\nu(M-1)}\left(\sum_{\mathbf{j}\neq\mathbf{0}}\left(\nu A_1[\rho_{\mathbf{0},\mathbf{j}}] + \Lambda^{(1)}_{\mathbf{0},\mathbf{j}}\right) - H'_{\mathbf{0}}\right)$
13: $z \leftarrow m^{-1}(1 - \mathrm{Tr}[\rho'_{\mathbf{0}}])$
14: $\rho_{\mathbf{0}} \leftarrow \rho'_{\mathbf{0}} + zI_m$
15: **for** each $\mathbf{j} \neq \mathbf{0}$ **do**
16: $\Lambda_{\mathbf{0},\mathbf{j}} \leftarrow \Lambda_{\mathbf{0},\mathbf{j}} + \mu\left(\tilde{\rho}_{\mathbf{0},\mathbf{j}} - \rho_{\mathbf{0},\mathbf{j}}\right)$
17: $\Lambda^{(1)}_{\mathbf{0},\mathbf{j}} \leftarrow \Lambda^{(1)}_{\mathbf{0},\mathbf{j}} + \nu(A_1[\rho_{\mathbf{0},\mathbf{j}}] - \rho_{\mathbf{0}})$
18: $\Lambda^{(2)}_{\mathbf{0},\mathbf{j}} \leftarrow \Lambda^{(2)}_{\mathbf{0},\mathbf{j}} + \nu(A_2[\rho_{\mathbf{0},\mathbf{j}}] - \rho_{\mathbf{0}})$
19: **end for**
20: $X'_{\mathbf{0},\mathbf{0}} \leftarrow X_{\mathbf{0},\mathbf{0}} + \varepsilon G_{\mathbf{0},\mathbf{0}}\left[\rho_{\mathbf{0}}\right]$
21: **for** each $\mathbf{j} \neq \mathbf{0}$ **do**
22: $X'_{\mathbf{0},\mathbf{j}} \leftarrow X_{\mathbf{0},\mathbf{j}} + \varepsilon G_{\mathbf{0},\mathbf{j}}\left[\rho_{\mathbf{0},\mathbf{j}}\right]$
23: **end for**
24: **for** each $\mathbf{k}$ **do**
25: $\hat{X}'_{\mathbf{k}} = \frac{1}{\sqrt{M^d}}\sum_{\mathbf{j}} \exp\left(-\imath\frac{2\pi\mathbf{j}\cdot\mathbf{k}}{M}\right) X'_{\mathbf{0},\mathbf{j}}$
26: $Y_{\mathbf{k}} \leftarrow \Pi_{\succeq 0}\left[\hat{X}'_{\mathbf{k}}\right]$
27: **end for**
28: **for** each $\mathbf{j}$ **do**
29: $X_{\mathbf{0},\mathbf{j}} \leftarrow \frac{1}{\sqrt{M^d}}\sum_{\mathbf{j}} \exp\left(\imath\frac{2\pi\mathbf{j}\cdot\mathbf{k}}{M}\right) Y_{\mathbf{k}}$
30: **end for**
31: **end while**

26 is $O(m^6)$ via full diagonalization, and forming the $X_{\mathbf{0},\mathbf{j}}$ in terms of the $Y_{\mathbf{k}}$ also costs $O(dm^4 M \log M)$ in total via FFT. Hence the cost of updating X (i.e., steps 20 through 29) is $O(m^4 M \log M + m^6 M)$. Hence the total cost per iteration of Algorithm 6, under the assumption that the sites are supersites each composed of L spin-$\frac{1}{2}$ sites, is $O(2^{4L} M \log M + 2^{6L} M)$, where L is the cluster size and M is the number of clusters.

The exponential scaling in the cluster size is unavoidable in our formulation due to the exact treatment of reduced density operators on the clusters (i.e., the cluster marginals). In this work we consider clusters of size no larger than size $L = 4$. Future work will investigate the possibility of treating larger clusters by introducing further relaxation and/or compression of the optimization variables to avoid exponential scaling in the cluster size.

As a final comment, observe that every for-loop in Algorithms 5 and 6 can be run fully in parallel.

4. Numerical experiments

The numerical experiments were implemented in MATLAB following Algorithm 6. (We shall consider only translation-invariant Hamiltonians.) We present results for the transverse-field Ising (TFI) model (2.2), the anti-ferromagnetic Heisenberg (AFH) model (2.3), the spinless fermion (SF) model (A.1), and the long-range spinless fermion (LRSF) model (A.2). Note that Algorithm 6 can be applied in the fermionic case *mutatis mutandi* to the problem (A.3). Throughout we fix the value of the algorithmic parameters to be $\mu = \nu = 10$, $\varepsilon = 2$ (i.e., we do not tune them specifically to different problems). The dual variables $\{\Lambda_{ij}, \Lambda^{(1)}_{ij}, \Lambda^{(2)}_{ij}\}_{i\neq j}$ are all initialized to be zero, and the primal density operator variables $\{\rho_i\}$, $\{\tilde{\rho}_{ij}\}_{i\neq j}$ are all initialized as multiples of the identity with unit trace. X is initialized as the identity. We run Algorithm 6 for 10,000 iterations. (The convergence behavior will be studied in detail below.)

First we consider the TFI model on a periodic 20×1 lattice, which is small enough to be solved by exact diagonalization of (2.2). We benchmark the per-site energy error of the two-marginal relaxation with clusters of size 1×1, 2×1, and 4×1. (In these cases the semidefinite matrix variables ρ_{ij} are *each* of size 4, 16, and 256, respectively; refer to Section 3.4 for further discussion of scaling.). The results are shown in Fig. 4.1. Observe that the approximations yield lower bounds for the energy as the theory requires, and these lower bounds become tighter as the cluster size is increased. In the same figure we also consider the TFI model on a periodic 4×4 lattice. Here we benchmark the energy error of the two-marginal relaxation with clusters of size 1×1, 2×1, and 2×2.

We perform completely analogous experiments for the AFH model with similar conclusions. The results are shown in Tables 1 and 2.

We also benchmark the SF and LRSF models on a 20×1 periodic lattice, with results pictured in Fig. 4.2. Note that the fermionic relaxations are exact for $U = 0$, as guaranteed in [25].

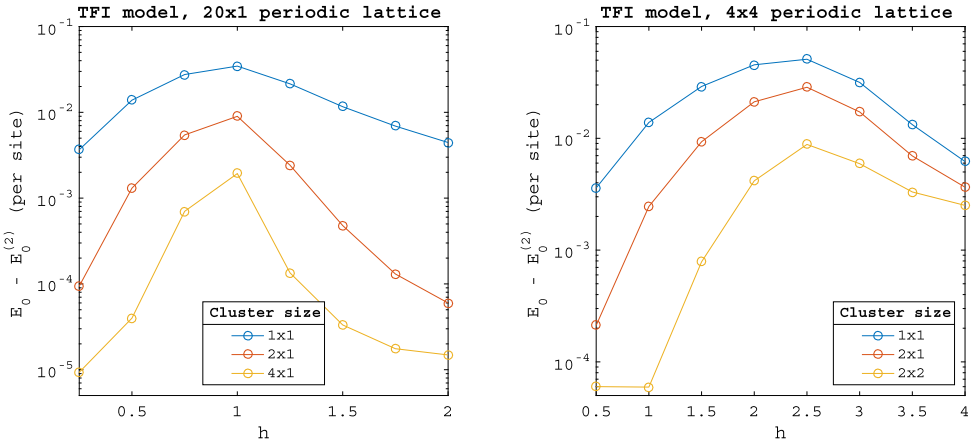

Fig. 4.1. Relaxation error per site for the TFI model on 20×1 and 4×4 periodic lattices, pictured left and right respectively. Several cluster sizes are considered according to the legends. Note that the relaxation is exact at $h = 0$ (not pictured). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Table 1
Relaxation error per site for the AFH model on a 20×1 periodic lattice for various cluster sizes.

1×1 clusters	2×1 clusters	4×1 clusters
0.5383	0.0521	0.0034

Table 2
Relaxation error per site for the AFH model on a 4×4 periodic lattice for various cluster sizes.

1×1 clusters	2×1 clusters	2×2 clusters
0.6634	0.1851	0.0034

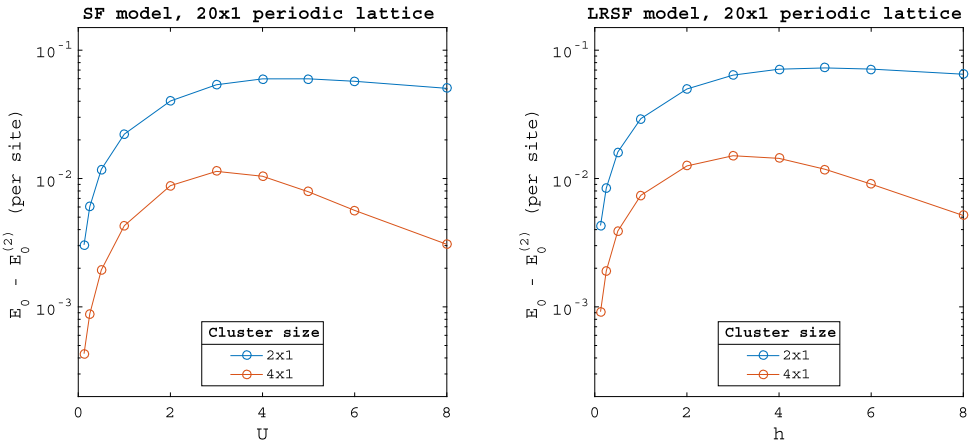

Fig. 4.2. Relaxation error for the (short-range) spinless fermion (A.1) and long-range spinless fermion (A.2) models on a 20×1 periodic lattice, pictured left and right respectively. Several cluster sizes are considered according to the legends. Note that the relaxation is exact at $U = 0$ (not pictured).

Next we consider the TFI model on a periodic 100×1 lattice for $h = 0.5, 1, 1.5$. This problem is too large to solve by exact diagonalization. We report the relaxation energy for several cluster sizes in Table 3.

We track convergence behavior of Algorithm 6 on these same problems. We use two different quantities to track convergence. The first is the per-site primal objective (i.e., energy) change between subsequent iterations. The second is the per-cluster feasibility error for the equality constraints, defined as

Table 3
Relaxation energy per site for the TFI model with $h = 0.5, 1, 1.5$ on a 100×1 periodic lattice for various cluster sizes.

	1×1 clusters	2×1 clusters	4×1 clusters
$h = 0.5$	-1.0763	-1.0648	-1.0636
$h = 1$	-1.3084	-1.2829	-1.2761
$h = 1.5$	-1.6835	-1.6724	-1.6720

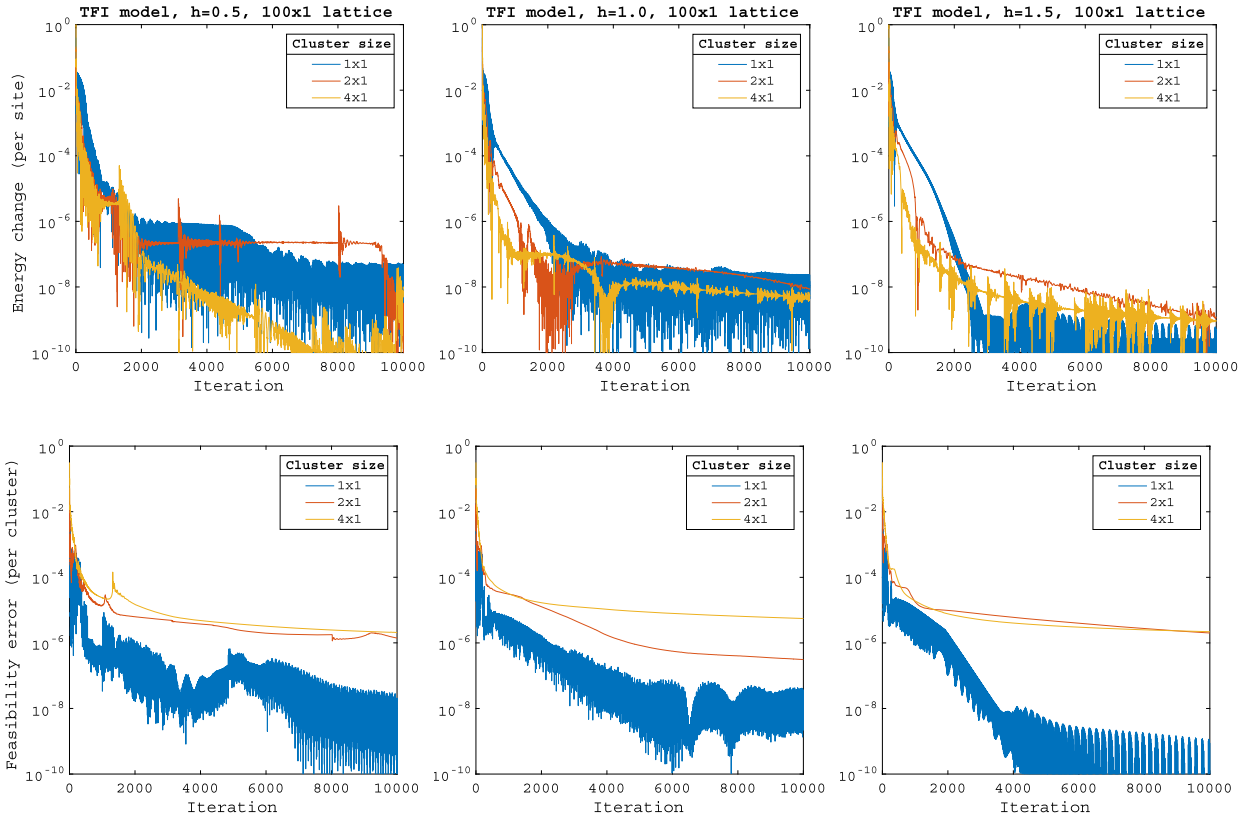

Fig. 4.3. Per-site iteration-over-iteration energy change for 100×1 periodic TFI model with $h = 0.5, 1, 1.5$ and several cluster sizes (top row). Per-cluster feasibility error (4.1) for same problems (bottom row).

$$\sqrt{\frac{1}{M-1}\sum_{\mathbf{j}\neq 0}\left(\|A_1[\rho_{\mathbf{0},\mathbf{j}}]-\rho_{\mathbf{0}}\|_{\mathrm{F}}^2+\|A_2[\rho_{\mathbf{0},\mathbf{j}}]-\rho_{\mathbf{0}}\|_{\mathrm{F}}^2+\|\rho_{\mathbf{0},\mathbf{j}}-\tilde{\rho}_{\mathbf{0},\mathbf{j}}\|_{\mathrm{F}}^2\right)}. \tag{4.1}$$

We plot these quantities as functions of the iteration count in Fig. 4.3. We also recommend that these quantities be used to define a practical stopping criterion for the optimizer, i.e., to stop when both converge within some specified numerical tolerance. In Appendix D we present additional experiments demonstrating the satisfaction of the global semidefinite constraints.

It is possible that tuning the parameters μ, ν, ε to a specific problem and specific choice of clusters could yield smoother convergence profiles. However, even using our fixed choice for all problems, we achieve convergence of the per-site energy within 10^{-6} (which is dominated by the relaxation error itself) in a number of iterations that does not seem to grow with the cluster size. Though the asymptotic convergence of ADMM-type methods can be slow, for practical purposes we only need the convergence accuracy to be dominated by the relaxation error of the relaxed problem. If convergence to very high accuracy is required, alternative approaches such as inexact ALM [33] might be more appropriate.

Then we fix clusters of size 2×1 and vary the system size of the TFI model with $h = 0.5, 1, 1.5$, to investigate the effect of system size on convergence. The results are shown in Fig. 4.4. The system size does not appear to have any obvious detrimental effect on the convergence rate.

To conclude, we report the relaxation energies obtained from these last experiments in Table 4. We observe that the relaxation energy approaches a limiting value in this thermodynamic limit, i.e., limit of infinite volume.

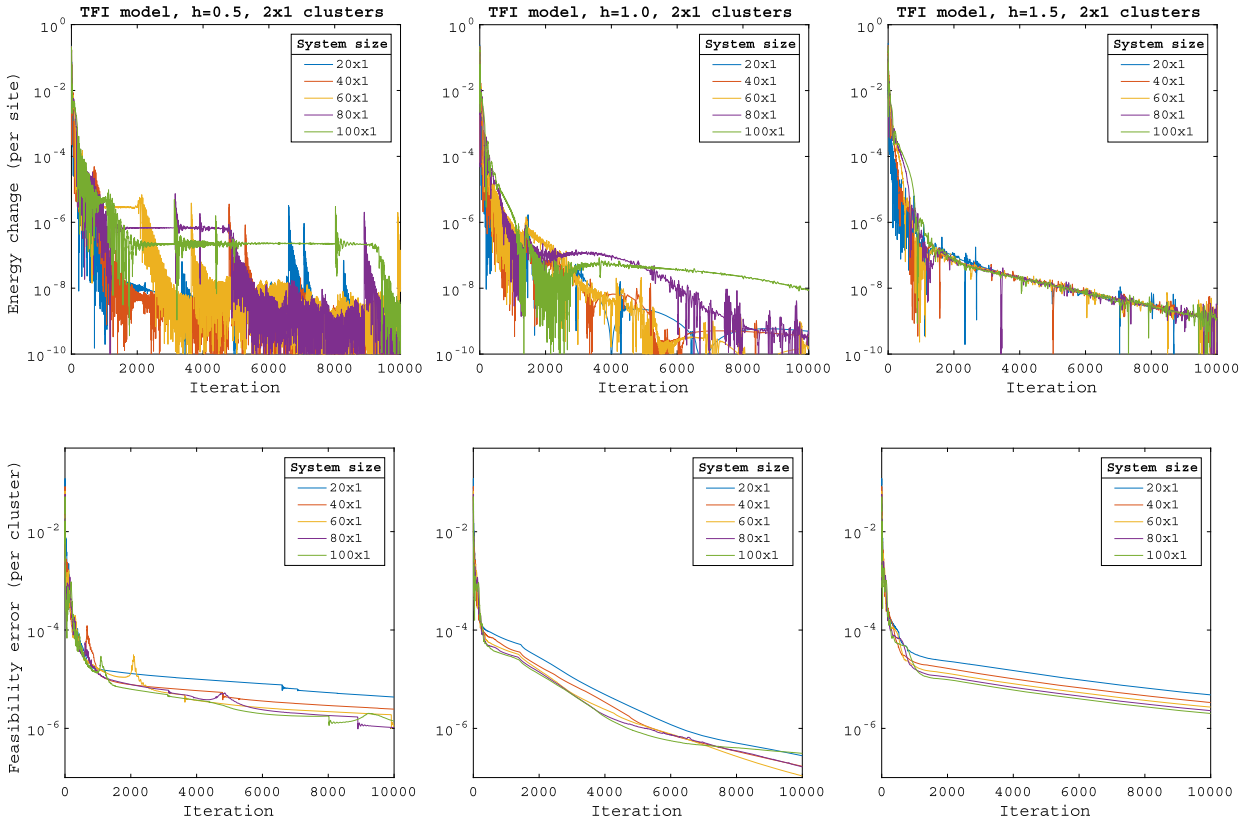

Fig. 4.4. Per-site iteration-over-iteration energy change for periodic TFI models of different sizes with $h = 0.5, 1, 1.5$ and fixed 2×1 cluster size (top row). Per-cluster feasibility error (4.1) for same problems (bottom row).

Table 4
Relaxation energy per site for TFI models of several sizes with $h = 0.5, 1, 1.5$ and clusters of fixed size 2×1.

	20×1 lattice	40×1 lattice	60×1 lattice	80×1 lattice	100×1 lattice
$h = 0.5$	−1.064851	−1.064795	−1.064786	−1.064779	−1.064776
$h = 1$	−1.283534	−1.283083	−1.283003	−1.282975	−1.282949
$h = 1.5$	−1.672407	−1.672394	−1.672393	−1.672393	−1.672394

CRediT authorship contribution statement

Yuehaw Khoo: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Michael Lindsey:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare no competing interests.

Data availability

Data will be made available on request.

Acknowledgements

This work was partially supported by the National Science Foundation under Award No. 1903031 (M.L.), the Applied Mathematics Program of the US Department of Energy (DOE) Office of Advanced Scientific Computing Research under contract number DE-AC02-05CH11231 (M.L.), the National Science Foundation under Award No. 2111563 (Y.K.), and the Department of Energy under Award No. DE-SC0022232 (Y.K.). We thank Lin Lin for helpful discussions.

Appendix A. Fermions

In this appendix we present the relevant background on fermionic many-body systems, following [25].

A.1. Background

Fermionic many-body problems in second quantization are specified in terms of the creation operators $a_1^\dagger, \dots, a_M^\dagger$ and their Hermitian adjoints, the annihilation operators a_i, which can be viewed as operators on a vector space of dimension 2^M called the Fock space $\mathcal{F}$. The key properties of these operators are the canonical anticommutation relations

$$\{a_i, a_j^\dagger\} = \delta_{ij}, \quad \{a_i, a_j\} = \{a_i^\dagger, a_j^\dagger\} = 0,$$

where $\{\cdot, \cdot\}$ denotes the anticommutator. In terms of these operators we also define the number operators $\hat{n}_i := a_i^\dagger a_i$, and the total number operator by $\hat{N} := \sum_{i=1}^M \hat{n}_i$.

One can identify the Fock space with a quantum spin-$\frac{1}{2}$ state space, i.e., identify $\mathcal{F} \simeq \bigotimes^M \mathbb{C}^2 \simeq \mathbb{C}^{2^M}$, via the correspondence

$$a_i^\dagger \rightsquigarrow \underbrace{\sigma^z \otimes \cdots \otimes \sigma^z}_{i-1 \text{ factors}} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes I_2 \otimes \cdots \otimes I_2,$$

known as the Jordan-Wigner transformation (JWT). This transformation is unnatural in the sense that it depends on the ordering of the states. More precisely, permuting the states before the JWT is not equivalent to permuting the tensor factors after the JWT. Moreover, a fermionic operator such as $a_i^\dagger a_j$ involving only two sites corresponds in general to a quantum spin operator involving potentially many more sites; hence the pairwise fermionic Hamiltonians that we consider below cannot be viewed in general as pairwise spin-$\frac{1}{2}$ Hamiltonians.

Next we define the notion of a pairwise fermionic Hamiltonian, and then we provide some examples. Unfortunately we cannot simply treat clusters of sites as 'supersites' without breaking the fermionic structure, so we approach the cluster framework directly, writing $\{1, \dots, M\}$ as a disjoint union of clusters $\bigcup_{\gamma=1}^{N_c} C_\gamma$ specified by the user. We let

$$\mathcal{A} := \langle 1, a_1, \dots, a_M, a_1^\dagger, \dots, a_M^\dagger \rangle$$

denote the star-algebra generated by the creation and annihilation operators subject to the canonical anticommutation relations $\{a_i, a_j^\dagger\} = \delta_{ij}$, $\{a_i, a_j\} = \{a_i^\dagger, a_j^\dagger\} = 0$, and similarly, we let

$$\mathcal{A}_C := \left\langle \{1\} \cup \{a_i, a_i^\dagger \,:\, i \in C\} \right\rangle$$

denote the subalgebra corresponding to a subset $C \subset \{1, \dots, M\}$. Then we consider pairwise Hamiltonians $\hat{H} \in \mathcal{A}$ of the form

$$\hat{H} = \sum_\gamma \hat{H}_\gamma + \sum_{\gamma<\delta} \hat{H}_{\gamma\delta},$$

where $\hat{H}_\gamma \in \mathcal{A}_{C_\gamma}$ and $\hat{H}_{\gamma\delta} \in \mathcal{A}_{C_\gamma \cup C_\delta}$ are Hermitian operators. We are interested in the ground-state energy

$$E_0 = \inf \left\{ \langle \psi | \hat{H} | \psi \rangle \,:\, |\psi\rangle \in \mathcal{F}, \langle \psi | \psi \rangle = 1 \right\}.$$

For particle-number conserving Hamiltonians (i.e., Hamiltonians that commute with $\hat{N}$), one may also consider the N-particle ground-state energy defined as

$$E_0(N) = \inf \left\{ \langle \psi | \hat{H} | \psi \rangle \,:\, |\psi\rangle \in \mathcal{F}, \langle \psi | \psi \rangle = 1, \langle \psi | \hat{N} | \psi \rangle = N \right\},$$

though note that this is formally equivalent to the unconstrained ground-state energy E_0 after subtracting $\mu \hat{N}$ from the Hamiltonian, where the Lagrange multiplier μ is called the chemical potential.

A.2. Examples

In this work we shall consider the half-filled lattice of spinless fermions [43] specified by the Hamiltonian

$$\hat{H} = \sum_{i \sim j} \left[-a_i^\dagger a_j - a_j^\dagger a_i + U \left(\hat{n}_i - \frac{1}{2} \right) \left(\hat{n}_j - \frac{1}{2} \right) \right], \tag{A.1}$$

where U is a scalar parameter (the 'interaction strength') and the notion of adjacency $i \sim j$ is defined relative to a graph (usually a rectangular lattice) on the sites $\{1, \dots, M\}$. This operator is pairwise relative to any cluster decomposition. One can also consider an analogous model with long-range Coulomb interaction

$$\hat{H} = \sum_{i \sim j} \left[-a_i^\dagger a_j - a_j^\dagger a_i \right] + U \sum_{i \neq j} \frac{1}{d(i,j)} \left(\hat{n}_i - \frac{1}{2} \right) \left(\hat{n}_j - \frac{1}{2} \right), \tag{A.2}$$

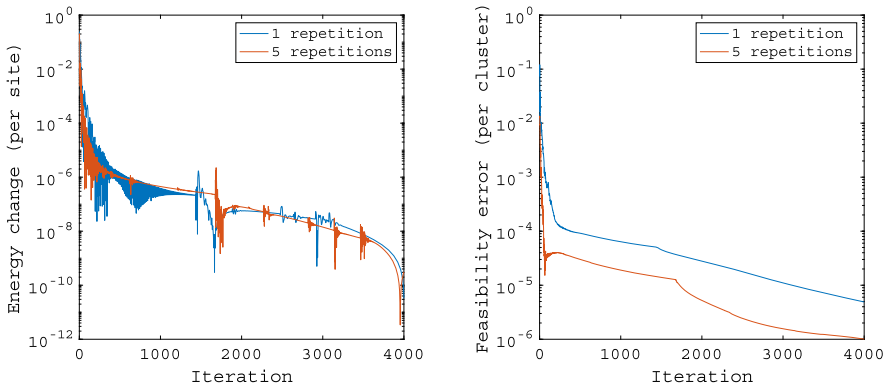

Fig. B.1. Left: per-site iteration-over-iteration energy change for 20×1 periodic TFI model with $h = 1$ and 2×1 clusters. Right: per-cluster feasibility error (4.1) for same problem. Note that the computational cost of each iteration within the red curves is about 5 times greater than that of each iteration within the blue curves.

where $d(i, j)$ is the Euclidean distance between sites i and j on the lattice.

As outlined in [25], one can also consider spinful systems such as the Hubbard model [18], as well as quantum chemistry Hamiltonians arising from electronic structure problems after a suitable choice of basis [38], e.g., the recently developed discontinuous Galerkin basis [8].

A.3. The two-marginal relaxation

In order to realize the two-marginal relaxation as a concrete semidefinite program, it is necessary to choose a JWT *for each pair of clusters* as follows. (Note that there will be no need to consider any global JWT.) For $1 \le \gamma \le N_c$, let $L_\gamma := |C_\gamma|$, and let $\kappa_\gamma : C_\gamma \to \{1, \dots, L_\gamma\}$ be a bijection specifying an ordering for the sites in the γ-th cluster. For $1 \le \gamma < \delta \le N_c$, let $L_{\gamma\delta} := |C_\gamma| + |C_\delta|$, and let $\kappa_{\gamma\delta} : C_\gamma \cup C_\delta \to \{1, \dots, L_{\gamma\delta}\}$ be the bijection specifying an ordering for the sites in the (γ, δ)-th pair of clusters, uniquely specified by the conditions that $\kappa_{\gamma\delta}|_{C_\gamma} = \kappa_\gamma$, $\kappa_{\gamma\delta}|_{C_\delta} = \kappa_\delta$, and $\kappa_{\gamma\delta}(C_\gamma) < \kappa_{\gamma\delta}(C_\delta)$. These orderings fix algebra isomorphisms $\mathcal{J}_\gamma : \mathcal{A}_{C_\gamma} \to \mathrm{End}\left(\bigotimes_{i=1}^{L_\gamma} \mathbb{C}^2\right)$ and $\mathcal{J}_{\gamma\delta} : \mathcal{A}_{C_\gamma \cup C_\delta} \to \mathrm{End}\left(\bigotimes_{i=1}^{L_{\gamma\delta}} \mathbb{C}^2\right)$ via the appropriate JWTs. Then the two-marginal relaxation is given concretely by

$$\begin{aligned} \underset{\{\rho_\gamma\}, \{\rho_{\gamma\delta}\}_{\gamma<\delta}}{\text{minimize}} \quad & \sum_\gamma \mathrm{Tr}\left[\mathcal{J}_\gamma\left(\hat{H}_\gamma\right)\rho_\gamma\right] + \sum_{\gamma<\delta} \mathrm{Tr}\left[\mathcal{J}_{\gamma\delta}\left(\hat{H}_{\gamma\delta}\right)\rho_{\gamma\delta}\right], \\ \text{subject to} \quad & \rho_{\gamma\delta} \succeq 0, \quad 1 \le \gamma < \delta \le N_c, \\ & \rho_\gamma = \mathrm{Tr}_{\kappa_{\gamma\delta}(C_\delta)}[\rho_{\gamma\delta}], \quad \rho_\delta = \mathrm{Tr}_{\kappa_{\gamma\delta}(C_\gamma)}[\rho_{\gamma\delta}], \quad 1 \le \gamma < \delta \le N_c, \\ & \mathrm{Tr}[\rho_\gamma] = 1, \quad \gamma = 1, \dots, N_c, \\ & G\left[\{\rho_\gamma\}, \{\rho_{\gamma\delta}\}_{\gamma<\delta}\right] \succeq 0, \end{aligned} \tag{A.3}$$

where G is specified blockwise subordinate to a collection $\{\hat{A}_{\gamma,\alpha} : \alpha = 1, \dots, n_\gamma\} \subset \mathcal{A}_{C_\gamma}$ of operators on each cluster via

$$\left(G_{\gamma\delta}\right)_{\alpha\beta} = \begin{cases} \mathrm{Tr}\left(\left[\mathcal{J}_\gamma\left(\hat{A}_{\gamma,\alpha}\right)\right]^\dagger \left[\mathcal{J}_\gamma\left(\hat{A}_{\gamma,\beta}\right)\right]\rho_\gamma\right), & \gamma = \delta \\ \mathrm{Tr}\left(\left[\mathcal{J}_{\gamma\delta}\left(\hat{A}_{\gamma,\alpha}\right)\right]^\dagger \left[\mathcal{J}_{\gamma\delta}\left(\hat{A}_{\delta,\beta}\right)\right]\rho_{\gamma\delta}\right), & \gamma \ne \delta. \end{cases}$$

Hence after the appropriate Jordan-Wignerized operators are formed, (A.3) is of identical form to (2.4).

Appendix B. Iterated primal update

In this section we explore the impact of repeating the block of primal updates within Algorithm 5 before each block of dual updates, as discussed in Section 3.2. We otherwise retain all the parameter choices used throughout Section 4. For illustration we consider the transverse-field Ising (TFI) model (2.2) on a 20×1 periodic lattice with $h = 1$ and 2×1 clusters. The convergence of the iteration-over-iteration energy change and the feasibility error (4.1) are plotted in Fig. B.1 for an experiment in which the primal block is repeated once per dual update (default) and one in which it is repeated 5 times. Note that the per-iteration cost of the latter is roughly 5 times greater than the former, so Fig. B.1 clearly suggests that the default strategy is more efficient.

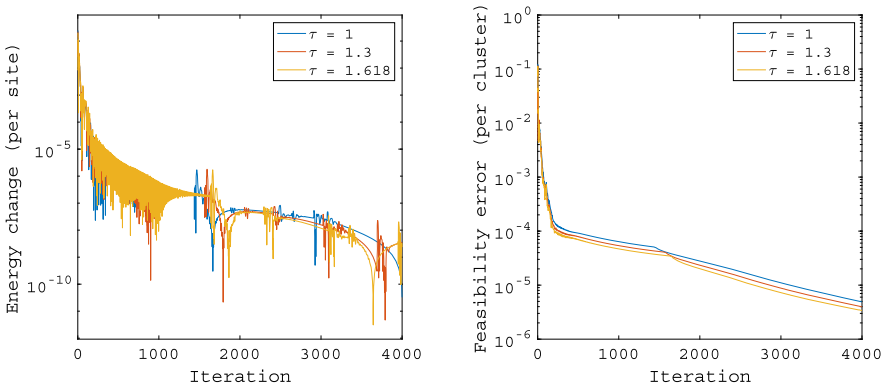

Fig. C.1. Left: per-site iteration-over-iteration energy change for 20×1 periodic TFI model with $h = 1$ and 2×1 clusters. Right: per-cluster feasibility error (4.1) for same problem. Curves are plotted for several values of the dual step size parameter τ.

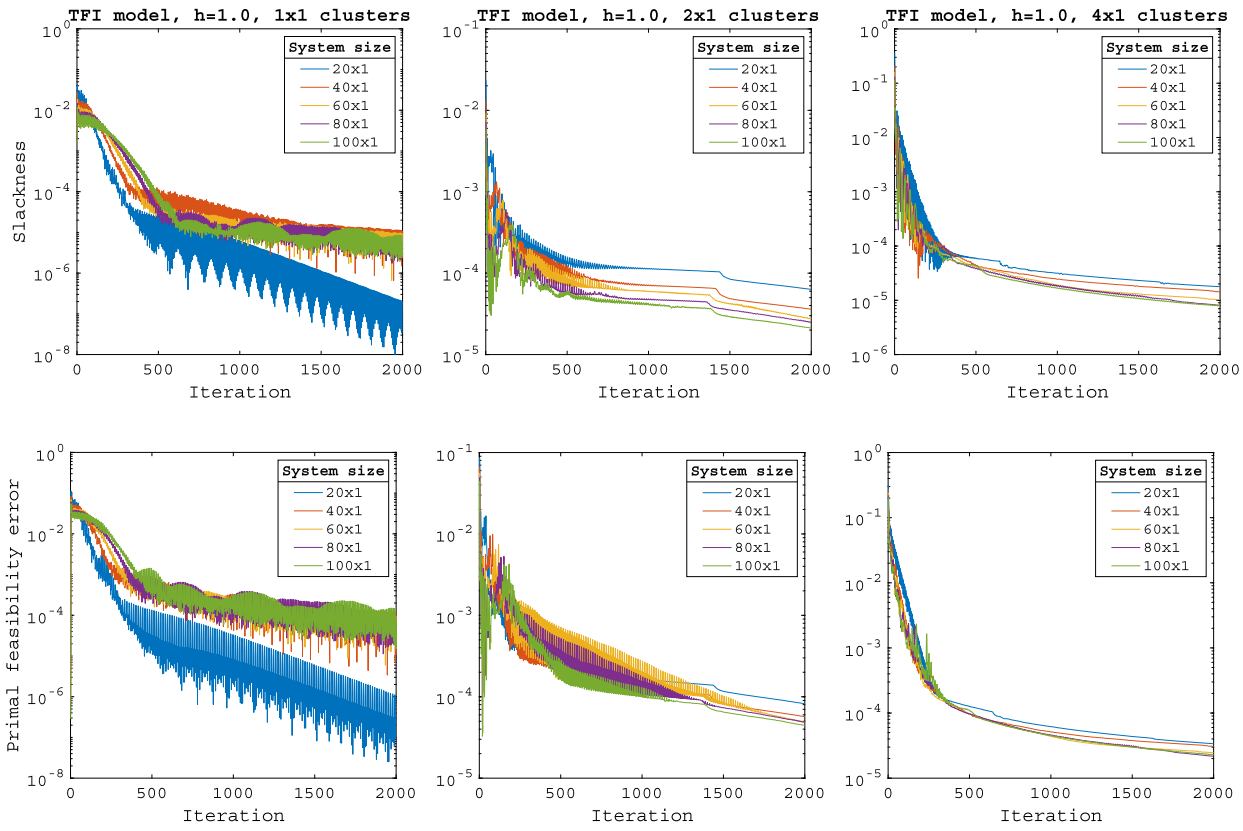

Fig. D.1. Slackness (D.2) and primal feasibility error (D.1), at top and bottom, respectively, over the course of optimization for the TFI model with $h = 1.0$. Left, middle, and right panels correspond to clusters of size 1, 2, and 4 respectively.

Appendix C. Dual step size

In this section we explore the impact of a nontrivial dual step size τ, as discussed in Section 3.2. We otherwise retain all the parameter choices used throughout Section 4. For illustration we consider again the transverse-field Ising (TFI) model (2.2) on a 20×1 periodic lattice with $h = 1$ and 2×1 clusters. The convergence of the iteration-over-iteration energy change and the feasibility error (4.1) are plotted in Fig. C.1 for the choices $\tau = 1$, $\tau = 1.3$, and $\tau = (1 + \sqrt{5})/2 \approx 1.618$. The performance is similar in all cases, so we adopt $\tau = 1$ by default for simplicity.

Appendix D. Checking the global semidefinite constraints

In this section we monitor the satisfaction of the global semidefinite constraints. Note that dual feasibility (i.e., $X \succeq 0$) is satisfied automatically by projection. Hence we are interested in the failure primal feasibility (i.e., $G = G[\{\rho_i\}, \{\rho_{ij}\}_{i<j}] \succeq 0$).

In the translation-invariant case we can define $G_{\mathbf{k}}$ in terms of G just as $X_{\mathbf{k}}$ is defined in terms of X in Algorithm 6, and primal feasibility reduces to the condition that $G_{\mathbf{k}} \succeq 0$ for each mode $\mathbf{k}$. Then we measure the primal feasibility error as

$$\max_{\mathbf{k}} \frac{\|G_{\mathbf{k}} - \Pi_{\succeq 0}(G_{\mathbf{k}})\|_{\mathrm{F}}}{1 + \|G_{\mathbf{k}}\|_{\mathrm{F}}} \tag{D.1}$$

We also monitor complementary slackness, i.e., the condition that $\mathrm{Tr}[XG] = 0$. Under translation invariance, this reduces to the condition that $\mathrm{Tr}[X_{\mathbf{k}} G_{\mathbf{k}}] = 0$ for all $\mathbf{k}$. We measure the deviation as

$$\max_{\mathbf{k}} \left| \frac{\mathrm{Tr}[X_{\mathbf{k}} G_{\mathbf{k}}]}{1 + \|X_{\mathbf{k}}\|_{\mathrm{F}} + \|G_{\mathbf{k}}\|_{\mathrm{F}}} \right| \tag{D.2}$$

In Fig. D.1 we plot the convergence of both of these measures over the course of optimization for the TFI model with $h = 1.0$, comparing several system sizes and trajectories, yielding errors on the order of 10^{-4} across all experiments after around 1000 iterations, after which additional accuracy is slower to attain. Note that we use the same optimization hyperparameters as in the main text and do not tune to the individual problems.

References

[1] J.S. Anderson, M. Nakata, R. Igarashi, K. Fujisawa, M. Yamashita, The second-order reduced density matrix method and the two-dimensional Hubbard model, Comput. Theor. Chem. 1003 (2013) 22–27.
[2] D.P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods, Academic Press, 2014.
[3] S. Boyd, N. Parikh, E. Chu, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Now Publishers Inc, 2011.
[4] E. Cances, G. Stoltz, M. Lewin, The electronic ground-state energy problem: a new reduced density matrix approach, J. Chem. Phys. 125 (2006) 064101.
[5] G. Carleo, M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science 355 (2017) 602–606.
[6] Y. Chen, Y. Khoo, M. Lindsey, Multiscale semidefinite programming approach to positioning problems with pairwise structure, arXiv:2012.10046.
[7] A.E. DePrince, D.A. Mazziotti, Exploiting the spatial locality of electron correlation within the parametric two-electron reduced-density-matrix method, J. Chem. Phys. 132 (2010) 034110.
[8] F. Faulstich, X. Wu, L. Lin, Discontinuous Galerkin method with Voronoi partitioning for quantum simulation of chemistry, arXiv:2011.00367.
[9] A.J. Ferris, D. Poulin, Algorithms for the Markov entropy decomposition, Phys. Rev. B 87 (2013) 205126.
[10] E. Fertitta, G. Booth, Rigorous wave function embedding with dynamical fluctuations, Phys. Rev. B 98 (2018) 235132.
[11] E. Fertitta, G. Booth, Energy-weighted density matrix embedding of open correlated chemical fragments, J. Chem. Phys. 151 (2019) 014115.
[12] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, Comput. Math. Appl. 2 (1976) 17–40.
[13] A. Georges, G. Kotliar, W. Krauth, M.J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, Rev. Mod. Phys. 68 (1996) 13.
[14] R. Glowinski, Lectures on Numerical Methods for Non-Linear Variational Problems, Springer, 1984.
[15] R. Glowinski, A. Marroco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires, ESAIM: Math. Model. Numer. Anal. (Modél. Math. Anal. Numér.) 9 (1975) 41–76.
[16] A. Haim, R. Kueng, G. Refael, Variational-correlations approach to quantum many-body problems, arXiv:2001.06510.
[17] M.R. Hestenes, Multiplier and gradient methods, J. Optim. Theory Appl. 4 (1969) 303–320.
[18] J. Hubbard, Electron correlations in narrow energy bands, Proc. R. Soc. Lond. 276 (1963) 1375.
[19] Y. Khoo, L. Lin, M. Lindsey, L. Ying, Semidefinite relaxation of multi-marginal optimal transport for strictly correlated electrons in second quantization, SIAM J. Sci. Comput. 42 (2020) B1462–B1489.
[20] G. Knizia, G. Chan, Density matrix embedding: a simple alternative to dynamical mean-field theory, Phys. Rev. Lett. 109 (2012) 186404.
[21] G. Knizia, G.K.-L. Chan, Density matrix embedding: a strong-coupling quantum embedding theory, J. Chem. Theory Comput. 9 (2013) 1428–1432.
[22] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, C.A. Marianetti, Electronic structure calculations with dynamical mean-field theory, Rev. Mod. Phys. 78 (2006) 865.
[23] M.S. Leifer, D. Poulin, Quantum graphical models and belief propagation, Ann. Phys. 323 (2008) 1899.
[24] Y. Li, Z. Wen, C. Yang, Y.-x. Yuan, A semismooth Newton method for semidefinite programs and its applications in electronic structure calculations, SIAM J. Sci. Comput. 40 (2018) A4131–A4157.
[25] L. Lin, M. Lindsey, Variational embedding for quantum many-body problems, Commun. Pure Appl. Math. 75 (9) (2022) 2033–2068.
[26] D. Mazziotti, Realization of quantum chemistry without wave functions through first-order semidefinite programming, Phys. Rev. Lett. 93 (2004) 213001.
[27] D. Mazziotti, Structure of fermionic density matrices: complete N-representability conditions, Phys. Rev. Lett. 108 (2012) 263002.
[28] D.A. Mazziotti, Contracted Schrödinger equation: determining quantum energies and two-particle density matrices without wave functions, Phys. Rev. A 57 (1998) 4219.
[29] M. Nakata, H. Nakatsuji, M. Ehara, M. Fukuda, K. Nakata, K. Fujisawa, Variational calculations of fermion second-order reduced density matrices by semidefinite programming algorithm, J. Chem. Phys. 114 (2001) 8282–8292.
[30] D. Poulin, M.B. Hastings, Markov entropy decomposition: a variational dual for quantum belief propagation, Phys. Rev. Lett. 106 (2011) 080403.
[31] M.J. Powell, A method for nonlinear constraints in minimization problems, Optimization (1969) 283–298.
[32] R.T. Rockafellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, SIAM J. Control 12 (1974) 268–285.
[33] R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Math. Oper. Res. 1 (1976) 97.
[34] S. Sachdev, Quantum Phase Transitions, Cambridge Univ. Pr., 2011.
[35] P. Sriluckshmy, M. Nusspickel, E. Fertitta, G. Booth, Fully algebraic and self-consistent effective dynamics in a static quantum embedding, arXiv:2012.05837.
[36] D. Sun, K.-C. Toh, L. Yang, A convergent 3-block semiproximal alternating direction method of multipliers for conic programming with 4-type constraints, SIAM J. Optim. 25 (2015) 882–915.
[37] D. Sun, K.-C. Toh, Y. Yuan, X.-Y. Zhao, Sdpnal+: a Matlab software for semidefinite programming with bound constraints (version 1.0), Optim. Methods Softw. 35 (2020) 87–115.

[38] A. Szabo, N. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, McGraw-Hill, New York, 1989.
[39] M.J. Wainwright, M.I. Jordan, Graphical models, exponential families, and variational inference, Found. Trends Mach. Learn. 1–2 (2008) 1–305.
[40] Z. Wen, D. Goldfarb, W. Yin, Alternating direction augmented Lagrangian methods for semidefinite programming, Math. Program. Comput. 2 (2010) 203–230.
[41] A. Yurtsever, O. Fercoq, V. Cevher, A conditional-gradient-based augmented Lagrangian framework, in: Proceedings of the 36th International Conference on Machine Learning, vol. 97, 2019, pp. 7272–7281.
[42] Z. Zhao, B.J. Braams, M. Fukuda, M.L. Overton, J.K. Percus, The reduced density matrix method for electronic structure calculations and the role of three-index representability conditions, J. Chem. Phys. 120 (2004) 2095–2104.
[43] A.K. Zhuravlev, M.I. Katsnelson, One-dimensional spinless fermion model with competing interactions beyond half-filling, Phys. Rev. B 64 (2001) 033102.