Variational Embedding for Quantum Many-Body Problems

LIN LIN
University of California, Berkeley
Computational Research Division, Lawrence Berkeley National Laboratory
MICHAEL LINDSEY
Courant Institute

Abstract

Quantum embedding theories are powerful tools for approximately solving large-scale, strongly correlated quantum many-body problems. The main idea of quantum embedding is to glue together a highly accurate quantum theory at the local scale and a less accurate quantum theory at the global scale. We introduce the first quantum embedding theory that is also variational, in that it is guaranteed to provide a one-sided bound for the exact ground-state energy. Our method, which we call the variational embedding method, provides a lower bound for this quantity. The method relaxes the representability conditions for quantum marginals to a set of linear and semidefinite constraints that operate at both local and global scales, resulting in a semidefinite program (SDP) to be solved numerically. The accuracy of the method can be systematically improved. The method is versatile and can be applied, in particular, to quantum many-body problems for both quantum spin systems and fermionic systems, such as those arising from electronic structure calculations. We describe how the proper notion of quantum marginal, sufficiently general to accommodate both of these settings, should be phrased in terms of certain algebras of operators. We also investigate the duality theory for our SDPs, which offers valuable perspective on our method as an embedding theory. As a byproduct of this investigation, we describe a formulation for efficiently implementing the variational embedding method via a partial dualization procedure and the solution of quantum analogues of the Kantorovich problem from optimal transport theory.

1 Introduction

Quantum many-body problems, such as the problem of computing the ground state of a system of quantum spins or fermions, have far-reaching applications in physics, chemistry, materials science, and beyond. Certain such problems, including those involving fermions in the "strongly correlated" regime, are among the most challenging problems in scientific computing. Roughly speaking, a ground state of a quantum many-body problem is specified by a wavefunction $|\Phi\rangle$ obtained

Communications on Pure and Applied Mathematics, Vol. LXXV, 2033–2068 (2022)

as a minimizer of the following optimization problem:

$$E_0 = \min_{|\Phi\rangle \in H, \langle \Phi | \Phi \rangle = 1} \langle \Phi | \hat{H} | \Phi \rangle \,, \tag{1.1}$$

in which we have employed the Dirac bra-ket notation, and H is the Hilbert space whose elements are quantum states. The optimization problem (1.1) is equivalent to a linear eigenvalue problem, with the ground state $|\Phi\rangle$ given by the eigenvector corresponding to the smallest eigenvalue (assuming the eigenvalue is simple) of $\hat{H}$. The cost of directly finding $|\Phi\rangle$ generally scales exponentially with respect to the system size. It is therefore of paramount interest to reduce the computational complexity of this task by accepting some controlled sacrifice of accuracy.

Among all the approaches to solving the problem (1.1), some are *variational* in the sense that they provide an approximation for E_0 that is guaranteed to be either an upper or lower bound. For example, methods that restrict the optimization over $|\Phi\rangle \in H$ to some computationally tractable subset provide *upper bounds* for E_0. Examples of such methods include the Hartree-Fock approximation [39], matrix product states (MPS) (also known as tensor trains) [30, 43], and other tensor network methods such as projected entangled-pair states (PEPS) [29, 40]. Meanwhile, other approaches attempt to formulate tractable *relaxations* of the variational principle (1.1). The idea of such approaches is to reformulate (1.1) as an equivalent optimization problem in terms of density matrices, in which the difficulty is encoded in the constraints, and then to enforce only a computationally tractable subset of these constraints. Such procedures yield guaranteed *lower bounds* for E_0. The most well-known example of such an approach is the two-electron reduced density matrix (2-RDM) theory for fermionic systems [1, 6, 9, 22, 24–27, 46].

Another category of approaches to the quantum many-body problem is that of the *quantum embedding theories* [38]. Notable examples include the dynamical mean-field theory (DMFT) [11, 19] and the density matrix embedding theory (DMET) [16, 17]. These methods divide the global system into a set of local clusters (sometimes called fragments), where the size of each cluster is taken to be independent of the global system size. Then one derives a modified quantum many-body problem for each cluster, which can be solved directly or using approximate (but highly accurate) methods. The information from all of the clusters is then "glued" together using global reduced quantities, such as the one-electron reduced density matrix (1-RDM) in DMET, or the single-particle Green's function in DMFT. The method can be solved self-consistently to remove the discrepancy between these global quantities and local fragment data.

In this work we propose an approach to the quantum many-body problem that is the first example to our knowledge of a quantum embedding method that is also variational. We therefore call it the variational embedding method, which we develop below for quantum spin systems and second-quantized fermionic systems. (Note that our framework for quantum spin systems formally includes the setting

of second-quantized bosonic systems as an infinite-dimensional limit.) The fundamental objects considered in our approach are quantum marginals, which are defined with respect to a decomposition of the global system into clusters. The quantum marginals are referred to as such because they are analogous to marginal distributions in the setting of classical probability theory. In the setting of quantum spins, these are just the reduced density operators, which are defined as partial traces of a global density operator. In the fermionic setting, a more general perspective is introduced to define the analogous quantities. This perspective views marginals as functionals on appropriate operator algebras.

Our approach is in particular a relaxation of the variational principle (1.1), and hence yields a *lower bound* for E_0. It is an embedding method in the sense that clusters are represented with high fidelity and glued together via some reduced global data. The accuracy of the variational embedding method can be systematically improved by increasing the cluster size *or* by considering marginals for larger groups of clusters, e.g., pairs, triples, etc. The relaxed optimization problem defining the variational embedding method is a semidefinite program (SDP), whose cost scales polynomially with respect to the system size (for fixed cluster size). Treating this relaxation as the primal problem, we derive the dual problem and show that the duality gap is zero. We also introduce a partial dualization of the primal problem, in which the interpretation as an embedding method becomes even clearer. In particular, we see the emergence of effective Hamiltonians for embedded problems, which are coupled only via the global determination of these effective Hamiltonians. The embedded problems are themselves quantum analogues of the Kantorovich problem of optimal transport [41]. Although our presentation of this quantum Kantorovich problem, which emphasizes general cost operators, differs somewhat from that of the existing literature, the same basic problem has appeared in [5,7,12,34,47].

We also describe how variational embedding adapts to the scenario of overlapping clusters. It can be readily seen that allowing for overlapping clusters tightens the constraints, yielding tighter lower bounds for the ground-state energy at comparable computational cost. This point may be of interest because the value-add of overlapping clusters in embedding theories such as DMET and DMFT is not yet clear [3,45]. We also describe how translation invariance can be exploited in the implementation of variational embedding.

As proof-of-principle, we demonstrate the performance of the variational embedding method for two quantum spin models (the transverse Ising model and the antiferromagnetic Heisenberg model) and one fermionic model (the Hubbard model). The system size is small due to the limitations of the preliminary implementation in CVX [13] within MATLAB®, and we plan to develop more efficient implementations to accommodate larger systems in the near future. In the numerical experiments, we solve the primal problem directly, but the partial dualization mentioned above suggests more efficient methods for solving the variational embedding method, with tractable scaling for extended systems.

1.1 Related work

In the fermionic setting, the aforementioned 2-RDM theory is the closest relative of variational embedding. Nonetheless, we point out that our "fermionic marginals" are different from the 2-RDM. In general, neither the variational embedding method nor the 2-RDM theory adopts a strictly tighter relaxation than the other. Roughly speaking, the variational embedding method enforces tighter constraints "within clusters" but weaker constraints "across clusters" relative to the most accurate 2-RDM theories. Therefore we expect that variational embedding can be more efficient for treating strong correlation effects that are relatively local in nature. That said, both frameworks are highly modular. In fact, it may be possible to adapt existing 2-RDM theories as methods for solving the embedded problems obtained in the variational embedding method. Finally, we comment that the partial dual formulation holds promise for scaling to extended systems, where 2-RDM theories can become prohibitively computationally expensive.

The approach of this paper can also be understood as an approximate method for solving the "quantum marginal problem," [15,35], i.e., the problem of determining whether a set of quantum marginals could have been obtained from a global quantum density operator. In general, the exact solution of this problem is intractable, so approximate methods must be adopted.

Finding approximate solutions to the quantum marginal problem can be viewed as a quantum analogue of the problem of finding outer bounds to the marginal polytope in classical probability [42]. In our approach, we derive two main types of constraints: local consistency constraints (which are linear) and global semidefinite constraints. The local consistency constraints, which enforce compatibility between marginals that share sites, are so termed by analogy to the constraints of the same name appearing in relaxations of the classical marginal polytope [42]. These constraints alone can be viewed as underlying the belief propagation (BP) [31] approximation for classical graphical models (see, also, e.g., [42] for reference). Note with caution that BP should be thought of as an *algorithm*, in addition to a set of modeling assumptions. Also note that BP involves an implicit approximation of the entropy, which is not relevant in the zero-temperature setting, i.e., the setting of this work.

BP has been generalized to the quantum setting (specifically, the setting of quantum spin systems in the sense of this paper) [21], and other works [10, 32] have more carefully studied quantum entropy approximation for quantum spin systems in the context of the local consistency constraints that are featured in BP. Meanwhile, [2] considers a semidefinite relaxation in a zero-temperature, translation-invariant setting for both quantum spins and fermions. In our language, one can view [2] as implicitly considering overlapping clusters for which local consistency constraints (which are generally more complicated to enforce due to cluster overlap) are automatically satisfied without need for explicit enforcement due to the translation invariance. None of these cluster-based works can be viewed as considering an analogue of the global semidefinite constraints introduced in this work.

Moreover, these works only support local Hamiltonians and cannot support long-range (e.g., Coulomb-type) interactions. In fact, the global semidefinite constraints improve the quality of the relaxation even in the case of local Hamiltonians (as we shall demonstrate in Section 4 below), but more dramatically they open the door to cluster-based semidefinite relaxations for long-range Hamiltonians and potentially *ab initio* electronic structure problems.

Another point of comparison is the Lasserre hierarchy [20, 42] of semidefinite relaxations, often considered as means for approximating the marginal polytope in classical probability. Our method is not the quantum analogue of any relaxation from this Lasserre hierarchy in the classical setting, nor is our method recovered from the Lasserre hierarchy as applied directly to the quantum many-body problem. In fact, the variational embedding method can be understood as advancing different systematically improvable hierarchies, both in the cluster size and in the sizes of the groups of clusters for which marginals are considered.

The variational embedding method can also be understood as a way to tighten the variational lower bound obtained in [14] for fermionic many-body problems based on the strictly correlated electron (SCE) formulation [36, 37]. There are two sources of error in the approach of [14]: a model error (which only vanishes in the "strictly correlated" limit of infinitely strong Coulomb repulsion) and an additional relaxation error that emerges from the relaxation of a *classical* marginal problem. The variational embedding method introduced in this paper can be viewed as a fully quantum version of this relaxation. It avoids any analogous notion of model error and can be shown to provide energies at least as tight as those obtained in [14].

1.2 Outline

In Section 2 we formulate variational embedding for quantum spin systems. After preliminary discussion in Section 2.1, we go on to introduce the local consistency constraints and global semidefinite constraints in Sections 2.2 and 2.3, respectively. In Section 2.4 we discuss a more abstract perspective on the global semidefinite constraints that is, in particular, more portable to the fermionic setting to appear later on. In Section 2.5 we introduce variational embedding constraints for higher marginals (i.e., marginals for higher tuples of sites), and in Section 2.6 we introduce the cluster perspective on variational embedding. In Section 2.7 we discuss how variational embedding can accommodate overlapping clusters for tighter relaxations, and in Section 2.8 we discuss how translation-invariance can be exploited, as well as additional "periodicity constraints" that can be imposed in this setting.

Section 3 concerns the formulation of variational embedding for fermionic systems in second quantization. After discussing preliminaries in Section 3.1, we employ the language of star-algebras to define appropriate fermionic marginals in Section 3.2. Using this language, we provide an abstract formulation of variational embedding for fermions in Section 3.3, which we show is exact for noninteracting problems (i.e., problems specified by single-body Hamiltonians) in Section 3.4. In

Section 3.5, we demonstrate how the abstract formulation can be practically implemented as an SDP.

Section 4 presents various numerical experiments. In Sections 4.1, 4.2, and 4.3 we treat the transverse-field Ising, antiferromagnetic Heisenberg, and Hubbard models, respectively.

Finally, we conclude in Section 5 with a discussion of duality for the SDP of variational embedding. To prepare for the formulation of the dual problem, we discuss in Section 5.1 a quantum analogue of the Kantorovich problem from optimal transport. Then in Section 5.2 we introduce a partially dualized SDP, which reveals that the variational embedding solution can be obtained as the solution of several quantum Kantorovich problems specified by "effective Hamiltonians," which are completely decoupled from one another apart from the determination of these effective Hamiltonians. In Section 5.3 we discuss the computational implications of this observation, and in Section 5.4 we close with a derivation of the full dual problem and a discussion of strong duality.

2 Quantum Spins

2.1 Preliminaries

Let $i = 1, \dots, M$ index the sites, and for each site i let X_i be the classical state space (discrete, for simplicity). For each site, the quantum state space is $Q_i := \mathbb{C}^{X_i}$, and the global quantum state space is

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

where $\mathcal{X} := \prod_{i=1}^{M} X_i$. Let H_i denote a Hermitian operator $Q_i \to Q_i$, and let H_{ij} denote a Hermitian operator $Q_i \otimes Q_j \to Q_i \otimes Q_j$. We will use the hatted notation $\widehat{H}_i$ to denote the operator $\mathcal{Q} \to \mathcal{Q}$ obtained by tensoring H_i by the identity operator on all sites $k \neq i$, and likewise we identify $\widehat{H}_{ij}$ with the operator $\mathcal{Q} \to \mathcal{Q}$ obtained by tensoring H_{ij} with the identity on all sites $k \notin \{i, j\}$. Then we consider a Hamiltonian $\widehat{H} : \mathcal{Q} \to \mathcal{Q}$ of the form

$$\widehat{H} = \sum_i \widehat{H}_i + \sum_{i<j} \widehat{H}_{ij}.$$

REMARK 2.1. *We shall introduce several examples of interest in the case* $X_i = \{-1, 1\}$*, i.e., the case of quantum spin-*$\frac{1}{2}$ *systems. The Pauli matrices*

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

together with the identity I_2 *form a basis for Hermitian operators on* $\mathbb{C}^2$*. Now let* $\sigma_i^{x/y/z} \in \mathcal{H}(\bigotimes_i \mathbb{C}^2) \simeq \bigotimes_i \mathcal{H}(\mathbb{C}^2)$ *be obtained by tensoring a copy of* $\sigma^{x/y/z}$ *for the* i^{th} *site with the identity* I_2 *on all the other sites. Two examples of the*

quantum spin systems are the transverse-field Ising (TFI) Hamiltonian and antiferromagnetic Heisenberg (AFH) Hamiltonian, specified by the Hamiltonians

$$\widehat{H}_{\mathrm{TFI}} = -h\sum_i \sigma_i^x - \sum_{i\sim j}\sigma_i^z\sigma_j^z, \tag{2.1}$$

$$\widehat{H}_{\mathrm{AFH}} = \sum_{i\sim j}\left[\sigma_i^x\sigma_j^x + \sigma_i^y\sigma_j^y + \sigma_i^z\sigma_j^z\right]. \tag{2.2}$$

where the summation of $i \sim j$ indicates summation over all pairs of indices that are adjacent in a graph defined on the index set (usually the graph is a square lattice). In the TFI Hamiltonian, $h \in \mathbb{R}$ is a scalar parameter.

We are interested in computing the ground-state energy

$$E_0 = \inf\{\langle\Phi|\widehat{H}|\Phi\rangle : |\Phi\rangle \in \mathcal{Q},\ \langle\Phi|\Phi\rangle = 1\}.$$

It can be equivalently recast as

$$E_0 = \inf_{\rho\in\mathcal{D}(\mathcal{Q})} \mathrm{Tr}[\widehat{H}\rho],$$

where $\mathcal{D}(\mathcal{Q})$ denotes the set of density operators on $\mathcal{Q}$ (i.e., positive semidefinite Hermitian operators $\mathcal{Q} \to \mathcal{Q}$ of unit trace). Assuming that there exists a unique ground state $|\Phi_0\rangle$, the infimum is attained at $\rho = |\Phi_0\rangle\langle\Phi_0|$. Now we can write

$$E_0 = \inf_{\{\rho_{ij}\}_{i<j}\in\mathbf{QM}_2(\mathcal{Q})}\Big(\sum_i \mathrm{Tr}[H_i\rho_i] + \sum_{i<j}\mathrm{Tr}[H_{ij}\rho_{ij}]\Big), \tag{2.3}$$

where $\mathbf{QM}_2(\mathcal{Q})$ denotes the set of collections $\{\rho_{ij}\}_{i<j}$ of representable quantum two-marginals, i.e., those collections $\{\rho_{ij}\}$ that can be obtained as reduced density operators of a single $\rho \in \mathcal{D}(\mathcal{Q})$ via the partial trace, as in

$$\rho_{ij} = \mathrm{Tr}_{\{1,\ldots,M\}\setminus\{i,j\}}[\rho],$$

where $i < j$.

To clarify, here we view ρ as being equipped with labels $1,\ldots,M$ for its indices as $\rho = \rho_{i_1\cdots i_M, j_1\cdots j_M}$, and for any subset $S \subset \{1,\ldots,M\}$, $\rho_S = \mathrm{Tr}_{\{1,\ldots,M\}\setminus S}[\rho]$ denotes the reduced density operator obtained by tracing out the indices contained in S, with the remaining labels maintained. We comment that the partial trace ρ_S may be equivalently defined as the unique operator on $\bigotimes_{i\in S}\mathcal{Q}_i$ such that $\mathrm{Tr}[\widehat{A}\rho_S] = \mathrm{Tr}[\widehat{A}\rho]$ for all operators $\widehat{A}$ on $\bigotimes_{i\in S}\mathcal{Q}_i$ (alternatively viewed as operators on $\mathcal{Q}$ by tensoring with the identity). This perspective illustrates the relationship between marginalization in the quantum spin setting (i.e., computing the partial trace) and the more abstract notion of marginalization that is necessary for the treatment of fermions in Section 3 below.

For convenience, we denote $\rho_{ij} = \rho_{\{i,j\}}$ for $i < j$ as above. It is convenient to then define ρ_{ij} for $i > j$ via the stipulation that $\sigma_{ij}\rho_{ij}\sigma_{ji} = \rho_{ji}$, where $\sigma_{ij} : \mathcal{Q}_i \otimes \mathcal{Q}_j \to \mathcal{Q}_j \otimes \mathcal{Q}_i$ is the linear operator defined by $\sigma_{ij}(\phi_i \otimes \phi_j) = \phi_j \otimes \phi_i$. Finally, we remark that the one-marginals $\rho_i = \mathrm{Tr}_{\{1,\ldots,M\}\setminus\{i\}}[\rho]$ are determined

by the two-marginals via $\rho_i = \mathrm{Tr}_{\{j\}}[\rho_{ij}]$, and this dependence is meant to be understood implicitly in (2.3). We will occasionally denote $\rho_{ii} := \rho_i$.

2.2 Local consistency constraints

Now it is of interest to determine necessary conditions satisfied by collections in $\mathbf{QM}_2(\mathcal{Q})$. By enforcing a set of necessary conditions as a proxy for membership in $\mathbf{QM}_2(\mathcal{Q})$, we can obtain a lower bound on the ground state energy.

To begin with, the ρ_{ij} are themselves density operators on $Q_i \otimes Q_j$, i.e., $\rho_{ij} \succeq 0$ with $\mathrm{Tr}[\rho_{ij}] = 1$. Moreover, we must have $\mathrm{Tr}_j[\rho_{ij}] = \mathrm{Tr}_{j'}[\rho_{ij'}]$ for all i and j, $j' \neq i$, and we must have $\sigma_{ij}\rho_{ij}\sigma_{ji} = \rho_{ji}$. These constraints define the set of *locally consistent quantum two-marginals*. Call this set $\mathbf{LQM}_2(\mathcal{Q})$. In practice, we define auxiliary variable ρ_i for the one-marginals, constrained to satisfy $\rho_i = \mathrm{Tr}_j[\rho_{ij}] = \mathrm{Tr}_j[\rho_{ji}]$. The constraints $\mathrm{Tr}[\rho_{ij}] = 1$ for all i, j can in fact be enforced by requiring $\mathrm{Tr}[\rho_i] = 1$ for all i, since $\mathrm{Tr}[\rho_{ij}] = \mathrm{Tr}[\mathrm{Tr}_j[\rho_{ij}]]$.

Note that the local consistency constraint $\mathrm{Tr}_j[\rho_{ij}] = \rho_i$ is equivalent to insisting that $\mathrm{Tr}[\hat{A}\rho_{ij}] = \mathrm{Tr}[\hat{A}\rho_i]$ for all operators $\hat{A}$ on Q_i (considered also as operators on $Q_i \otimes Q_j$ by tensoring with the identity). This perspective highlights the connection to the abstract local consistency constraints appearing in the discussion of fermionic systems in Section 3 below.

2.3 Global semidefinite constraints and the two-marginal SDP

We can derive a further constraint, more global in nature, as follows. Consider operators $\hat{O} : \mathcal{Q} \to \mathcal{Q}$ (not necessarily Hermitian) of the form $\hat{O} = \sum_i \hat{O}_i$, where each $\hat{O}_i$ is a one-body operator on $\mathcal{Q}$, i.e., obtained by tensoring an operator O_i on Q_i with the identity. Now $\hat{O}^\dagger \hat{O} \succeq 0$, so

$$\mathrm{Tr}\left[\rho\, \hat{O}^\dagger \hat{O}\right] \geq 0 \tag{2.4}$$

for any $\rho \in \mathcal{D}(\mathcal{Q})$. We will expand the left-hand side to obtain a constraint on the quantum two-marginals, which can be phrased as a semidefinite matrix constraint.

First compute

$$\begin{aligned} 0 \leq \mathrm{Tr}\left[\rho\, \hat{O}^\dagger \hat{O}\right] &= \mathrm{Tr}\left[\rho \sum_{ij} \hat{O}_i^\dagger \hat{O}_j\right] \\ &= \sum_i \mathrm{Tr}\left[\rho_i O_i^\dagger O_i\right] + \sum_{i \neq j} \mathrm{Tr}\left[\rho_{ij} O_i^\dagger \otimes O_j\right]. \end{aligned}$$

Now without loss of generality, we can identify X_i with $\{1, \ldots, m_i\}$ where $m_i = |X_i|$. Hence we can think of O_i as an arbitrary complex matrix

$$O_i = (O_{i,kl})_{k,l=1,\ldots,m_i}.$$

We will use square brackets to indicate entries of an operator as in $[O_i]_{kl} = O_{i,kl}$. Note that the two-marginal ρ_{ij} is an operator $Q_i \otimes Q_j \to Q_i \otimes Q_j$, so we denote

its $((k,p),(l,q))$ entry by $[\rho_{ij}]_{kp,lq}$ for $k,l = 1,\dots,m_i$ and $p,q = 1,\dots,m_j$. Finally, for $i \neq j$, observe that

$$\left[O_i^\dagger \otimes O_j\right]_{kp,lq} = [O_i^\dagger]_{kl}[O_j]_{pq} = \overline{O_{i,lk}} O_{j,pq}.$$

Then we expand the $i \neq j$ sum to obtain

$$\begin{aligned}\sum_{i\neq j} \mathrm{Tr}\left[\rho_{ij}\, O_i^\dagger \otimes O_j\right] &= \sum_{i\neq j}\sum_{k,l=1}^{m_i}\sum_{p,q=1}^{m_j} [\rho_{ij}]_{lq,kp}\left[O_i^\dagger \otimes O_j\right]_{kp,lq} \\ &= \sum_{i\neq j}\sum_{k,l=1}^{m_i}\sum_{p,q=1}^{m_j} [\rho_{ij}]_{lq,kp}\overline{O_{i,lk}} O_{j,pq} \\ &= \sum_{i,j=1}^{M}\sum_{k,l=1}^{m_i}\sum_{p,q=1}^{m_j} (1-\delta_{ij})[\rho_{ij}]_{lq,kp}\overline{O_{i,lk}} O_{j,pq}.\end{aligned}$$

Next expand the i sum:

$$\begin{aligned}\sum_{i} \mathrm{Tr}\left[\rho_i\, O_i^\dagger O_i\right] &= \sum_{i}\sum_{k=1}^{m_i}\sum_{q=1}^{m_i} [\rho_i]_{qk}\left[O_i^\dagger O_i\right]_{kq} \\ &= \sum_{i}\sum_{k,l=1}^{m_i}\sum_{q=1}^{m_i} [\rho_i]_{qk}\left[\hat{O}_i^\dagger\right]_{kl}\left[\hat{O}_i\right]_{lq} \\ &= \sum_{i}\sum_{k,l=1}^{m_i}\sum_{q=1}^{m_i} [\rho_i]_{qk}\overline{O_{i,lk}} O_{i,lq} \\ &= \sum_{i}\sum_{k,l=1}^{m_i}\sum_{p,q=1}^{m_i} \delta_{lp}[\rho_i]_{qk}\overline{O_{i,lk}} O_{i,pq} \\ &= \sum_{i,j=1}^{M}\sum_{k,l=1}^{m_i}\sum_{p,q=1}^{m_j} \delta_{ij}\delta_{lp}[\rho_i]_{qk}\overline{O_{i,lk}} O_{i,pq}.\end{aligned}$$

Therefore we have derived

$$\sum_{i,j=1}^{M}\sum_{k,l=1}^{m_i}\sum_{p,q=1}^{m_j} \left[\delta_{ij}\delta_{lp}[\rho_i]_{qk} + (1-\delta_{ij})[\rho_{ij}]_{lq,kp}\right]\overline{O_{i,lk}} O_{j,pq} \geq 0.$$

We can think of $O_{j,pq}$ as a vector $O \in \prod_{i=1}^{M} \mathbb{C}^{m_i\times m_i} \simeq \mathbb{C}^{\sum_{i=1}^{M} m_i^2}$. The choice of such O was completely arbitrary. Therefore we have proved that

$$\left(\sum_{i=1}^{M} m_i^2\right) \times \left(\sum_{i=1}^{M} m_i^2\right) \text{ matrix } G^{(2)} = G^{(2)}[\{\rho_{ij}\}_{i\leq j}]$$

defined by

$$G^{(2)}_{ilk,jpq} := \delta_{ij}\delta_{lp}[\rho_i]_{qk} + (1-\delta_{ij})[\rho_{ij}]_{lq,kp}$$

is positive semidefinite. This matrix can be thought of as a linear operator $G^{(2)} : \prod_{i=1}^{M} \mathbb{C}^{m_i \times m_i} \to \prod_{i=1}^{M} \mathbb{C}^{m_i \times m_i}$. (One can readily check that $G^{(2)}$ is Hermitian.) For a quantum spin system, we have $m_i = 2$ for all i, so this is a semidefinite constraint on a $(4M) \times (4M)$ matrix, which can be efficiently enforced.

At last we have derived a semidefinite relaxation, which we shall call the *two-marginal SDP*:

$$E_0^{(2)} = \inf_{\{\rho_{ij}\}_{i<j} \in \mathbf{LQM}_2(\mathcal{Q}): G^{(2)}[\{\rho_{ij}\}_{i\le j}] \succeq 0} \Big(\sum_i \mathrm{Tr}[H_i \rho_i] + \sum_{i<j} \mathrm{Tr}[H_{ij}\rho_{ij}] \Big).$$

The relaxation yields the energy lower bound $E_0 \ge E_0^{(2)}$, as well as a minimizer $\rho^{(2)}$ that is expected to approximate the exact two-marginals.

The two-marginal SDP can be written, in expanded form, as

$$\text{(2.5)} \quad \underset{\{\rho_i\},\{\rho_{ij}\}_{i<j}}{\text{minimize}} \quad \sum_i \mathrm{Tr}[H_i\rho_i] + \sum_{i<j} \mathrm{Tr}[H_{ij}\rho_{ij}]$$

$$\text{(2.6)} \quad \text{subject to} \quad \rho_{ij} \succeq 0, \quad i,j = 1,\dots,M,$$

$$\text{(2.7)} \quad \rho_i = \mathrm{Tr}_{\{j\}}[\rho_{ij}],\ \rho_j = \mathrm{Tr}_{\{i\}}[\rho_{ij}], \quad i,j = 1,\dots,M,$$

$$\text{(2.8)} \quad \mathrm{Tr}[\rho_i] = 1, \quad i = 1,\dots,M,$$

$$\text{(2.9)} \quad G^{(2)}[\{\rho_{ij}\}_{i\le j}] \succeq 0.$$

Although there are several ways to write constraints yielding the same feasible set, the dual SDP is actually influenced by the choice of constraints used to define this set. The choices made here will yield interesting dual structure, to be explored in Section 5.

2.4 Abstract perspective on the global semidefinite constraints

More abstractly, it is useful to think of $G = G[\{\rho_{ij}\}]$ as being composed of blocks $G_{ij}[\rho_{ij}]$ (indexed by marginal pairs i, j), defined by

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

where $\{O_{i,\alpha}\}_{\alpha=1}^{m_i^2}$ is a basis for the set of one-body operators on site i. By considering α as a multi-index $\alpha = (k,l)$ and choosing $(O_{i,(k,l)})_{k',l'} = \delta_{kk'}\delta_{ll'}$ to be the "standard unit vectors" in $\mathbb{C}^{m_i \times m_i}$, we exactly recover our former explicit representation of $G[\{\rho_{ij}\}]$.

REMARK 2.2 (Restricted operator sets). *The more abstract perspective suggests a natural framework for further relaxation. Suppose that for each $i = 1,\dots,M$, we are given a linearly independent collection $\{O_{i,\alpha}\}_{\alpha\in\mathcal{I}_i}$ of one-body operators for the i^{th} site, where $\mathcal{I}_i$ is a given index set. Then we can define $G = G[\rho^{(2)}]$ in terms*

of blocks as above, where the block $G_{ij}[\rho_{ij}]$ is a matrix of size $|\mathcal{I}_i| \times |\mathcal{I}_j|$, defined once again by

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

for $\alpha \in \mathcal{I}_i$, $\beta \in \mathcal{I}_j$. In principle one can consider restricted index sets with $|\mathcal{I}_i| < m_i^2$ containing only the most physically important operators. Such restricted structure will correspond to interesting structure from the perspective of the dual problem to be considered below.

REMARK 2.3 (Quasi-local constraints). *In order to improve the efficiency of the semidefinite introduced above, one could enforce the semidefiniteness of certain principal submatrices of G. For example, for each k, one could define a submatrix $G^{(k)}$ of G by restricting the block indices i, j to those satisfying $d(i,k), d(j,k) \leq d_{\max}$, where $d(\cdot,\cdot)$ is an appropriate notion of distance between indices (e.g., graph distance for a lattice model) and $d_{\max}$ is a locality parameter. Then one enforces $G^{(k)}[\{\rho_{ij}\}] \succeq 0$ for all k. For constant $d_{\max}$ suitably large, in principle such constraints could achieve good performance while maintaining linear scaling in M of the SDP problem size for suitably local Hamiltonians, by omitting ρ_{ij} from the optimization variables for $d(i,j) > d_{\max}$.*

2.5 Higher marginal constraints

A tighter SDP relaxation can be derived by considering a set $\{\rho_{ijk}\}_{i<j<k}$ of quantum three-marginals as the optimization variable. One may enforce the suitably defined local consistency constraints, denoted $\{\rho_{ijk}\}_{i<j<k} \in \mathbf{LQM_3}(\mathcal{Q})$, then defining variables ρ_{ij} in terms of the ρ_{ijk} via partial traces, and additionally enforce $G[\{\rho_{ij}\}_{i\leq j}] \succeq 0$. We refer to the corresponding semidefinite relaxation as the *three-marginal SDP*.

To derive the corresponding semidefinite constraints, we have to keep track of the four-marginals. Suitable necessary conditions can derived by enforcing $\mathrm{Tr}[\rho\, \widehat{O}^\dagger \widehat{O}] \geq 0$ for all $\widehat{O}$ of the form $\widehat{O} = \sum_{i,i'} \widehat{O}_{i,i'}$, where the $\widehat{O}_{i,i'}$ are *two-body* operators. As such one may define the *four-marginal SDP*, and so on. Note that, e.g., the four-marginal SDP can in fact accommodate more general Hamiltonians, i.e., Hamiltonians including additional four-body terms.

2.6 Cluster perspective

In order to systematically improve the accuracy of the two-marginal SDP, instead of considering higher marginals we may alternately consider *increasing the cluster size*. Formally, such considerations will yield problems can still be accommodated as special cases of our previously introduced setting. However, the

difference in perspective is noteworthy, and the generalization to the case of overlapping clusters (considered in the next section) is *not* accommodated as such a special case.

Suppose that our site index set is written as a union of cluster index sets C_γ, i.e.,

$$\{1, \ldots, M\} = \bigcup_{\gamma=1}^{N_c} C_\gamma,$$

where the cluster index sets C_γ are *disjoint*. Then one can define

$$Y_\gamma := \prod_{i \in C_\gamma} X_i$$

to be the classical state space for the γ^{th} cluster. Then by considering the clusters now as *sites* with classical state spaces Y_γ and following the derivation of the two-marginal SDP, we may derive the *cluster two-marginal SDP*, relative to the cluster decomposition $\{C_\gamma\}$. Note that this problem may be viewed formally as a two-marginal SDP; however, the distinction makes sense when we think of the limit of expanding clusters for a problem that is otherwise fixed. Higher-marginal cluster SDPs can be derived similarly.

2.7 Overlapping clusters

We now demonstrate the treatment of overlapping clusters. Suppose again that

$$\{1, \ldots, M\} = \bigcup_{\gamma=1}^{N_c} C_\gamma,$$

but now relax the assumption that the C_γ are disjoint. Since the overlap of two clusters might even be a single site of the original model, we can no longer just "coarse-grain" clusters and neglect all of their intracluster structure. In particular, imposition of necessary local consistency constraints demands a bit more care.

Now the primary objects in our relaxation will be the two-*cluster* marginals, denoted $\rho_{\gamma\delta} := \rho_{C_\gamma \cup C_\delta}$ for $\gamma \leq \delta$. Each $\rho_{\gamma\delta}$ is an operator on the quantum state space specified by the *union* of sites $C_\gamma \cup C_\delta$, which may of course be smaller in size than $|C_\gamma| + |C_\delta|$. Then the one-cluster marginals $\rho_\gamma := \rho_{C_\gamma}$ (which we sometimes also denote by $\rho_{\gamma\gamma}$) are obtained in terms of the two-cluster marginals via

$$\rho_\gamma = \mathrm{Tr}_{C_\delta \setminus C_\gamma}[\rho_{\gamma\delta}], \quad \rho_\delta = \mathrm{Tr}_{C_\gamma \setminus C_\delta}[\rho_{\gamma\delta}].$$

These identities yield consistency constraints analogous to the local consistency constraints introduced earlier. However, we can also include the *overlap* constraints by introducing the variable $\rho_{(\gamma\delta)\cap(\gamma'\delta')}$ representing the marginal corresponding to the set $(C_\gamma \cup C_\delta) \cap (C_{\gamma'} \cup C_{\delta'})$, for all $\gamma < \delta$, $\gamma' < \delta'$, constrained by

$$\rho_{(\gamma\delta)\cap(\gamma'\delta')} = \mathrm{Tr}_{(C_\gamma \cup C_\delta)\setminus(C_{\gamma'} \cup C_{\delta'})}\left[\rho_{\gamma\delta}\right] = \mathrm{Tr}_{(C_{\gamma'} \cup C_{\delta'})\setminus(C_\gamma \cap C_\delta)}\left[\rho_{\gamma'\delta'}\right].$$

Note that these constraints are nontrivial only if the intersection $(C_\gamma \cup C_\delta) \cap (C_{\gamma'} \cup C_{\delta'})$ of cluster pairs is nonempty.

To complete the discussion of the overlapping cluster two-marginal SDP, we need to derive the semidefinite constraint. This is derived by observing the necessary condition $\mathrm{Tr}[\rho\, \widehat{O}^\dagger \widehat{O}] \geq 0$ for all $\widehat{O}$ of the form $\widehat{O} = \sum_\gamma \widehat{O}_\gamma$, where $\widehat{O}_\gamma$ is a one-cluster operator, i.e., an operator on $\bigotimes_{i \in C_\gamma} Q_i$, interpreted also (abusing notation slightly) as an operator on $\mathcal{Q}$ by tensoring with the identity on all sites outside of C_γ.

In fact, given a collection of one-cluster operators $\{O_{\gamma,\alpha}\}_{\alpha \in \mathcal{I}_\gamma}$ for the γ^{th} cluster (i.e., operators on $\bigotimes_{i \in C_\gamma} Q_i$), we build $G[\{\rho_{\gamma\delta}\}]$ blockwise by defining

$$(G_{\gamma\delta}[\rho_{\gamma\delta}])_{\alpha\beta} = \mathrm{Tr}\big[\rho_{\gamma\delta}\, \widetilde{O}^\dagger_{\gamma,\alpha} \widetilde{O}_{\delta,\beta}\big]$$

for $\alpha \in \mathcal{I}_\gamma$, $\beta \in \mathcal{I}_\delta$, and $\gamma < \delta$ (extending to $\gamma > \delta$ by hermiticity), where $\widetilde{O}_{\gamma,\alpha}$ is an operator on $\bigotimes_{i \in C_\gamma \cup C_\delta} Q_i$ obtained from $O_{\gamma,\alpha}$ by tensoring with the identity operator over all sites in $C_\delta \setminus C_\gamma$. For example, if $C_\gamma = \{1, 2\}$ and $C_\delta = \{2, 3\}$, then we can represent $\widetilde{O}_{\gamma,\alpha} = O_{\gamma,\alpha} \otimes I_{m_3}$ and $\widetilde{O}_{\delta,\beta} = I_{m_1} \otimes O_{\delta,\beta}$ (recall that here $O_{\gamma,\alpha}$ is an operator on $Q_1 \otimes Q_2$ and $O_{\delta,\beta}$ is an operator on $Q_2 \otimes Q_3$).

The semidefinite constraint is, as before, $G[\{\rho_{\gamma\delta}\}] \succeq 0$. The resulting SDP can accommodate Hamiltonians of the form

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

where $\widehat{H}_\gamma$ and $\widehat{H}_{\gamma\delta}$ are one-cluster and two-cluster operators, respectively.

Suitable analogous relaxations with higher overlapping cluster marginal constraints may also be derived. We remark that the treatment of overlapping clusters here is significantly simpler and more principled than several other quantum embedding theories, including the dynamical mean-field theory (DMFT) and density matrix embedding theory (DMET).

2.8 Translation-invariant setting

In this section we describe how translation-invariant structure can be exploited in a natural way in our semidefinite relaxation framework. For simplicity we focus only on the case of the two-marginal SDP for a translation-invariant Hamiltonian in one dimension. Extension to higher dimensions is straightforward.

For the purposes of this section it is convenient to adopt a zero-indexing convention for our site indices (usually denoted by i, j); i.e., we index our sites as $i = 0, \ldots, M - 1$. We obtain a translation-invariant Hamiltonian by assuming that $\widehat{H}_i = \widehat{H}_0$ for all i and $\widehat{H}_{ij} = \widehat{H}_{0,j-i}$ for all $i < j$. In turn, we are guaranteed translation-invariance of the ground-state density operator (note: symmetry-breaking cannot occur for systems of finite size). In particular, we have $\rho_i = \rho_0$ for all i and $\rho_{ij} = \rho_{0,j-i}$ for all $i < j$, and it follows that we can constrain the matrix $G = G[\{\rho_{ij}\}]$ to be block-circulant, so that the block G_{ij} depends only

on $i - j \pmod M$. Hence all of the information of G is contained in the first row of blocks, and moreover G can be block-diagonalized by taking the blockwise discrete Fourier transform of the first row of blocks. Indeed, these diagonal blocks are obtained as

$$\widetilde{G}_k = \frac{1}{\sqrt{M}} \sum_{j=0}^{M-1} \exp\left(\imath \frac{2\pi jk}{M}\right) G_{0j},$$

$k = 0, \ldots, M-1$, where we use "$\imath$" to denote the imaginary unit to avoid confusion with our indexing notation. Now the constraint $G \succeq 0$ is equivalent to the constraint that $\widetilde{G}_k \succeq 0$ for all k. Hence we arrive at the periodic two-marginal SDP:

$$\begin{aligned}
\underset{\rho_0, \{\rho_{0j}\}_{j=0,\ldots,M-1}}{\text{minimize}} \quad & \mathrm{Tr}[H_0 \rho_0] + \sum_{j=1}^{M-1} \mathrm{Tr}[H_{0j} \rho_{0j}] \\
\text{subject to} \quad & \rho_{0j} \succeq 0, \quad j = 0, \ldots, M-1, \\
& \rho_0 = \mathrm{Tr}_{\{j\}}[\rho_{0j}],\ \rho_0 = \mathrm{Tr}_{\{0\}}[\rho_{0j}], \quad j = 0, \ldots, M-1, \\
& \mathrm{Tr}[\rho_0] = 1, \\
& \sum_{j=0}^{M-1} \exp\left(\imath \frac{2\pi jk}{M}\right) G_{0j}[\rho_{0j}] \succeq 0, \quad k = 0, \ldots, M-1.
\end{aligned}$$

Notice that we have economized significantly on optimization variables, and, moreover, we have exchanged a semidefinite constraint of size $\sim M$ for M semidefinite constraints of size constant in M. Moreover, a careful implementation of a solver for this SDP should be able to exploit the FFT in the implementation of the semidefinite constraints.

Periodicity constraints

If our sites are obtained as composite sites representing nonoverlapping clusters (as discussed in Section 2.6), and if, moreover, our Hamiltonian is translation-invariant with respect to these *underlying* sites, then we can impose further constraints to enforce the *internal* translation-invariance of our cluster marginals. To wit, in addition to our optimization variables $\{\rho^{\mathrm{C}}_{0\delta}\}$ for the two-*cluster* marginals, we can define additional optimization variables $\{\rho_{0j}\}$ for the two-*site* marginals and then enforce, for all $i \in C_0$, $j \in \{1, \ldots, M\}$, that

$$\rho_{0,j-i} = \mathrm{Tr}_{C_0 \cup C_{\delta(j)} \setminus \{i,j\}}\left[\rho^{\mathrm{C}}_{0,\delta(j)}\right],$$

where $\delta(j)$ is the index of the cluster containing site j. We refer to these additional constraints as *periodicity constraints*.

3 Fermions

3.1 Preliminaries

The fundamental objects of fermionic systems in the second quantized formulation (see, e.g., [28]) are the creation operators $a_1^\dagger, \dots, a_M^\dagger$ and their Hermitian adjoints, the annihilation operators a_i, which act on the Fock space $\mathcal{F}$ and satisfy 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. One defines the number operators by $\hat{n}_i := a_i^\dagger a_i$ and the total number operator by $\hat{N} := \sum_{i=1}^M \hat{n}_i$.

These objects can be concretely realized via the identification of Hilbert spaces $\mathcal{F} \simeq \bigotimes^M \mathbb{C}^2 \simeq \mathbb{C}^{2^M}$, under which the annihilation operators correspond to quantum spin-$\frac{1}{2}$ operators as

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

This identification of operators defines the Jordan-Wigner transformation (JWT) [28]. Note that the JWT depends on the ordering of the states in the sense that permuting the states before the JWT is not equivalent to permuting the tensor factors after the JWT.

After specifying a particle-number-conserving Hamiltonian $\widehat{H}$, i.e., a Hermitian operator on the Fock space that commutes with $\hat{N}$, and a fixed particle number N, we are interested in computing the energy

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

It is equivalent to solve

$$E_0(N) = \inf_{\rho \in \mathcal{D}(\mathcal{F}) : \mathrm{Tr}[\hat{N}\rho] = N} \mathrm{Tr}\big[\widehat{H}\rho\big],$$

where $\mathcal{D}(\mathcal{F})$ indicates the set of density operators on the Fock space (i.e., positive semidefinite Hermitian operators $\mathcal{F} \to \mathcal{F}$ of unit trace).

Observe that although $\mathcal{F}$ can be identified with a quantum-spin state space, the creation operators are *not* one-qubit operators in the sense of quantum spin systems, nor are hopping operators $a_i^\dagger a_j + a_j^\dagger a_i$ generically two-qubit operators. Moreover, the complexity of such operators after the JWT can depend unphysically on the ordering of the sites. Hence most second-quantized problems of interest (with the exception of local one-dimensional models) *simply do not* fit into the framework of variational embedding introduced above for quantum spin systems.

To illustrate this point and provide some concrete examples, we now describe several Hamiltonians of interest in this setting. Of particular note is the Hubbard model, whose states we enumerate via the orbital-spin index (i, σ), where $i =$

$1, \dots, M, \sigma = \uparrow, \downarrow$.

$$\widehat{H} = -t \sum_{ij\sigma} A_{ij} a_{i\sigma}^\dagger a_{j\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \tag{3.1}$$

where A_{ij} is the adjacency matrix of a graph with vertex set $\{1, \dots, M\}$, e.g., a one-dimensional chain or a two-dimensional square lattice. The Hubbard model plays a significant role in understanding strongly correlated quantum systems, such as the high temperature superconductivity [33].

More generally, one can consider a "generalized Coulomb model" of the form

$$\widehat{H} = \sum_{ij\sigma} h_{ij} a_{i\sigma}^\dagger a_{j\sigma} + \sum_{ij\sigma\tau} U_{ij} \hat{n}_{i\sigma} \hat{n}_{j\tau},$$

which includes in particular the Hubbard model and variants with longer-range interactions. In fact,nuum can be mapped to second-quantized Hamiltonians of this form. As we shall see, the generalized Coulomb model is accommodated naturally within the framework of fermionic variational embedding.

Broadening our view further still, consider a general two-body Hamiltonian $\widehat{H}$, written as

$$\widehat{H} = \sum_{ij} h_{ij} a_i^\dagger a_j + \frac{1}{2} \sum_{ijkl} v_{ijkl} a_i^\dagger a_j^\dagger a_l a_k.$$

Electronic structure problems in first quantization can be mapped to such Hamiltonians via an arbitrary choice of orbital basis $\{\phi_i\}$ for (a subspace of) $L^2(\mathbb{R}^d)$, where d is the physical dimension. If the basis functions have compact support, then v_{ijkl} can be nonzero only if both $\mathrm{supp}(\phi_i) \cap \mathrm{supp}(\phi_k) \neq \varnothing$ and $\mathrm{supp}(\phi_j) \cap \mathrm{supp}(\phi_l) \neq \varnothing$. It will follow that after a suitable choice of overlapping clusters, such Hamiltonians can also be accommodated within fermionic variational embedding. We leave investigation of *ab initio* quantum chemistry problems by these means to future work.

In order to define a convex relaxation of the fermionic Gibbs variational principle that is analogous to our relaxation for quantum spin systems, we adopt a more abstract (and indeed general) perspective in section 3.2, allowing for the derivation of a suitable two-cluster-marginal SDP in Section 3.3. We will in fact see in Section 3.4 that our relaxation is tight for noninteracting Hamiltonians, i.e., Hamiltonians that are quadratic in the creation and annihilation operators. To our knowledge this feature has no analogue in the quantum spin setting because there is no related notion of noninteracting systems. Then in Section 3.5, we will describe how one can translate our abstract convex optimization problem into an explicit SDP that can be implemented in practice.

3.2 Abstract perspective

The fundamental objects of interest in the abstract perspective is the *algebra of operators* on the Fock space. In fact, the Fock space itself plays no direct role in the following developments, nor does any global JWT. Marginalization will make

use of the notion of a *subalgebra* subordinate to each cluster. It is in the details of how these subalgebras lie within the global algebra that the quantum-spin and fermionic cases differ.

Now let

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

denote the unital star-algebra over the complex numbers[1] generated by the creation and annihilation operators subject to the canonical anticommutation relations. (Throughout we will use angle brackets to denote such generated algebras.) We let $\hat{n}_i = a_i^\dagger a_i$ denote the corresponding number operators and let $\hat{N} = \sum_i \hat{n}_i$ denote the total number operator. Recall from above that for spinful models such as the Hubbard model, the state index i can be thought of as a composite orbital-spin index, i.e., $i = (x, \sigma)$.

In fact, the algebra $\mathcal{A}$ comes equipped with a $\mathbb{Z}_2$-grading; i.e., we can write $\mathcal{A}$ as a direct sum of vector spaces $\mathcal{A} = \mathcal{A}^{\mathrm{e}} \oplus \mathcal{A}^{\mathrm{o}}$, where $\mathcal{A}^{\mathrm{e}}$ and $\mathcal{A}^{\mathrm{o}}$ denote the sets of even and odd operators, respectively. An operator is even (resp., odd) if it can be written as a sum of even (resp., odd) monomials in $a_1, \ldots, a_M, a_1^\dagger, \ldots, a_M^\dagger$. (The reader can check that this notion is well-defined.) The $\mathbb{Z}_2$-grading refers to the fact that $\mathcal{A}^{\mathrm{e}}\mathcal{A}^{\mathrm{e}} \subset \mathcal{A}^{\mathrm{e}}$, $\mathcal{A}^{\mathrm{o}}\mathcal{A}^{\mathrm{o}} \subset \mathcal{A}^{\mathrm{e}}$, $\mathcal{A}^{\mathrm{e}}\mathcal{A}^{\mathrm{o}} \subset \mathcal{A}^{\mathrm{o}}$, and $\mathcal{A}^{\mathrm{o}}\mathcal{A}^{\mathrm{e}} \subset \mathcal{A}^{\mathrm{o}}$.

For any subset $C \subset \{1, \ldots, M\}$. Let $\mathcal{A}_C$ denote the subalgebra

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

and let the even and odd components $\mathcal{A}_C^{\mathrm{e}}$ and $\mathcal{A}_C^{\mathrm{o}}$ be defined accordingly. Suppose that our site index set is written as a union of cluster index sets C_γ, i.e.,

$$\{1, \ldots, M\} = \bigcup_{\gamma=1}^{N_{\mathrm{c}}} C_\gamma,$$

where the cluster index sets C_γ are disjoint, for simplicity.

We comment that, in contrast to our exposition for the case of quantum spins, we shall directly work with general clusters (as opposed to clusters consisting of a single site). The reason is that in the quantum spin setting, it was possible to view nonoverlapping clusters as single sites (with enlarged local state spaces). Such a reduction is not natural in the fermionic setting. Hence we retain the index notation γ, δ for clusters and i, j for individual sites (of which the clusters are comprised).

[1] A star-algebra over $\mathbb{C}$ is essentially an associative algebra over $\mathbb{C}$ in which one can take adjoints, where the adjoints satisfy their usual algebraic properties. "Unital" means that $1 \in \mathcal{A}$. For further details, see, e.g., [4]. We will use no deep results from the theory of star-algebras but nonetheless find the perspective to be clarifying. Specifically, it is useful to view our algebra of fermionic operators independently from any Fock space on which it acts, and in fact the notion of the Fock space does not play any explicit role in our developments.

We assume the Hamiltonian $\widehat{H} \in \mathcal{A}$ can be written as a sum of one-cluster and two-cluster operators as

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

where $\widehat{H}_{\gamma} \in \mathcal{A}_{C_\gamma}$ and $\widehat{H}_{\gamma\delta} \in \mathcal{A}_{C_\gamma \cup C_\delta}$.

Note carefully for context that the subalgebra $\mathcal{A}_{C_\gamma}$ corresponds in our earlier setting of quantum spin systems to the subalgebra of operators on $\bigotimes_{i\in C_\gamma} Q_i$, viewed as operators on $\mathcal{Q}$ by tensoring with the identity. Clearly, even by viewing the fermionic system as a spin system via JWT, this subalgebra is inequivalent to the fermionic subalgebra above defined. The reader should keep this perspective on the developments of Section 2 in mind as we transpose them to the fermionic setting.

Next we turn to defining our notion of a statistical ensemble and its marginals. For this task we turn to the language of star-algebras. The role of our full ensemble is played by the *state*, a linear functional $\omega : \mathcal{A} \to \mathbb{C}$ such that $\omega(1) = 1$ and $\omega(A^\dagger A) \geq 0$ for any $A \in \mathcal{A}$. In our setting (which is finite-dimensional), the action of a state can be viewed as nothing more than tracing against a density operator on the Fock space, as can be verified readily via the Riesz representation theorem. In the quantum spin setting of Section 2, the action $\omega(\widehat{A})$ of the state corresponds to the trace $\mathrm{Tr}[\widehat{A}\rho]$ against the density operator ρ. For $\widehat{A}$ an operator on $\bigotimes_{i\in C} Q_i$, we have $\omega_C(\widehat{A}) = \omega(\widehat{A}) = \mathrm{Tr}[\widehat{A}\rho] = \mathrm{Tr}[\widehat{A}\rho_C]$, i.e., our notion of marginalization—applied to a cluster subalgebra in the quantum spin setting—precisely recovers the partial trace operation. However, the abstract perspective will be useful in defining the notion of a marginal because if we try to directly borrow the corresponding notion from the setting of quantum spins, i.e., the partial trace, then we find ourselves in need of a global JWT to proceed.

We let Ω denote the set of states on $\mathcal{A}$. Then in star-algebraic language, the N-particle ground-state energy $E_0(N)$ minimization problem is naturally recast as

$$E_0(N) = \inf_{\omega\in\Omega:\omega(\widehat{N})=N} \omega(\widehat{H}). \tag{3.2}$$

Next, our notion of a marginal in this setting is simply the restriction of a state to a subalgebra. That is, for a subset $C \subset \{1,\dots,M\}$, we define the marginal ω_C via

$$\omega_C := \omega|_{\mathcal{A}_C}.$$

Of course, ω_C is itself a state on $\mathcal{A}_C$. We let Ω_C denote the set of states on $\mathcal{A}_C$. Notice that, as follows immediately from the definition, these sets are *convex*.

3.3 The two-cluster-marginal SDP

In this section we shall derive an "abstract SDP" without describing how it can be realized on a computer. Later, in Section 3.5, we will describe how to achieve such realization (which makes use of JWTs only for each pair of clusters). For simplicity, we will only derive a relaxation that analogizes the (*nonoverlapping*)

two-cluster-marginal SDP. Further analogues can be derived by straightforward (though perhaps tedious) modifications of the arguments presented below.

For simplicity we denote the one-cluster marginals by $\omega_\gamma := \omega_{C_\gamma}$ and the two-cluster marginals by $\omega_{\gamma\delta} := \omega_{C_\gamma \cup C_\delta}$. Note carefully from the definitions here that $\omega_{\gamma\delta} = \omega_{\delta\gamma} : \mathcal{A}_{C_\gamma \cup C_\delta} \to \mathbb{C}$ and that $\omega_{\gamma\gamma} = \omega_\gamma : \mathcal{A}_{C_\gamma} \to \mathbb{C}$. Our one- and two-cluster marginals evidently satisfy the local consistency constraints

$$\omega_\gamma = \omega_{\gamma\delta}|_{\mathcal{A}_{C_\gamma}}, \quad \omega_\delta = \omega_{\gamma\delta}|_{\mathcal{A}_{C_\delta}},$$

via nested restriction operations. By analogy to (2.4), our semidefinite constraint will be derived from the observation that for any $\widehat{A} \in \mathcal{A}$ of the form $\widehat{A} = \sum_\gamma \widehat{A}_\gamma$, where $\widehat{A}_\gamma \in \mathcal{A}_{C_\gamma}$ for all γ,

$$0 \le \omega(\widehat{A}^\dagger \widehat{A}) = \omega\Big(\Big[\sum_\gamma \widehat{A}_\gamma\Big]^\dagger \Big[\sum_\delta \widehat{A}_\delta\Big]\Big) = \sum_{\gamma\delta} \omega(\widehat{A}_\gamma^\dagger \widehat{A}_\delta).$$

Therefore the two-cluster marginals satisfy

$$\sum_{\gamma\delta} \omega_{\gamma\delta}(\widehat{A}_\gamma^\dagger \widehat{A}_\delta) \ge 0$$

for all choices of $\{A_\gamma\}_{\gamma=1}^{N_c}$ for which $\widehat{A}_\gamma \in \mathcal{A}_{C_\gamma}$ for all γ.

More specifically, for each cluster γ consider a *list* $\{\widehat{A}_{\gamma,\alpha}\}_{\alpha \in \mathcal{I}_\gamma}$ of operators in $\mathcal{A}_{C_\gamma}$, possibly (but not necessarily) spanning the space of all operators in $\mathcal{A}_{C_\gamma}$. (Compare to the perspective of Section 2.4 on the global semidefinite constraints in the quantum spin setting.) Then one obtains $G[\{\omega_{\gamma\delta}\}] \succeq 0$, where $G = (G_{\gamma\delta})$ is specified blockwise by

$$(G_{\gamma\delta}[\omega_{\gamma\delta}])_{\alpha\beta} = \omega_{\gamma\delta}(\widehat{A}_{\gamma,\alpha}^\dagger \widehat{A}_{\delta,\beta}).$$

In fact, $G = G[\{\omega_{\gamma\delta}\}_{\gamma \le \delta}]$ depends only on $\omega_{\gamma\delta}$ for $\gamma \le \delta$ because the lower triangular part can be obtained from the upper triangular part via hermiticity.

Then we have derived the following relaxation of the variational principle (3.2), in which the ω_γ and $\omega_{\gamma\delta}$ are considered as optimization variables:

$$\begin{aligned} E_0^{(2)}(N) := \underset{\{\omega_\gamma\},\{\omega_{\gamma\delta}\}_{\gamma<\delta}}{\text{minimize}} \quad & \sum_\gamma \omega_\gamma\Big(\widehat{H}_\gamma\Big) + \sum_{\gamma<\delta} \omega_{\gamma\delta}\Big(\widehat{H}_{\gamma\delta}\Big), \\ \text{subject to} \quad & \omega_{\gamma\delta} \in \Omega_{C_\gamma \cup C_\delta}, \quad 1 \le \gamma < \delta \le N_c, \\ & \omega_\gamma = \omega_{\gamma\delta}|_{\mathcal{A}_{C_\gamma}}, \quad 1 \le \gamma < \delta \le N_c, \\ & \omega_\delta = \omega_{\gamma\delta}|_{\mathcal{A}_{C_\delta}}, \quad 1 \le \gamma < \delta \le N_c, \\ & N = \sum_\gamma \omega_\gamma(\widehat{N}_\gamma), \\ & G\big[\{\omega_{\gamma\delta}\}_{\gamma\le\delta}\big] \succeq 0, \end{aligned} \tag{3.3}$$

where $\widehat{N}_\gamma := \sum_{i \in C_\gamma} \widehat{n}_i$ denotes the γ^{th} cluster number operator. Since the constraints are convex, we have specified an abstract convex optimization problem.

Now that we know that this relaxation makes sense in principle, our hope is to express it later as a concrete semidefinite program.

It is computationally useful to realize a simplification. Physical fermionic Hamiltonians are always even (including the anomalous, or particle-number-*non*conserving, Hamiltonians that arise in effective descriptions of superconductivity), and hence one expects the action of a physical state on an *odd* operator in fact always yields zero. Hence

$$(G_{\gamma\delta}[\omega_{\gamma\delta}])_{\alpha\beta} = \omega_{\gamma\delta}\big(\widehat{A}^{\dagger}_{\gamma,\alpha}\widehat{A}_{\delta,\beta}\big)$$

is zero unless $\widehat{A}_{\gamma,\alpha}$ and $\widehat{A}_{\delta,\beta}$ are either both even or both odd. It follows that we can reduce the size of the semidefinite constraint by splitting our operator lists into even and odd subsets, which we denote by

$$\big\{\widehat{A}^{\mathrm{e}}_{\gamma,\alpha}\big\}_{\alpha\in\mathcal{I}^{\mathrm{e}}_{\gamma}} \quad \text{and} \quad \big\{\widehat{A}^{\mathrm{o}}_{\gamma,\alpha}\big\}_{\alpha\in\mathcal{I}^{\mathrm{o}}_{\gamma}},$$

respectively. Then we define separate matrices G^{e} and G^{o} blockwise by

$$\big(G^{\mathrm{e/o}}_{\gamma\delta}[\omega_{\gamma\delta}]\big)_{\alpha\beta} = \omega_{\gamma\delta}\big(\big[\widehat{A}^{\mathrm{e/o}}_{\gamma,\alpha}\big]^{\dagger}\big[\widehat{A}^{\mathrm{e/o}}_{\delta,\beta}\big]\big). \tag{3.4}$$

Then we may equivalently substitute our semidefinite constraint $G \succeq 0$ with two semidefinite constraints $G^{\mathrm{e/o}} \succeq 0$, each of half (assuming that complete operator lists are chosen) the original size.

3.4 Exactness for noninteracting problems

In this section we assume that $\widehat{H}$ is noninteracting, i.e., of the form $\widehat{H} = \sum_{ij} h_{ij}\, a_i^{\dagger} a_j$, where $h = (h_{ij})$ is Hermitian. We want to show that in this setting $E_0^{(2)}(N) = E_0(N)$, i.e., the relaxation just introduced is tight, under the meager further assumption that for each $i \in \{1, \ldots, M\}$, the operators $a_i, a_i^{\dagger}$ are contained in some cluster's operator list.

Indeed, under this latter assumption it is not hard to see that the matrices

$$D(\omega_{\{i,j\}}) := \big(\omega_{\{i,j\}}\big(a_i^{\dagger}a_j\big)\big)_{i,j=1}^{M} \quad \text{and} \quad D'(\omega_{\{i,j\}}) := \big(\omega_{\{i,j\}}\big(a_i a_j^{\dagger}\big)\big)_{i,j=1}^{M}$$

appear as principal submatrices of $G^{\mathrm{o}}[\{\omega_{\gamma\delta}\}]$, where the two-site marginals $\omega_{\{i,j\}}$ are suitably obtained in terms of the two-cluster marginals $\omega_{\gamma\delta}$ by appropriate restriction. Note that by the fermionic anticommutation relations, in fact

$$D'(\omega_{\{i,j\}}) = I_M - D(\omega_{\{i,j\}})^{\top}.$$

Hence for any feasible solution to our SDP, we have $0 \preceq D(\omega_{\{i,j\}}) \preceq I_M$. Then it follows that $E_0^{(2)}(N)$ is an upper bound for the optimal value $E_0'(N)$ of the following (further relaxed) SDP:

$$\begin{aligned} 2E_0'(N) \;:=\; & \underset{D\in\mathbb{C}^{M\times M}}{\text{minimize}} && \mathrm{Tr}[D^{\top}h] \\ & \text{subject to} && 0 \preceq D \preceq I_M, \\ & && \mathrm{Tr}[D] = N. \end{aligned}$$

On the other hand, $E_0'(N) = \sum_{i=1}^{N} \lambda_i(h)$, where $\lambda_i(h)$ indicates the i^{th} lowest eigenvalue of h. For noninteracting problems this is precisely the value of $E_0(N)$. Hence we have shown $E_0(N) \geq E_0^{(2)}(N) \geq E_0'(N) = E_0(N)$, from which it follows that $E_0^{(2)}(N) = E_0(N)$.

For certain problems, one may also expect asymptotic tightness in the limit of strong interaction. For example, in the $t \to 0$ (or equivalently, $U \to \infty$) limit of the Hubbard model, the sites completely decouple, and it can be checked readily that our SDP is tight in this scenario.

3.5 Concrete perspective

In order to represent $\omega_{\gamma\delta}$ in concrete terms, note that $\omega_{\gamma\delta}$ is defined by its action on $\mathcal{A}_{C_\gamma \cup C_\delta}$. It is at this point that we introduce for computational purposes the JWT, though only for restricted fermionic algebras. Let $\mathrm{End}(V)$ denote the set of all endomorphisms of V. After specifying ordering the sites of $C_\gamma \cup C_\delta$, i.e., a labeling map $\kappa_{\gamma\delta} : C_\gamma \cup C_\delta \to \{1, \ldots, L_{\gamma\delta}\}$ where $L_{\gamma\delta} := |C_\gamma \cup C_\delta|$, the corresponding JWT fixes an algebra isomorphism

$$\mathcal{J}_{\gamma\delta} : \mathcal{A}_{C_\gamma \cup C_\delta} \to \mathrm{End}\left(\bigotimes_{i=1}^{L_{\gamma\delta}} \mathbb{C}^2\right),$$

and we define $c_{\kappa(i)}^{\gamma\delta} \in \mathrm{End}(\bigotimes_{i=1}^{L_{\gamma\delta}} \mathbb{C}^2)$ to be the image of a_i under this isomorphism for $i \in C_\gamma \cup C_\delta$. More specifically, the transformation $\mathcal{J}_{\gamma\delta}$ is specified by setting

$$\mathcal{J}_{\gamma\delta}\left(a_{\kappa_{\gamma\delta}^{-1}(i)}\right) = c_i^{\gamma\delta},$$

where

$$c_i^{\gamma\delta} := \underbrace{\sigma^z \otimes \cdots \otimes \sigma^z}_{(i-1)\text{ factors}} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \otimes \underbrace{I_2 \otimes \cdots \otimes I_2}_{(L_{\gamma\delta}-i)\text{ factors}}.$$

Notice that the case $\gamma = \delta$ makes perfect sense according to the above definitions, though we will also introduce the alternative notation $\mathcal{J}_\gamma := \mathcal{J}_{\gamma\gamma}$.

Let $\mathrm{Id} \in \mathrm{End}(\bigotimes_{i=1}^{L_{\gamma\delta}} \mathbb{C}^2)$ be the identity operator. Then $F_{\gamma\delta} := \mathcal{J}_{\gamma\delta} \circ \omega_{\gamma\delta} \circ \mathcal{J}_{\gamma\delta}^{-1}$ is a linear functional on $\mathrm{End}(\bigotimes_{i=1}^{L_{\gamma\delta}} \mathbb{C}^2)$ satisfying $F_{\gamma\delta}(\mathrm{Id}) = 1$ and $F_{\gamma\delta}(A^\dagger A) \geq 0$ for any $A \in \mathrm{End}(\bigotimes_{i=1}^{L_{\gamma\delta}} \mathbb{C}^2)$. It follows (via the Riesz representation theorem) that there exists a unique $\rho_{\gamma\delta} \succeq 0$ with $\mathrm{Tr}[\rho_{\gamma\delta}] = 1$ such that $F_{\gamma\delta}(A) = \mathrm{Tr}[A\rho_{\gamma\delta}]$ for all $A \in \mathrm{End}(\bigotimes_{i=1}^{L_{\gamma\delta}} \mathbb{C}^2)$. That is to say, $\omega_{\gamma\delta}(\widehat{A}) = \mathrm{Tr}[A\rho_{\gamma\delta}]$ whenever $A = \mathcal{J}_{\gamma\delta}(\widehat{A})$. Again, we introduce the alternative notation $\rho_\gamma = \rho_{\gamma\gamma}$ for conceptual clarity.

Motivated by the preceding, we shall replace optimization over *states* $\omega_{\gamma\delta} : \mathcal{A}_{C_\gamma \cup C_\delta} \to \mathbb{C}$ with optimization over *density operators*

$$\rho_{\gamma\delta} \in \mathrm{End}\left(\bigotimes_{i=1}^{L_{\gamma\delta}} \mathbb{C}^2\right).$$

Crucially, the correspondence between states and density operators has relied on a separate JWT for *each* pair (γ, δ), not a single global JWT that maps the global fermionic state to a global density operator. Neither should we obtain $\rho_{\gamma\delta}$ from a global density operator ρ via the standard definition of the partial trace, as in the case of quantum spin systems.

Under this correspondence $G^{\mathrm{e/o}}_{\gamma\delta}[\omega_{\gamma\delta}]$ as defined by (3.4) can be obtained as

$$\left(G^{\mathrm{e/o}}_{\gamma\delta}[\rho_{\gamma\delta}]\right)_{\alpha\beta} = \mathrm{Tr}\left(\left[\mathcal{J}_{\gamma\delta}\left(\hat{A}^{\mathrm{e/o}}_{\gamma,\alpha}\right)\right]^\dagger \left[\mathcal{J}_{\gamma\delta}\left(\hat{A}^{\mathrm{e/o}}_{\delta,\beta}\right)\right]\rho_{\gamma\delta}\right),$$

where we abuse notation slightly by identifying $G^{\mathrm{e/o}}_{\gamma\delta}[\rho_{\gamma\delta}]$ with $G^{\mathrm{e/o}}_{\gamma\delta}[\omega_{\gamma\delta}]$.

In order to write down a concrete realization of the optimization problem (3.3), the remaining hurdle is to encode the local consistency constraints $\omega_\gamma = \omega_{\gamma\delta}|_{\mathcal{A}_{C_\gamma}}$ and $\omega_\delta = \omega_{\gamma\delta}|_{\mathcal{A}_{C_\delta}}$ for $\gamma < \delta$, which require us to further "marginalize" our fermionic states.

We did this by first assuming that the labeling map $\kappa_{\gamma\delta}$ satisfies $\kappa_{\gamma\delta}(C_\gamma) < \kappa_{\gamma\delta}(C_\delta)$ in the sense that every element of the left-hand side is less than every element of the right-hand side. In the case of overlapping clusters, which (as previously mentioned) we shall not discuss in full detail, the relevant generalization ensures that $\kappa_{\gamma\delta}(C_\gamma) < \kappa_{\gamma\delta}([C_\gamma \cup C_\delta] \setminus C_\gamma)$. For simplicity we also assume that $\kappa_{\gamma\delta}|_{C_\gamma} = \kappa_{\gamma\gamma}$ for all $\gamma < \delta$, and from now on we think of the labeling maps $\kappa_{\gamma\delta}$ as fixed. It is always possible to choose a labeling that satisfies these assumptions.

Then it follows from the definition of the JWT that for any $\hat{A} \in \mathcal{A}_{C_\gamma}$, $A := \mathcal{J}_{\gamma\delta}(\hat{A})$ is of the form

$$A = B \otimes \mathrm{Id}_{\bigotimes_{i=1}^{|C_\delta|} \mathbb{C}^2} = B \otimes \underbrace{I_2 \otimes \cdots \otimes I_2}_{|C_\delta| \text{ factors}},$$

where $B = \mathcal{J}_\gamma(\hat{A}) \in \mathrm{End}(\bigotimes_{i=1}^{L_\gamma} \mathbb{C}^2)$. Then

$$\omega_{\gamma\delta}(\hat{A}) = \mathrm{Tr}[A\rho_{\gamma\delta}] = \mathrm{Tr}[B\tilde{\rho}_\gamma],$$

where $\tilde{\rho}_\gamma := \mathrm{Tr}_{\kappa_{\gamma\delta}(C_\gamma)}[\rho_{\gamma\delta}]$. Meanwhile, we have $\mathcal{J}_\gamma(\hat{A}) = B$, and $\omega_\gamma(\hat{A}) = \mathrm{Tr}[B\rho_\gamma]$. Hence the constraint $\omega_\gamma = \omega_{\gamma\delta}|_{\mathcal{A}_{C_\gamma}}$ for $\gamma < \delta$ is equivalent to the stipulation that $\mathrm{Tr}[B\tilde{\rho}_\gamma] = \mathrm{Tr}[B\rho_\gamma]$ for all B, i.e., that

$$\rho_\gamma = \mathrm{Tr}_{\kappa_{\gamma\delta}(C_\delta)}[\rho_{\gamma\delta}].$$

Here $\mathrm{Tr}_{\kappa_{\gamma\delta}(C_\delta)}(\cdot)$ *is* the standard partial trace.

Meanwhile, for any $A = \mathcal{J}_{\gamma\delta}(\widehat{A})$ where $\widehat{A} \in \mathcal{A}^{\mathrm{e}}_{C_\delta}$ is *even*, we can write

$$A = \underbrace{I_2 \otimes \cdots \otimes I_2}_{|C_\gamma| \text{ factors}} \otimes B,$$

where $B = \mathcal{J}_\delta(\widehat{A}) \in \mathrm{End}(\bigotimes_{i=1}^{L_\delta} \mathbb{C}^2)$. Hence for all $\widehat{A} \in \mathcal{A}^{\mathrm{e}}_{C_\delta}$, we derive as above that $\omega_\delta(\widehat{A}) = \mathrm{Tr}[B\widetilde{\rho}_\delta]$, where $\widetilde{\rho}_\delta := \mathrm{Tr}_{\kappa_{\gamma\delta}(C_\delta)}[\rho_{\gamma\delta}]$. But for $\widehat{A} \in \mathcal{A}^{\mathrm{o}}_{C_\delta}$, as mentioned above we can assume $\omega_{\gamma\delta}(\widehat{A}) = \omega_\delta(\widehat{A}) = 0$ (because this identity is a necessary condition satisfied by the exact marginals) and hence also that $\mathrm{Tr}[B\rho_\delta] = 0 = \mathrm{Tr}[B\widetilde{\rho}_\delta]$ for all $B \in \mathcal{J}_\delta(\mathcal{A}^{\mathrm{o}}_{C_\delta})$. Thus the constraint $\omega_\delta = \omega_{\gamma\delta}|_{\mathcal{A}_{C_\delta}}$ for $\gamma < \delta$ is equivalent to the stipulation that $\mathrm{Tr}[B\widetilde{\rho}_\delta] = \mathrm{Tr}[B\rho_\delta]$ for all B, i.e., that

$$\rho_\delta = \mathrm{Tr}_{\kappa_{\gamma\delta}(C_\delta)}[\rho_{\gamma\delta}].$$

Finally, note that the constraint $\mathrm{Tr}[\rho_{\gamma\delta}] = 1$ can simply be encoded, given our first local consistency constraint, by $\mathrm{Tr}[\rho_\gamma] = 1$. Then we obtain the following concrete realization of (3.3):

$$\begin{aligned} E_0^{(2)}(N) := \underset{\{\rho_\gamma\},\{\rho_{\gamma\delta}\}_{\gamma<\delta}}{\text{minimize}} \quad & \sum_\gamma \mathrm{Tr}\big[\mathcal{J}_\gamma\big(\widehat{H}_\gamma\big)\rho_\gamma\big] + \sum_{\gamma<\delta} \mathrm{Tr}\big[\mathcal{J}_{\gamma\delta}\big(\widehat{H}_{\gamma\delta}\big)\rho_{\gamma\delta}\big], \\ \text{subject to} \quad & \rho_{\gamma\delta} \succeq 0, \quad 1 \le \gamma < \delta \le N_{\mathrm{c}}, \\ & \rho_\gamma = \mathrm{Tr}_{\kappa_{\gamma\delta}(C_\delta)}[\rho_{\gamma\delta}], \quad 1 \le \gamma < \delta \le N_{\mathrm{c}}, \\ & \rho_\delta = \mathrm{Tr}_{\kappa_{\gamma\delta}(C_\gamma)}[\rho_{\gamma\delta}], \quad 1 \le \gamma < \delta \le N_{\mathrm{c}}, \\ & \mathrm{Tr}[\rho_\gamma] = 1, \quad \gamma = 1, \ldots, N_{\mathrm{c}}, \\ & N = \sum_\gamma \mathrm{Tr}\big[\mathcal{J}_\gamma\big(\widehat{N}_\gamma\big)\rho_\gamma\big], \\ & G\big[\{\rho_{\gamma\delta}\}_{\gamma\le\delta}\big] \succeq 0. \end{aligned}$$

4 Numerical Results

All numerical results were computed in MATLAB® with CVX [13] for performing SDP calculations. We limit our experiments to problems that are small enough to validate by exact diagonalization. In particular, we will illustrate numerically the fact that all of our relaxations must yield lower bounds for the exact energy. We will also show that the omission of the global semidefinite constraints results in looser lower bounds; i.e., the global semidefinite constraints are nontrivial, even though the Hamiltonians are all local. As discussed in Section 5.3 below, a more scalable implementation should be possible, but such an implementation (as well as an accompanying numerical study of properties of larger systems, e.g., approaching a thermodynamic limit) will be left to future work.

4.1 Transverse-field Ising model

First we consider the transverse-field Ising (TFI) model (2.1) on a periodic 12×1 lattice, comparing results of the two-cluster-marginal SDP for various cluster sizes.

We also test the periodicity constraints of Section 2.8 and the case of overlapping clusters. The results are shown in Figure 4.1.

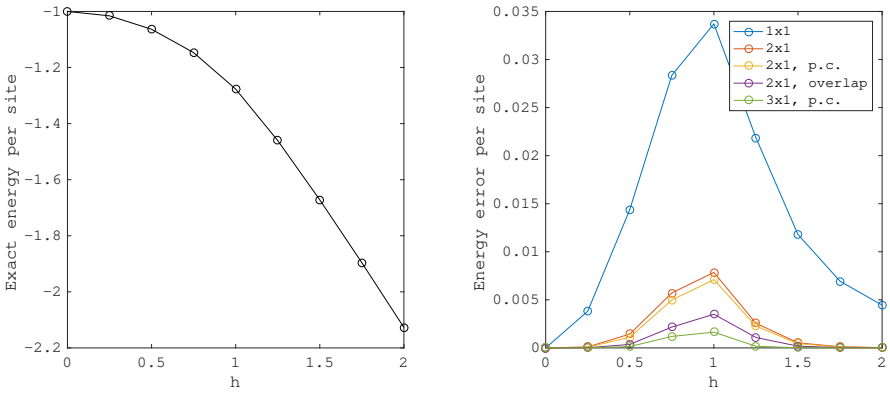

FIGURE 4.1. TFI model on periodic 12×1 lattice. Approximate energies are computed via the two-cluster-marginal relaxation. Note that "p.c." indicates the inclusion of the periodicity constraints introduced in Section 2.8, and "overlap" indicates the choice of overlapping 2×1 clusters, i.e., $\{1,2\},\{2,3\},\{3,4\},\ldots,\{11,12\},\{12,1\}$.

Note that, as the theory requires, all approximations do indeed yield lower bounds for the exact energy. Moreover, these bounds become tighter for larger cluster sizes. Also notice that the case of overlapping 2×1 clusters compares favorably to the case of nonoverlapping 2×1 clusters, achieving an energy error roughly twice as small. (In the case of overlapping clusters, the periodicity constraints of Section 2.8 are satisfied automatically by the solution, and there is no need to enforce them explicitly. Hence from Figure 4.1 it is clear that most of the improvement yielded by allowing for overlap is *not* merely due to these constraints.)

In Figure 4.2 we test the same relaxations on the same model problem except that we *omit the global semidefinite constraints*. Neglecting the global semidefinite constraints corresponds to the use of belief propagation (BP) [31] in the classical setting and its quantum generalization [2, 10, 21, 32]. Note that the omission of these constraints results in a significant degradation of the lower bound, even though the Hamiltonian is local.

Next we consider the TFI model on a periodic 4×3 square lattice, comparing results of the two-cluster-marginal SDP for various cluster sizes. The results are shown in Figure 4.3. Here we are more limited by the preliminary implementation in what can be tested, though the observations are compatible with those preceding remarks that are applicable. In Figure 4.4, we once again test the effect of removing the global semidefinite constraints, and similar conclusions apply.

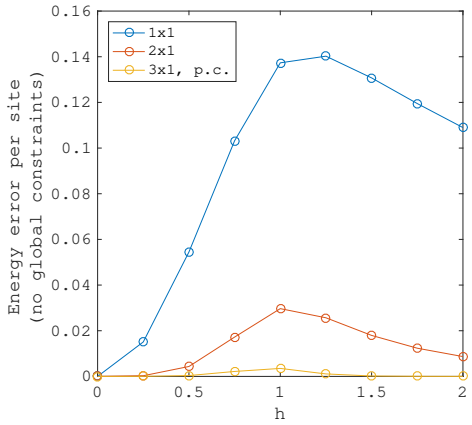

FIGURE 4.2. Results for the same model and same relaxations as in Figure 4.1, with the modification that the global semidefinite constraints are omitted in all cases. In this experiment the curves for "2×1, p.c." and "2×1, overlap" coincide with that of "2×1." Note the change of scale of the vertical axis relative to the analogous plot of Figure 4.1. For clarity, we remark that the value of the "3×1, p.c." curve at $h = 1$ is 0.0035, compared to the corresponding value (with global constraints active) of 0.0016 depicted in Figure 4.1.

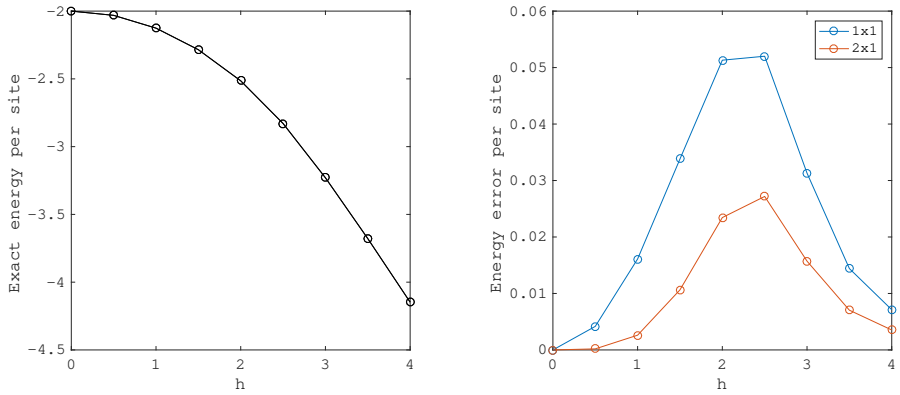

FIGURE 4.3. TFI model on periodic 4×3 lattice. Approximate energies are computed via the two-cluster-marginal relaxation.

4.2 Antiferromagnetic Heisenberg model

First we consider the antiferromagnetic Heisenberg model (2.2) on a periodic 12×1 lattice, comparing results of the two-cluster-marginal SDP for various cluster sizes. We also test the periodicity constraints of Section 2.8 and the case of

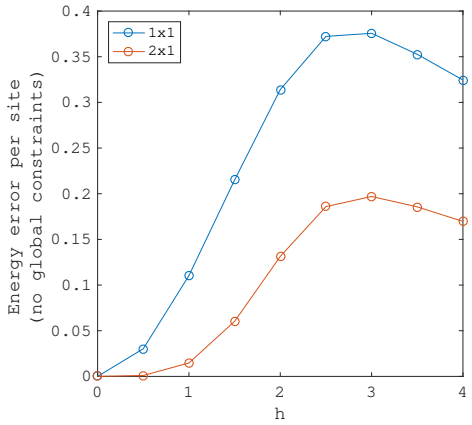

FIGURE 4.4. Results for the same model and same relaxations as in Figure 4.3, with the modification that the global semidefinite constraints are omitted in all cases. Note the change of scale of the vertical axis relative to the analogous plot of Figure 4.3.

overlapping clusters, as well as the effect of omitting the global semidefinite constraints. The results are shown in Table 4.1.

In Table 4.2 we show results for the AFH model on a periodic 4×3 lattice for various cluster sizes. For these experiments, the observations are qualitatively similar to those reported for the TFI model, though the relative energy errors are larger. In particular, the errors for 1×1 clusters are quite large, though the error falls dramatically as the cluster size is increased. Moreover, the global constraints achieve significant error reduction even though the Hamiltonian is local.

TABLE 4.1. Energy error by cluster specification for the AFH model on periodic 12×1 lattice, with and without global semidefinite constraints. For reference, the exact ground state energy is -1.7958. Approximate energies for the first line are computed via the two-cluster-marginal relaxation. Note that "p.c." indicates the inclusion of the periodicity constraints introduced in Section 2.8, and "overlap" indicates the choice of overlapping 2×1 clusters, i.e., $\{1,2\},\{2,3\},\{3,4\},\ldots,\{11,12\}$. For the results of the second line, the global semidefinite constraints were omitted.

	1×1	2×1	2×1, p.c.	2×1, overlap	3×1, p.c.
global	0.6017	0.0634	0.0462	0.0159	0.0048
no global	1.2042	0.2042	0.2042	0.2042	0.0310

TABLE 4.2. Energy error by cluster specification for the AFH model on periodic 4×3 lattice, with and without global semidefinite constraints. For reference, the exact ground state energy is -2.4561. Approximate energies for the first line are computed via the two-cluster-marginal relaxation. Approximate energies for the second line are obtained by omitting the global semidefinite constraints.

	1×1 clusters	2×1 clusters	1×3 clusters
global	1.0439	0.3937	0.0410
no global	3.5439	2.1897	0.8773

4.3 Hubbard model

Finally, we consider the Hubbard model (3.1) on a nonperiodic 8×1 lattice with particle numbers $N = 6, 7, 8, 9, 10$ and interaction strengths $U \in [0, 12]$. In Figure 4.5, we plot results for the two-cluster-marginal relaxation with 1×1 clusters $C_i := \{(i, \uparrow), (i, \downarrow)\}$. Observe that for $U = 0$, the system is noninteracting and the energy is exact, as guaranteed by the discussion in Section 3.4. Furthermore, the error of the energy decreases with respect to U (even without normalizing by U). We remark that the error of the energy per site is on par with that of DMET [16] when the same cluster sizes are used. In comparison to DMET, variational embedding is less accurate for intermediate U (i.e., $U \approx 4$) but scales more gracefully

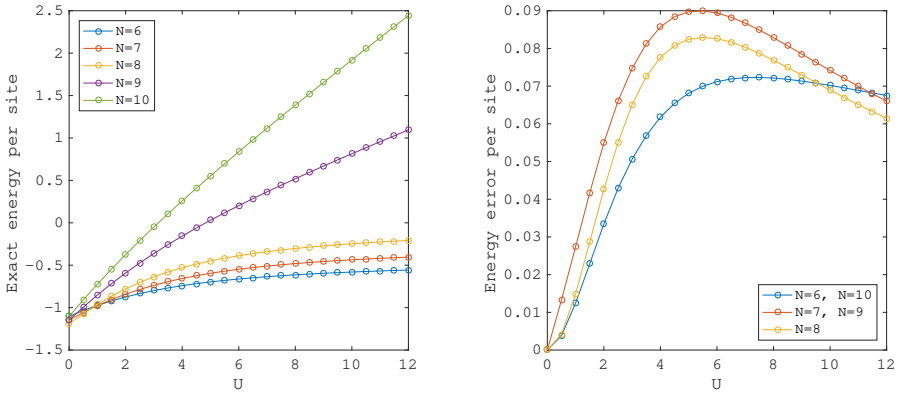

FIGURE 4.5. Hubbard model on nonperiodic 8×1 lattice. Approximate energies are computed via the two-cluster-marginal relaxation with 1×1 clusters $C_i := \{(i, \uparrow), (i, \downarrow)\}$. Note that the energy errors in the cases $N = 6$ and $N = 7$ coincide with the errors in the cases $N = 10$ and $N = 9$, respectively, due to the particle-hole symmetry.

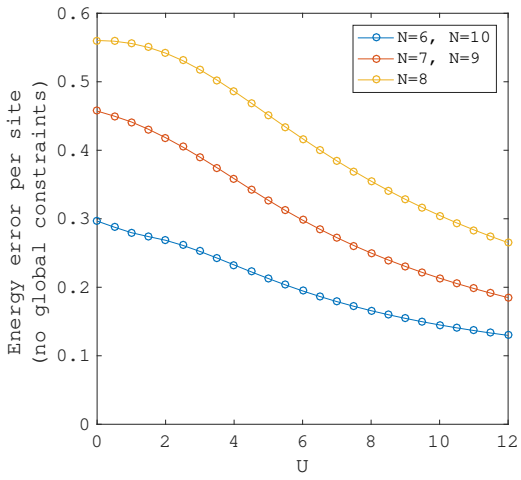

FIGURE 4.6. Results for the same model and same relaxation as in Figure 4.5, with the modification that the global semidefinite constraints are omitted in all cases. Note the change of scale of the vertical axis relative to the analogous plot of Figure 4.5.

in the regime of large U (i.e., $U \gtrsim 8$). However, a thorough comparison of variational embedding with other embedding methods will be a matter for future work following more careful implementation.

In Figure 4.6 we test the same relaxation on the same model problems, except that once again we *omit the global semidefinite constraints*. Once again we observe significant degradation of the lower bound. Note, moreover, that the omission of these constraints breaks the exactness of the relaxation energy for $U = 0$.

5 Duality and the Effective Hamiltonian Perspective

In order to reduce the computational cost for solving the SDP in the variational embedding (called the primal problem), we may consider the associated dual problem. For simplicity, we consider duality only for the two-marginal SDP in the quantum spin setting, and it will be convenient to take the "abstract perspective" of Section 2.4, with possibly restricted operator sets as in Remark 2.2. Duality in other settings can be approached by similar means.

5.1 The quantum Kantorovich problem

In preparation for our discussion of the duality of the two-marginal SDP, we first introduce the notion of the quantum Kantorovich problem, which is a direct quantum analogue (and in fact generalization) of the Kantorovich problem of optimal transport [41]. See also in [5, 7, 12, 34, 47] for related, though different, presentations.

The analogy to classical optimal transport is defined by replacing probability measures with density operators, a cost function with a cost operator $\widehat{C}$, and classical marginalization with quantum marginalization (i.e., the partial trace). Given operators $\mu_i \in \mathrm{End}(Q_i)$ for $i = 1, 2$ of unit trace, we may define the optimal quantum Kantorovich cost via the SDP

$$\begin{aligned} \mathbf{QK}[\widehat{C}\,;\mu_1,\mu_2] := \underset{\pi\in\mathrm{End}(Q_1\otimes Q_2)}{\text{minimize}} \quad & \mathrm{Tr}[C\pi] \\ \text{subject to} \quad & \pi \succeq 0 \\ & \mu_1 = \mathrm{Tr}_{\{2\}}[\pi],\ \mu_2 = \mathrm{Tr}_{\{1\}}[\pi]. \end{aligned}$$

Note that if $\mu_1 \not\succeq 0$ or $\mu_2 \not\succeq 0$, then since $\pi \succeq 0$ implies that $\mathrm{Tr}_{\{i\}}[\pi] \succeq 0$, the problem is infeasible, i.e., $\mathbf{QK}[C\,;\mu_1,\mu_2] = +\infty$. Hence without loss of generality one may assume that $\mu_i \succeq 0$, i.e., that the μ_i are indeed density operators on Q_i. Nonetheless, the slightly relaxed perspective will be of some use below. In fact, conversely, the program *is* feasible whenever $\mu_1, \mu_2 \succeq 0$ because in this case $\pi = \mu_1 \otimes \mu_2$ is a feasible point.

There is a notion of quantum Kantorovich duality that analogizes the usual notion, as follows. Let the *Hermitian* operators $A \in \mathrm{End}(Q_1)$ and $B \in \mathrm{End}(Q_2)$ be dual variables for the first and second marginal constraints, respectively. These will be the "quantum Kantorovich potentials." Dualizing these constraints yields the Lagrangian

$$\mathcal{L}_{\mathrm{QK}}(\pi, A, B) = \mathrm{Tr}[C\pi] + \mathrm{Tr}[A(\mu_1 - \mathrm{Tr}_{\{2\}}[\pi])] + \mathrm{Tr}[B(\mu_2 - \mathrm{Tr}_{\{1\}}[\pi])]$$

still constrained by $\pi \succeq 0$. Using the fact that $\mathrm{Tr}[A\,\mathrm{Tr}_{\{2\}}[\pi]] = \mathrm{Tr}[(A\otimes \mathrm{Id})\pi]$ and $\mathrm{Tr}[B\,\mathrm{Tr}_{\{1\}}[\pi]] = \mathrm{Tr}[(\mathrm{Id}\otimes B)\pi]$, we obtain

$$\mathcal{L}_{\mathrm{QK}}(\pi, A, B) = \mathrm{Tr}[A\mu_1] + \mathrm{Tr}[B\mu_2] + \mathrm{Tr}[(C - A\otimes \mathrm{Id} - \mathrm{Id}\otimes B)\pi].$$

Now for fixed A, B, we have

$$\inf_{\pi\succeq 0} \mathrm{Tr}[(C - A\otimes \mathrm{Id} - \mathrm{Id}\otimes B)\pi] = \begin{cases} 0, & C - A\otimes \mathrm{Id} - \mathrm{Id}\otimes B \succeq 0, \\ -\infty, & \text{otherwise.} \end{cases}$$

Hence we have derived the Kantorovich dual problem

$$\begin{aligned} \underset{A,B\ \text{Hermitian}}{\text{maximize}} \quad & \mathrm{Tr}[A\mu_1] + \mathrm{Tr}[B\mu_2] \\ \text{subject to} \quad & A\otimes \mathrm{Id} + \mathrm{Id}\otimes B \preceq C. \end{aligned} \tag{5.1}$$

Strong duality holds by Sion's minimax theorem [18] (together with the compactness of the feasible set of the primal problem).

Let π be the minimizer for the primal problem, and suppose that the dual problem admits a maximizer (A, B). Then let $M = C - A\otimes \mathrm{Id} - \mathrm{Id}\otimes B$, so

$$\begin{aligned} \mathrm{Tr}[M\pi] &= \mathrm{Tr}[C\pi] - \mathrm{Tr}[(A\otimes \mathrm{Id})\pi] - \mathrm{Tr}[(\mathrm{Id}\otimes B)\pi] \\ &= \mathrm{Tr}[C\pi] - \mathrm{Tr}[A\mu_1] - \mathrm{Tr}[B\mu_2] \\ &= 0, \end{aligned}$$

by primal and dual optimality. But $\pi \succeq 0$, so we can write $\pi = \sum_{i=1}^m p_i \phi_i \phi_i^*$ where $p_i > 0$, and $\mathrm{Tr}[M\pi] = \sum_{i=1}^m p_i \phi_i^* M \phi_i$. But also $M \succeq 0$, so $p_i \phi_i^* M \phi_i \geq 0$ for all $i = 1, \ldots, m$. Then since $\mathrm{Tr}[M\pi] = 0$ it follows that $\phi_i^* M \phi_i = 0$ for all $i = 1, \ldots, m$, and since $M \succeq 0$ this means that $M\phi_i = 0$ for all $i = 1, \ldots, m$.

Therefore π is a convex combination of orthogonal projectors onto mutually orthogonal, degenerate ground state eigenvectors of the Hamiltonian $C - A \otimes \mathrm{Id} - \mathrm{Id} \otimes B$. For the reader familiar with optimal transport, we remark that this observation generalizes the corresponding observation [41] in the classical setting on the support of the Kantorovich coupling, i.e., that $\pi_{ij} \geq 0$ only if $\phi_i + \psi_j = c_{ij}$, where $\pi = (\pi_{ij})$, $\phi = (\phi_i)$, and $\psi = (\psi_i)$ are the Kantorovich potentials, and $c = (c_{ij})$ is the cost matrix.

In fact, one can consider a regularization of the primal problem by a von Neumann entropy penalty (scaled by β), for which the solution can be shown to be of the form

$$\pi_\beta = \frac{1}{Z_\beta} \exp[-\beta(C - A_\beta \otimes \mathrm{Id} - \mathrm{Id} \otimes B_\beta)],$$

where A_β and B_β are the unique operators chosen to yield the desired marginals μ_1, μ_2. This is the quantum analogy of the entropic regularization of classical optimal transport [8]. In the "zero-temperature" limit $\beta \to \infty$ one expects $\pi_\beta \to \pi$, $A_\beta \to A$, and $B_\beta \to B$.

5.2 Partial duality

Before any derivations, we comment that strong duality (i.e., the fact that there is zero gap between the optimal values of the primal and dual problems for the two-marginal SDP) can be understood as follows. In the original primal problem (2.5), the feasible domain for $\{\rho_i\}$, $\{\rho_{ij}\}_{i<j}$ in this problem is compact, so strong duality holds simply by Sion's minimax theorem [18]. The question of whether the dual optimizer is attained is more subtle and will be deferred to future work, though see [14] for the discussion of strong duality in a similar setting.

Now we turn to the derivation of the partial dual problem. We adopt the "abstract" perspective on the global semidefinite constraints introduced in Section 2.4, as well as the notation of that section. Referring to (2.5), we first consider a *partial* Lagrangian obtained by dualizing *only* the constraint (2.9):

$$\mathcal{L}_{\mathrm{part}}(\{\rho_i\}, \{\rho_{ij}\}, X) = \sum_i \mathrm{Tr}[H_i \rho_i] + \sum_{i<j} \mathrm{Tr}[H_{ij} \rho_{ij}] - \mathrm{Tr}(G[\{\rho_{ij}\}]\, X),$$

whose domain is defined by $X \in \mathbb{C}^{(\sum_i m_i^2) \times (\sum_i m_i^2)}$ Hermitian positive semidefinite and $\{\rho_i\}, \{\rho_{ij}\}$ satisfying constraints (2.6), (2.7), and (2.8).

Now

$$\begin{aligned}\operatorname{Tr}\big(G[\{\rho_{ij}\}]\,X\big) &= \sum_{ij}\operatorname{Tr}\big(G_{ij}[\rho_{ij}]\,X_{ji}\big)\\ &= \sum_{i}\sum_{\alpha\beta}\operatorname{Tr}\big[\rho_i\,O_{i,\alpha}^{\dagger}O_{i,\beta}\big](X_{ii})_{\beta\alpha}\\ &\quad+\sum_{i\neq j}\sum_{\alpha\beta}\operatorname{Tr}\big[\rho_{ij}\big(O_{i,\alpha}^{\dagger}\otimes O_{j,\beta}\big)\big](X_{ji})_{\beta\alpha}\\ &= \sum_{i}\sum_{\alpha\beta}\operatorname{Tr}\big[\rho_i\,O_{i,\alpha}^{\dagger}O_{i,\beta}\big](X_{ii})_{\beta\alpha}\\ &\quad+\sum_{i<j}\sum_{\alpha\beta}\big\{\operatorname{Tr}\big[\rho_{ij}\big(O_{i,\alpha}^{\dagger}\otimes O_{j,\beta}\big)\big](X_{ji})_{\beta\alpha}\\ &\qquad\qquad+\operatorname{Tr}\big[\rho_{ji}\big(O_{j,\beta}^{\dagger}\otimes O_{i,\alpha}\big)\big](X_{ij})_{\alpha\beta}\big\}.\end{aligned}$$

Now by the hermiticity of X we have $(X_{ji})_{\beta\alpha} = \overline{(X_{ij})_{\alpha\beta}}$, and we also have the identity

$$\operatorname{Tr}\big[\rho_{ji}\big(O_{j,\beta}^{\dagger}\otimes O_{i,\alpha}\big)\big] = \operatorname{Tr}\big[\rho_{ij}\big(O_{i,\alpha}\otimes O_{j,\beta}^{\dagger}\big)\big].$$

Therefore

$$\operatorname{Tr}\big(G[\{\rho_{ij}\}]\,X\big) = \sum_{i}\operatorname{Tr}[Y_i(X_{ii})\,\rho_i] + \sum_{i<j}\operatorname{Tr}\big[Y_{ij}(X_{ij})\,\rho_{ij}\big],$$

where we have defined the functions

$$Y_i : \mathbb{C}^{|\mathcal{I}_i|\times|\mathcal{I}_i|}\to\operatorname{End}(Q_i) \quad\text{and}\quad Y_{ij} : \mathbb{C}^{|\mathcal{I}_i|\times|\mathcal{I}_j|}\to\operatorname{End}(Q_i\otimes Q_j)$$

by

$$Y_i(M) = \sum_{\alpha\beta}\overline{M}_{\alpha\beta}\,O_{i,\alpha}^{\dagger}O_{i,\beta}, \quad Y_{ij}(M) = \Big[\sum_{\alpha\beta}\overline{M}_{\alpha\beta}\big(O_{i,\alpha}^{\dagger}\otimes O_{j,\beta}\big)\Big] + \text{h.c.},$$

where "h.c." denotes the Hermitian conjugate. Note that if M is Hermitian, then $Y_i(M)$ is Hermitian as well; hence $Y_i(X_{ii})$ and $Y_{ij}(X_{ij})$ are Hermitian operators.

By applying Sion's minimax theorem [18] and then separating the infimum over $\{\rho_i\}, \{\rho_{ij}\}$ into an outer infimum over $\{\rho_i\}$ (subject to constraint (2.8)) and an inner infimum over $\{\rho_{ij}\}$ (subject to constraints (2.6) and (2.7)), we may rewrite the two-marginal SDP energy as

$$E_0^{(2)} = \sup_{X\succeq 0}\ \inf_{\{\rho_i\}:\operatorname{Tr}[\rho_i]=1,\forall i}\ \mathcal{F}(X,\{\rho_i\}), \tag{5.2}$$

where

$$\mathcal{F}(X,\{\rho_i\}) := \sum_i \mathrm{Tr}[(H_i - Y_i(X_{ii}))\rho_i] + \sum_{i<j} \mathbf{QK}[H_{ij} - Y_{ij}(X_{ij}); \rho_i, \rho_j]. \tag{5.3}$$

This is the form of a concave-convex maxmin problem. The effective domain of the minimization over $\{\rho_i\}$ is in fact specified by the constraints $\mathrm{Tr}[\rho_i] = 1$, $\rho_i \succeq 0$ for all i, because if $\rho_i \not\succeq 0$ for some i, then at least one of the quantum Kantorovich problems in the expression for $\mathcal{F}(X,\{\rho_i\})$ is infeasible, i.e., of infinite optimal cost. The significance of this form is that for fixed $X, \{\rho_i\}$, the two-marginals ρ_{ij} have been entirely decoupled from one another in the evaluation of $\mathcal{F}(X,\{\rho_i\})$. Moreover, for each pair $i < j$, we see the emergence of the effective Hamiltonians $H_i^{\mathrm{eff}}(X_{ii}) := H_i - Y_i(X_{ii})$ and $H_{ij}^{\mathrm{eff}}(X_{ij}) := H_{ij} - Y_{ij}(X_{ij})$ on Q_i and $Q_i \otimes Q_j$, respectively. Notice that the new contributions to these effective Hamiltonians are linear combinations of operators of the form $O_{i,\alpha}^\dagger O_{i,\beta}$ and $O_{i,\alpha}^\dagger \otimes O_{i,\beta}$, respectively. Thus we see how our choice of effective operator lists is reflected in the richness of our class of possible effective Hamiltonians.

5.3 Computational perspective

From the computational point of view, the partial dual formulation can be much more efficient to solve than the primal formulation. Although general results guarantee that the complexity of solving the two-marginal SDP (2.5) is only polynomial in M, direct solution of the primal problem (by, e.g., interior-point methods) may still scale quite poorly in practice. One might hope that the complexity should be limited only by $O(M^3)$ per iteration, i.e., the cost of diagonalizing a matrix of size proportional to M, since the SDP constraint (2.9) concerns a matrix of size proportional to M. However, since the semidefinite matrix G is entangled with further equality constraints, the best guarantees for interior-point methods are far more pessimistic. One can interpret our discussion of duality thus far as revealing a special structure of these equality constraints that allows us in principle to design methods achieving a cost of $O(M^3)$ per iteration. (We remark that similar considerations could be expected to achieve a cost of $O(M)$ per iteration for the quasi-local two-marginal SDP with fixed $d_{\max}$, as described in Remark 2.3, though we omit details for simplicity.)

Now we describe how to compute gradients of $\mathcal{F}(X,\{\rho_i\})$ in order to apply, e.g., gradient ascent-descent methods. For fixed $X, \{\rho_i\}$, let $(A_{ij}^\star, B_{ij}^\star)$ be the unique dual optimizer (assuming that it exists) for the Kantorovich dual formulation of $\mathbf{QK}[H_{ij} - Y_{ij}(X_{ij}); \rho_i, \rho_j]$. Then it follows that

$$\frac{\partial \mathcal{F}}{\partial \rho_k}(X,\{\rho_i\}) = H_k - Y_k(X_{kk}) + \sum_{j>k} A_{kj}^\star + \sum_{i<k} B_{ik}^\star.$$

(Note that if the dual optimizer is not unique, one only gets a supergradient.) One may take a gradient descent step for ρ_k in the direction of the *traceless part* of $\frac{\partial \mathcal{F}}{\partial \rho_k}$, adjusting the step size if necessary to guarantee that $\rho_k \succeq 0$. Moreover, letting $\rho_{ij}^\star$ be the primal solution of the Kantorovich problem indicated by $\mathbf{QK}[H_{ij} - Y_{ij}(X_{ij}); \rho_i, \rho_j]$, we have

$$\frac{\partial \mathcal{F}}{\partial (\overline{X}_{ii})_{\alpha\beta}}(X, \{\rho_i\}) = -\mathrm{Tr}\big[O_{i,\alpha}^\dagger O_{i,\beta} \rho_i\big],$$
$$\frac{\partial \mathcal{F}}{\partial (\overline{X}_{ij})_{\alpha\beta}}(X, \{\rho_i\}) = -\mathrm{Tr}\big[\big(O_{i,\alpha}^\dagger \otimes O_{j,\beta}\big)\rho_{ij}^\star\big].$$

(If the primal optimizer is not unique, one only gets a subgradient.) After taking a gradient ascent step in X, one may project onto the feasible domain $\{X \succeq 0\}$ by diagonalizing X and zeroing all negative eigenvalues.

Efficient methods for solving the primal and dual quantum Kantorovich problems (beyond black-box SDP solvers) will be explored in future work. In particular, preliminary results indicate promise for a quantum analogue of the classical Sinkhorn scaling algorithm [8], for which the computational cost per iteration is roughly given by the cost of diagonalizing certain operators on $Q_i \otimes Q_j$.

5.4 Full duality

For completeness we also derive the full dual problem to the original two-marginal SDP. We first introduce dual variables $\lambda_i \in \mathbb{R}$ for the constraints $\mathrm{Tr}[\rho_i] = 1$ appearing in the minimization within (5.2), and then exchange the resulting internal supremum over λ with the infimum over $\{\rho_i\}$ to obtain the problem

$$\begin{aligned}
&\sup_{X \succeq 0, \lambda} \inf_{\{\rho_i\}} \Big\{ \sum_i \lambda_i (1 - \mathrm{Tr}[\rho_i]) + \mathcal{F}(X, \{\rho_i\}) \Big\} \\
&\quad = \sup_{X \succeq 0, \lambda} \Big\{ \sum_i \lambda_i + \inf_{\{\rho_i\}} \Big\{ \sum_i \mathrm{Tr}[(H_i - Y_i(X_{ii}) - \lambda_i)\rho_i] \\
&\qquad\qquad + \sum_{i<j} \mathbf{QK}[H_{ij} - Y_{ij}(X_{ij}); \rho_i, \rho_j] \Big\} \Big\}.
\end{aligned}$$

Now by substituting the Kantorovich dual expression (5.1) for $\mathbf{QK}$ and then exchanging maximization and minimization, we obtain the problem

$$\begin{aligned}
\underset{X \succeq 0, \lambda \in \mathbb{R}^M, \{A_{ij}\}, \{B_{ij}\}}{\text{maximize}} \quad & \sum_i \lambda_i + \inf_{\{\rho_i\}} \Big\{ \sum_i \mathrm{Tr}[(H_i - Y_i(X_{ii}) - \lambda_i)\rho_i] \\
& \qquad + \sum_{i<j} \mathrm{Tr}\big[A_{ij}\rho_i\big] + \sum_{i<j} \mathrm{Tr}\big[B_{ij}\rho_j\big] \Big\} \\
\text{subject to} \quad & A_{ij} \otimes \mathrm{Id} + \mathrm{Id} \otimes B_{ij} \preceq H_{ij} - Y_{ij}(X_{ij}), \quad i < j, \\
& X \succeq 0.
\end{aligned}$$

Now the expression within the infimum in the objective function can be rewritten as

$$\sum_i \mathrm{Tr}\Big[\Big(H_i - Y_i(X_{ii}) - \lambda_i + \sum_{j>i} A_{ij} + \sum_{j<i} B_{ji}\Big)\rho_i\Big],$$

so carrying out the infimum within the objective function, we arrive at the full dual:

$$\begin{aligned} \underset{X\succeq 0,\, \lambda\in\mathbb{R}^M, \{A_{ij}\}, \{B_{ij}\}}{\text{maximize}} \quad & \mathbf{1}^\top \lambda \\ \text{subject to} \quad & H_i - Y_i(X_{ii}) - \lambda_i + \sum_{j>i} A_{ij} + \sum_{j<i} B_{ji} = 0, \\ & i = 1, \dots, M, \\ & A_{ij} \otimes \mathrm{Id} + \mathrm{Id} \otimes B_{ij} \preceq H_{ij} - Y_{ij}(X_{ij}), \\ & 1 \le i < j \le M, \quad X \succeq 0, \end{aligned}$$

where the optimization variables $A_{ij} \in \mathrm{End}(Q_i)$ and $B_{ij} \in \mathrm{End}(Q_j)$ are understood to be Hermitian.

Acknowledgment. This work was partially supported by the Department of Energy under grant no. DE-SC0017867 and no. DE-AC02-05CH11231 (L.L.), by the Air Force Office of Scientific Research under award number FA9550-18-1-0095 (L.L. and M.L.), by the National Science Foundation Graduate Research Fellowship Program under grant DGE-1106400, and by the National Science Foundation under award no. 1903031 (M.L.). We thank Garnet Chan, Jianfeng Lu, and Lexing Ying for helpful discussions.

Bibliography

[1] Anderson, J. S.; Nakata, M.; Igarashi, R.; Fujisawa, K.; Yamashita, M. The second-order reduced density matrix method and the two-dimensional hubbard model. *Comput. Theor. Chem.* **1003** (2013), 22–27. doi:10.1016/j.comptc.2012.08.018

[2] Barthel, T.; Hübener, R. Solving condensed-matter ground-state problems by semidefinite relaxations. *Phys. Rev. Lett.* **108** (2012), no. 20, 200404. doi:10.1103/PhysRevLett.108.200404

[3] Biroli, G.; Parcollet, O.; Kotliar, G. Cluster dynamical mean-field theories: Causality and classical limit. *Phys. Rev. B* **69** (2004), no. 20, 205108. doi:10.1103/PhysRevB.69.205108

[4] Bratteli, O.; Robinson, D. W. *Operator algebras and quantum statistical mechanics. 1. C^*- and W^*-algebras, symmetry groups, decomposition of states*. Second edition. Texts and Monographs in Physics. Springer, New York, 1987. doi:10.1007/978-3-662-02520-8

[5] Caglioti, E.; Golse, F.; Paul, T. Toward optimal transport for quantum densities. Preprint, 2018. Available at: https://hal.archives-ouvertes.fr/hal-01963667

[6] Cances, E.; Stoltz, G.; Lewin, M. The electronic ground-state energy problem: A new reduced density matrix approach. *J. Chem. Phys.* **125** (2006), no. 6, 064101. doi:10.1063/1.2222358

[7] Chen, Y.; Gangbo, W.; Georgiou, T.; Tannenbaum, A. On the matrix Monge-Kantorovich problem. *European J. Appl. Math* **31** (2020), no. 4, 574-–600. doi:10.1017/s0956792519000172

[8] Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems* **26** (2013), 2292–2300.

[9] DePrince, A. E.; Mazziotti, D. A. Exploiting the spatial locality of electron correlation within the parametric two-electron reduced-density-matrix method. *J. Chem. Phys.* **132** (2010), no. 3, 034110. doi:10.1063/1.3283052

[10] Ferris, A. J.; Poulin, D. Algorithms for the Markov entropy decomposition. *Phys. Rev. B* **87** (2013), no. 20, 205126. doi:10.1103/PhysRevB.87.205126

[11] Georges, A.; Kotliar, G.; Krauth, W.; Rozenberg, M. J. Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Rev. Modern Phys.* **68** (1996), no. 1, 13–125. doi:10.1103/RevModPhys.68.13

[12] Golse, F.; Mouhot, C.; Paul, T. On the mean field and classical limits of quantum mechanics. *Comm. Math. Phys.* **343** (2016), no. 1, 165–205. doi:10.1007/s00220-015-2485-7

[13] Grant, M.; Boyd, S.: CVX: Matlab software for disciplined convex programming. 2013.

[14] Khoo, Y.; Lin, L.; Lindsey, M.; Ying, L. Semidefinite relaxation of multi-marginal optimal transport for strictly correlated electrons in second quantization. *SIAM J. Sci. Comput.* **42** (2020), no. 6, B1462–B1489. doi:10.1137/20M1310977

[15] Klyachko, A. A. Quantum marginal problem and N-representability. *J. Phys.: Conf. Ser.* **36** (2006), 72. doi:10.1088/1742-6596/36/1/014

[16] Knizia, G.; Chan, G. K.-L. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Phys. Rev. Lett.* **109** (2012), no. 18, 186404. doi:10.1103/PhysRevLett.109.186404

[17] Knizia, G.; Chan, G. K.-L. Density matrix embedding: A strong-coupling quantum embedding theory. *J. Chem. Theory Comput.* **9** (2013), no. 3, 1428–1432. doi:10.1021/ct301044e

[18] Komiya, H. Elementary proof for Sion's minimax theorem. *Kodai Math. J.* **11** (1988), no. 1, 5–7. doi:10.2996/kmj/1138038812

[19] Kotliar, G.; Savrasov, S. Y.; Haule, K.; Oudovenko, V. S.; Parcollet, O.; Marianetti, C. A. Electronic structure calculations with dynamical mean-field theory. *Rev. Mod. Phys.* **78** (2006), no. 3, 865. doi:10.1103/RevModPhys.78.865

[20] Lasserre, J. B. *Moments, positive polynomials and their applications*. Imperial College Press Optimization Series, 1. Imperial College Press, London, 2009.

[21] Leifer, M. S.; Poulin, D. Quantum graphical models and belief propagation. *Ann. Phys.* **323** (2008), no. 8, 1899–1946. doi:10.1016/j.aop.2007.10.001

[22] Li, Y.; Wen, Z.; Yang, C.; Yuan, Y.-x. A semismooth Newton method for semidefinite programs and its applications in electronic structure calculations. *SIAM J. Sci. Comput.* **40** (2018), no. 6, A4131–A4157. doi:10.1137/18M1188069

[23] Mardirossian, N.; McClain, J. D.; Chan, G. Lowering of the complexity of quantum chemistry methods by choice of representation. *J. Chem. Phys.* **148** (2018), no. 4, 044106. doi:10.1063/1.5007779

[24] Mazziotti, D. A. Contracted Schrödinger equation: Determining quantum energies and two-particle density matrices without wave functions. *Phys. Rev. A* **57** (1998), no. 6, 4219. doi:10.1103/PhysRevA.57.4219

[25] Mazziotti, D. Realization of quantum chemistry without wave functions through first-order semidefinite programming. *Phys. Rev. Lett.* **93** (2004), no. 21, 213001. doi:10.1103/PhysRevLett.93.213001

[26] Mazziotti, D. Structure of fermionic density matrices: Complete N-representability conditions. *Phys. Rev. Lett.* **108** (2012), no. 26, 263002. doi:10.1103/PhysRevLett.108.263002

[27] Nakata, M.; Nakatsuji, H.; Ehara, M.; Fukuda, M.; Nakata, K.; Fujisawa, K. Variational calculations of fermion second-order reduced density matrices by semidefinite programming algorithm. *J. Chem. Phys.* **114** (2001), no. 19, 8282–8292. doi:10.1063/1.1360199

[28] Negele, J. W.; Orland, H. *Quantum many-particle systems*. Frontiers in Physics, 68. Addison-Wesley, Advanced Book Program, Redwood City, CA, 1988. doi:10.1142/S0218001488000273

[29] Orús, R. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Ann. Phys.* **349** (2014), 117–158. doi:10.1016/j.aop.2014.06.013

[30] Oseledets, I. V.; Tyrtyshnikov, E. E. Breaking the curse of dimensionality, or how to use svd in many dimensions. *SIAM J. Sci. Comput.* **31** (2009), no. 5, 3744–3759. doi:10.1137/090748330

[31] Pearl, J. Reverend Bayes on inference engines: a distributed hierarchical approach. *Proceedings of the Second National Conference on Artificial Intelligence* (1982), 133–136.

[32] Poulin, D.; Hastings, M. B. Markov entropy decomposition: A variational dual for quantum belief propagation. *Phys. Rev. Lett.* **106** (2011), no. 8, 080403. doi:10.1103/PhysRevLett.106.080403

[33] Raghu, S.; Kivelson, S.; Scalapino, D. Superconductivity in the repulsive hubbard model: An asymptotically exact weak-coupling solution. *Phy. Rev. B* **81** (2010), no. 22, 224505. doi:10.1103/PhysRevB.81.224505

[34] Rouzé, C.; Datta, N. Concentration of quantum states from quantum functional and transportation cost inequalities. *J. Math. Phys.* **60** (2019), no. 1, 012202. doi:10.1063/1.5023210

[35] Schilling, C. The quantum marginal problem. *Mathematical results in quantum mechanics*, 165–176. World Scientific, Hackensack, NJ, 2015.

[36] Seidl, M.; Gori-Giorgi, P.; Savin, A. Strictly correlated electrons in density-functional theory: A general formulation with applications to spherical densities. *Phys. Rev. A* **75** (2007), no. 4, 042511. doi:10.1103/PhysRevA.75.042511

[37] Seidl, M.; Perdew, J. P.; Levy, M. Strictly correlated electrons in density-functional theory. *Phys. Rev. A* **59** (1999), no. 1, 51. doi:10.1103/PhysRevA.59.51

[38] Sun, Q.; Chan, G. K.-L. Quantum embedding theories. *Acc. Chem. Res.* **49** (2016), no. 12, 2705–2712. doi:10.1021/acs.accounts.6b00356

[39] Szabo, A.; Ostlund, N. *Modern quantum chemistry: Introduction to advanced electronic structure theory*. McGraw-Hill, New York, 1989.

[40] Verstraete, F.; Cirac, J. I. Renormalization algorithms for quantum-many body systems in two and higher dimensions. Preprint, 2004. arXiv:cond-mat/0407066 [cond-mat.str-el]

[41] Villani, C. *Optimal transport: Old and new*. Springer, Berlin-Heidelberg, 2009.

[42] Wainwright, M. J.; Jordan, M. I. *Graphical models, exponential families, and variational inference*. Foundations and Trends in Machine Learning, 1-2. Now, Boston-Delft, 2008.

[43] White, S. R. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.* **69** (1992), no. 19, 2863. doi:10.1103/PhysRevLett.69.2863

[44] White, S. R.; Stoudenmire, E. M. Multisliced gausslet basis sets for electronic structure. *Phys. Rev. B* **99** (2019), no. 8, 081110. doi:10.1103/PhysRevB.99.081110

[45] Ye, H.-Z.; Ricke, N. D.; Tran, H. K.; Van Voorhis, T. Bootstrap embedding for molecules. *J. Chem. Theory Comput.* **15** (2019), no. 8, 4497–4506. doi:10.1021/acs.jctc.9b00529

[46] Zhao, Z.; Braams, B. J.; Fukuda, M.; Overton, M. L.; Percus, J. K. The reduced density matrix method for electronic structure calculations and the role of three-index representability conditions. *J. Chem. Phys.* **120** (2004), no. 5, 2095–2104. doi:10.1063/1.1636721

[47] Zhou, L.; Ying, S.; Yu, N.; Ying, M. Strassen's theorem for quantum couplings. *Theor. Comput. Sci.* **802** (2020), 67-–76. doi:10.1016/j.tcs.2019.08.026

LIN LIN
Department of Mathematics
University of California, Berkeley
1083 Evans Hall
Berkeley, CA 94720
USA
E-mail: linlin@math.berkeley.edu

MICHAEL LINDSEY
Courant Institute
251 Mercer St.
New York, NY 10012
USA
E-mail: michael.lindsey@cims.nyu.edu

Received January 2020.