Semidefinite relaxation of multi-marginal optimal transport, with application to strictly correlated electrons in second quantization

Michael Lindsey, UC Berkeley
(Joint work with Yuehaw Khoo, Lin Lin, Lexing Ying)

Molecular simulation: quantum mechanical models
ICIAM, July 2019

Background

- Electronic structure theory concerns the N-particle Schrödinger operator

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\Delta_{x_i} + \sum_{i=1}^{N} V_{\mathrm{ext}}(x_i) + \lambda \sum_{i<j} V_{\mathrm{c}}(x_i - x_j),$$

  acting on the space $\mathcal{H} = \Lambda^N(L^2(\mathbb{R}^d;\mathbb{C}))$ of antisymmetric functions $\psi(x_1,\ldots,x_N) \in \mathbb{C}$
- One is often interested in determining the ground-state energy E_0; in particular, this allows one to compute molecular dynamics in the Born-Oppenheimer approximation
- Electron density $\rho : \mathbb{R}^d \to \mathbb{R}$ defined via

$$\rho(x) = N \int |\psi(x, x_2, \ldots, x_N)|^2 \, dx_2 \cdots dx_N$$

Background

- Density functional theory [see Kohn and Sham (1965), Levy (1979), Lieb (1983)]: there exists a universal (i.e., independent of V_{ext}) functional $F[\rho]$ such that

$$E_0 = \inf_{\rho} \left\{ \int V_{\text{ext}}(x)\rho(x)\,dx + F[\rho] \right\}$$

- In general $F[\rho]$ is unknown. (Note $\lambda = 0$ case is trivial)
- Meanwhile $\lambda \to +\infty$ limit is the regime of strictly correlated electrons (SCE), cf. Seidl et al (1999)
- In this limit, exact functional can be expressed in terms of a multi-marginal optimal transport (MMOT) problem with N marginals, marginal state space $\mathbb{R}^d$, and pairwise cost function $\sum_{i<j} V_{\text{c}}(x_i - x_j)$

Background

- One hopes that SCE solution could be productively baked into practical DFT functionals, but MMOT is hard
- See preprint for references; especially relevant motivation is Khoo and Ying (2018)
- Real-space grid seems fundamental to discretization of SCE (?), preventing one from making use of small but effective quantum chemistry basis sets
- Hence we try to come up with a notion of SCE *directly* in second quantization, after a choice of basis has already been made
- We can also consider model problems (e.g., Hubbard-like models) not directly derived via a choice of basis for a first-quantized problem
- Will be similar but different: still MMOT, but now L marginals (where L is number of sites or basis elements in model) each of state space $\{0, 1\}$

Outline

- Extend the formalism of 'strictly correlated electrons' to the setting of second quantization
  - Exact expression for density functional in the limit of infinitely strong electronic repulsion
  - Expression involves a multi-marginal optimal transport (MMOT) problem with pairwise cost
- Introduce convex relaxation method for approximately solving **general** MMOT problems with pairwise cost
- Discuss dual structure and interpretation
- Numerical experiments on model problems

Second quantization

- State space is called the Fock space, denoted by $\mathcal{F}$. The occupation number (orthonormal) basis set for the Fock space is

$$\{|s_1, \ldots, s_L\rangle\}_{s_i \in \{0,1\}, i=1,\ldots,L}$$

- State $|\psi\rangle \in \mathcal{F}$ will be written as a linear combination of occupation number basis elements as follows:

$$|\psi\rangle = \sum_{s_1,\ldots,s_L \in \{0,1\}} \psi(s_1, \ldots, s_L) \, |s_1, \ldots, s_L\rangle \,, \quad \psi(s_1, \ldots, s_L) \in \mathbb{C}$$

- Hence the state vector $|\psi\rangle$ can be identified with a vector $\psi \in \bigotimes^L \mathbb{C}^2 \simeq \mathbb{C}^{2^L}$
- The fermionic creation operators are defined (via this perspective) as

$$a_p^\dagger = \underbrace{\sigma^z \otimes \cdots \otimes \sigma^z}_{p-1 \text{ factors}} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \otimes I_2 \otimes I_2 \otimes \cdots \,,$$

and the annihilation operators a_p are the Hermitian adjoints

Second quantization

- More important (and fundamental) than this definition are the CAR:

$$\{a_p, a_q^\dagger\} = \delta_{pq}, \quad \{a_p^\dagger, a_q^\dagger\} = \{a_p, a_q\} = 0$$

- Number operator defined as $\hat{n}_p := \hat{a}_p^\dagger \hat{a}_p$, total number operator defined $\hat{N} = \sum_p \hat{n}_p$
- The Hamiltonian operator is assumed to take the following form:

$$\hat{H} = \sum_{p,q=1}^{L} t_{pq} \hat{a}_p^\dagger \hat{a}_q + \sum_{p=1}^{L} w_p \hat{n}_p + \sum_{p,q=1}^{L} v_{pq} \hat{n}_p \hat{n}_q$$

- Note the restriction of the form of the two-body interaction
  - We do not consider the general form $\sum_{p,q,r,s=1}^{L} v_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_s \hat{a}_r$ that results from arbitrary quantum chemistry basis set
  - Special choices of basis (e.g., Gausslets [White (2017)]) can achieve the restricted form
  - Also covers model problems such as Hubbard model with long-range Coulomb interaction
- Following DFT, one could think of t, v as fixed, w as problem-specific

Site occupation functional theory (SOFT) / lattice DFT

- Analog of Levy-Lieb in second quantization: SOFT / lattice DFT [Schönhammer 1995]
- N-particle ground state energy given by

$$E_0 = \inf_{|\psi\rangle \in \mathcal{F} \,:\, \langle\psi|\psi\rangle = 1,\, \langle\psi|\hat{N}|\psi\rangle = N} \langle\psi|\, \hat{H}\, |\psi\rangle$$

- Electron density $\rho \in \mathbb{R}^L$ is defined as $\rho_p = \langle\psi|\hat{n}_p|\psi\rangle$, which satisfies $\sum_p \rho_p = N$

Site occupation functional theory (SOFT) / lattice DFT

Then we follow the Levy-Lieb constrained minimization approach [Levy (1979),Lieb (1983)] and rewrite the ground state minimization problem as follows:

$$E_0 = \inf_{\rho\in\mathcal{J}_N}\left\{\sum_p \rho_p w_p + \left(\inf_{|\psi\rangle\mapsto\rho,|\psi\rangle\in\mathcal{F}} \langle\psi|\sum_{pq} t_{pq}\hat{a}_p^\dagger\hat{a}_q + \sum_{pq} v_{pq}\hat{n}_p\hat{n}_q\,|\psi\rangle\right)\right\}$$
$$= \inf_{\rho\in\mathcal{J}_N}\{W[\rho] + F_{\mathsf{LL}}[\rho]\},$$

where

$$F_{\mathsf{LL}}[\rho] := \inf_{|\psi\rangle\mapsto\rho,|\psi\rangle\in\mathcal{F}} \langle\psi|\sum_{pq} t_{pq}\hat{a}_p^\dagger\hat{a}_q + \sum_{pq} v_{pq}\hat{n}_p\hat{n}_q\,|\psi\rangle\,.$$

Here

$$\mathcal{J}_N := \left\{\rho\in\mathbb{R}^L \;\middle|\; \rho\geq 0, \sum_p \rho_p = N\right\}.$$

SCE

- As in first-quantized SCE, we can lower-bound as

$$\begin{aligned} F_{\mathsf{LL}}[\rho] &\geq \inf_{|\psi\rangle\mapsto\rho} \langle\psi| \sum_{pq} t_{pq}\hat{a}_p^\dagger \hat{a}_q |\psi\rangle + \inf_{|\psi\rangle\mapsto\rho} \langle\psi| \sum_{pq} v_{pq}\hat{n}_p\hat{n}_q |\psi\rangle \\ &=: T[\rho] + E_{\mathsf{sce}}[\rho], \end{aligned}$$

- Then we can lower-bound the ground-state energy in terms of the 'Kohn-Sham SCE' energy:

$$E_0 \geq E_{\text{KS-SCE}} := \inf_{\rho\in\mathcal{J}_N} \{W[\rho] + T[\rho] + E_{\mathsf{sce}}[\rho]\} ..$$

- Can derive self-consistency condition via stationarity (or, more rigorously, convex duality): at optimal ρ, the effective single-particle Hamiltonian

$$\hat{H}_0[\rho] := \sum_{p,q=1}^{L} t_{pq}\hat{a}_p^\dagger\hat{a}_q + \sum_{p=1}^{L} \left[w_p + (v_{\text{sce}}[\rho])_p\right] \hat{n}_p$$

has ground state with density ρ, where $v_{\text{sce}}[\rho] := \nabla_\rho E_{\text{sce}}[\rho]$

SCE $\to$ MMOT

- Hence up to convergence of self-consistent loop, need only worry about computing $E_{\mathrm{sce}}[\rho]$ and its gradient
- To this end, rewrite

$$\begin{aligned} E_{\mathsf{sce}}[\rho] &= \inf_{|\psi\rangle \mapsto \rho} \langle \psi | \sum_{pq} v_{pq} \hat{n}_p \hat{n}_q \, |\psi\rangle \\ &= \inf_{|\psi\rangle \mapsto \rho} \sum_{s_1,\ldots,s_L} \sum_{pq} v_{pq} s_p s_q |\psi(s_1,\ldots,s_L)|^2 \\ &= \inf_{\mu \in \Pi(\rho)} \sum_{s_1,\ldots,s_L} \sum_{pq} v_{pq} s_p s_q \mu(s_1,\ldots,s_L), \end{aligned}$$

  where $\Pi(\rho)$ is the space of joint probability mass functions on $\{0,1\}^L$ with marginals satisfying

$$\mu_p^{(1)}(s) = (1-\rho_p)\delta_{s0} + \rho_p \delta_{s1}, \quad s = 0, 1.$$

- Considering the $\mu_p^{(1)}$ alternately as vectors, we also write

$$\mu_p^{(1)} = [1-\rho_p, \rho_p]^\top$$

SCE $\rightarrow$ MMOT

- Define the cost function $C : \{0,1\}^L \rightarrow \mathbb{R}$ by

$$C(s_1, \dots, s_L) := \sum_{pq} v_{pq} s_p s_q.$$

Then our SCE energy may be written

$$E_{\text{sce}}[\rho] = \inf_{\mu \in \Pi(\rho)} \sum_{s_1,\dots,s_L} C(s_1, \dots, s_L)\, \mu(s_1, \dots, s_L) = \inf_{\mu \in \Pi(\rho)} \langle C, \mu \rangle,$$

- This is precisely the form of a MMOT problem
- In other words, minimization of a linear functional of a joint probability measure subject to constraints on all of the marginals of the measure [see, e.g., Pass (2015)]
- Note that dimension of the feasible space for this problem is exponential in L
- Gradient $v_{\text{sce}}[\rho]$ defined in terms of the Kantorovich potentials (solutions to dual LP)

Pairwise cost and representability

- Efficient relaxation enabled by pairwise cost structure

$$C(s_1, \ldots, s_L) = \sum_{p \neq q} v_{pq} s_p s_q =: \sum_{p \neq q} C_{pq}(s_p, s_q).$$

- Accordingly, the objective function can be written as

$$E_{\mathsf{sce}}[\rho] = \inf_{\mu \in \Pi(\rho)} \sum_{p \neq q} \langle C_{pq}, \mu^{(2)}_{pq} \rangle,$$

  where the 2-marginals $\mu^{(2)}_{pq}$ are defined implicitly in terms of μ
- We also identify $\mu^{(2)}_{pq}$ and C_{pq} with 2×2 matrices, e.g.,

$$\mu^{(2)}_{pq} = \begin{bmatrix} \mu^{(2)}_{pq}(0,0) & \mu^{(2)}_{pq}(0,1) \\ \mu^{(2)}_{pq}(1,0) & \mu^{(2)}_{pq}(1,1) \end{bmatrix},$$

  so

$$E_{\mathsf{sce}}[\rho] = \inf_{\mu \in \Pi(\rho)} \sum_{p \neq q} \mathrm{Tr}[C_{pq} \mu^{(2)}_{pq}]$$

- The 2-marginals must be jointly *representable*; exponentially hard to enforce exactly

Relaxation of representability condition

- In the following we adopt the notation

$$\mathbf{s} = (s_1, \ldots, s_L) \in \{0,1\}^L$$

- For such $\mathbf{s}$, define $e_{\mathbf{s}} \in \mathbb{R}^{2\times 2\times \cdots \times 2}$, via

$$e_{\mathbf{s}} = e_{s_1} \otimes \cdots \otimes e_{s_L},$$

  where we adopt the (zero-indexing) convention $e_0 = [1,0]^\top$, $e_1 = [0,1]^\top$

- Any probability measure μ on $\{0,1\}^L$ can be written as a convex combination of the $e_{\mathbf{s}}$, i.e.,

$$\mu = \sum_{\mathbf{s}} a_{\mathbf{s}} e_{\mathbf{s}}, \quad \text{where} \quad \sum_{\mathbf{s}} a_{\mathbf{s}} = 1,\ a_{\mathbf{s}} \geq 0$$

Relaxation of representability condition

- It follows that the 1- and 2-marginals can be written

$$\mu_p^{(1)} = \sum_{\mathbf{s}} a_{\mathbf{s}}\, e_{s_p}, \quad \mu_{pq}^{(2)} = \sum_{\mathbf{s}} a_{\mathbf{s}}\, e_{s_p} \otimes e_{s_q} = \sum_{\mathbf{s}} a_{\mathbf{s}}\, e_{s_p} e_{s_q}^{\top}$$

- Now define

$$M = M(\{a_{\mathbf{s}}\}) = \sum_{\mathbf{s}} a_{\mathbf{s}} \begin{bmatrix} e_{\mathbf{s}_1} \\ \vdots \\ e_{\mathbf{s}_L} \end{bmatrix} \begin{bmatrix} e_{\mathbf{s}_1}^{\top} \cdots e_{\mathbf{s}_L}^{\top} \end{bmatrix},$$

- Then M is the matrix of 2×2 blocks M_{pq} given by

$$M_{pq} = \begin{cases} \operatorname{diag}(\mu_p^{(1)}), & p = q, \\ \mu_{pq}^{(2)}, & p \neq q \end{cases}$$

- Then let $C = (C_{pq}) \in \mathbb{R}^{(2L)\times(2L)}$ be the matrix of the 2×2 blocks C_{pq} defined above, which specifies the pairwise cost on each pair of marginals

Relaxation of representability condition

- Then objective can in fact be rewritten as

$$\sum_{p \neq q} \mathrm{Tr}[C_{pq} \mu_{pq}^{(2)}] = \mathrm{Tr}[CM]$$

- Hence the MMOT problem is equivalently

$$\underset{M \in \mathbb{R}^{(2L)\times(2L)},\, \{a_{\mathbf{s}}\}_{\mathbf{s}\in\{0,1\}^L}}{\text{minimize}} \quad \mathrm{Tr}(CM)$$

$$\text{subject to} \quad M = \sum_{\mathbf{s}} a_{\mathbf{s}} \begin{bmatrix} e_{\mathbf{s}_1} \\ \vdots \\ e_{\mathbf{s}_L} \end{bmatrix} \left[e_{\mathbf{s}_1}^\top \cdots e_{\mathbf{s}_L}^\top \right],$$

$$M_{pp} = \mathrm{diag}(\mu_p^{(1)}) \text{ for all } p = 1, \ldots, L,$$

$$\sum_{\mathbf{s}} a_{\mathbf{s}} = 1, \quad a_{\mathbf{s}} \geq 0 \text{ for all } \mathbf{s} \in \{0,1\}^L.$$

Relaxation of representability condition

- At this point, we have not alleviated its exponential complexity; indeed, note that $\{a_{\mathbf{s}}\}_{\mathbf{s}\in\{0,1\}^L}$ is a vector of size 2^L
- We relax by omitting $\{a_{\mathbf{s}}\}_{\mathbf{s}\in\{0,1\}^L}$ entirely from the optimization, retaining only M as an optimization variable and enforcing several necessary constraints on M that are satisfied by the solution of the exact problem
- First, note that $M \geq 0$ and $M \succeq 0$
- Second, the fact that the 1-marginals can be written in terms of the 2-marginals imposes additional *local consistency* constraints on M

Relaxation of representability condition

- Then we obtain the relaxation (the 'primal 2-marginal SDP'):

$$
\begin{aligned}
\underset{M \in \mathbb{R}^{(2L)\times(2L)}}{\text{minimize}} \quad & \operatorname{Tr}(CM) \\
\text{subject to} \quad & M \succeq 0, \\
& M_{pq} \geq 0 \ \text{ for all } \ p,q = 1,\dots,L \ (p<q), \\
& M_{pq}\mathbf{1}_2 = \mu_p^{(1)} \ \text{ for all } \ p,q = 1,\dots,L \ (p<q), \\
& M_{pq}^{\top}\mathbf{1}_2 = \mu_q^{(1)} \ \text{ for all } \ p,q = 1,\dots,L \ (p<q), \\
& M_{pp} = \operatorname{diag}(\mu_p^{(1)}) \ \text{ for all } \ p = 1,\dots,L
\end{aligned}
$$

- Note that the optimal value of the primal problem is in fact attained because the constraints define a compact feasible set
- The fact that the marginal state space is $\{0,1\}$ is immaterial to the derivation; i.e., works for arbitrary MMOT problem with pairwise cost

Duality

- Results will only be sketched
- Writing dual in the right way reveals interesting structure; can be interpreted as solving an ordinary OT problem for each pair of marginals; these OT problems are completely decoupled, modulo coupling via a globally determined effective contribution to their cost functions
- Suggests fast algorithms for future work, as well as hybridization with existing methods for ordinary OT
- Careful understanding of dual also allows one to prove not only strong duality, but also that dual optimizer is attained (nontrivial because Slater's condition fails)

Model problem

Here we consider a 1D spinless Hubbard-like model defined by the Hamiltonian of Eq. (7), in which we take

$$t_{pq} = \begin{cases} 1 & \text{if } |q-p| = 1, \\ 0 & \text{otherwise} \end{cases}$$

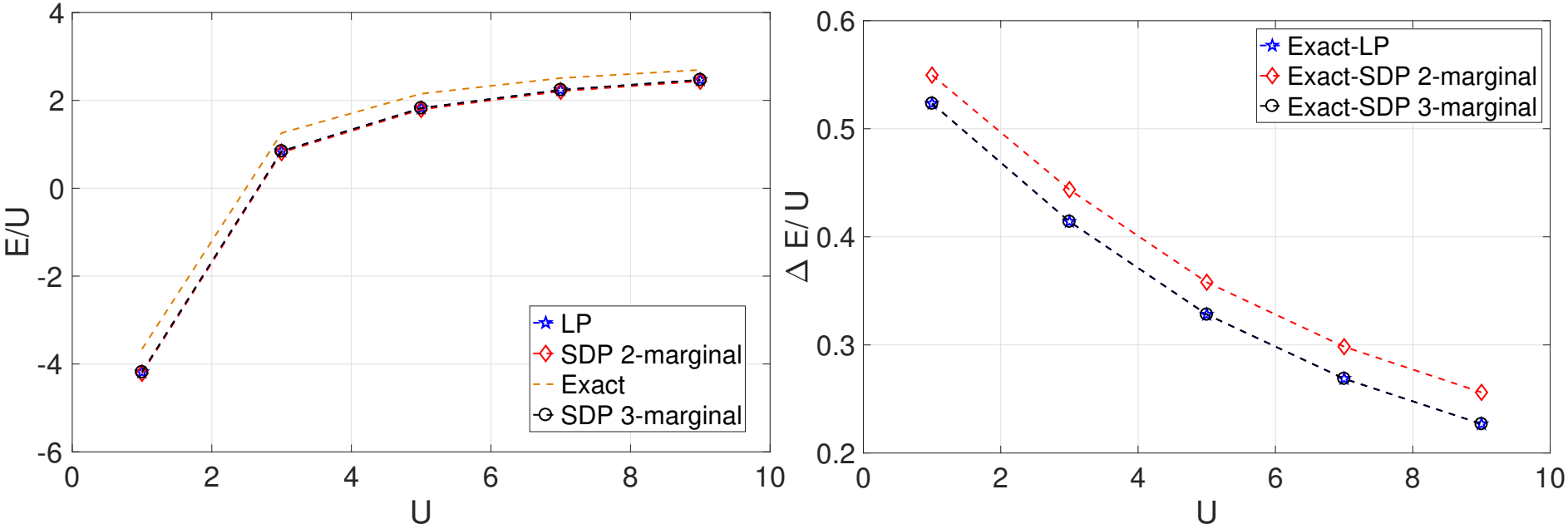
and consider two different cases of v, with next-nearest neighbor (NNN) interaction,

$$v_{pq} = \begin{cases} U/2 & \text{if } |q-p| = 1, \\ U/40 & \text{if } |q-p| = 2, \\ 0 & \text{otherwise} \end{cases}$$

and next-next-nearest neighbor interaction (NNNN)

$$v_{pq} = \begin{cases} U/2 & \text{if } |q-p| = 1, \\ U/20 & \text{if } |q-p| = 2, \\ U/200 & \text{if } |q-p| = 3, \\ 0 & \text{otherwise} \end{cases}$$

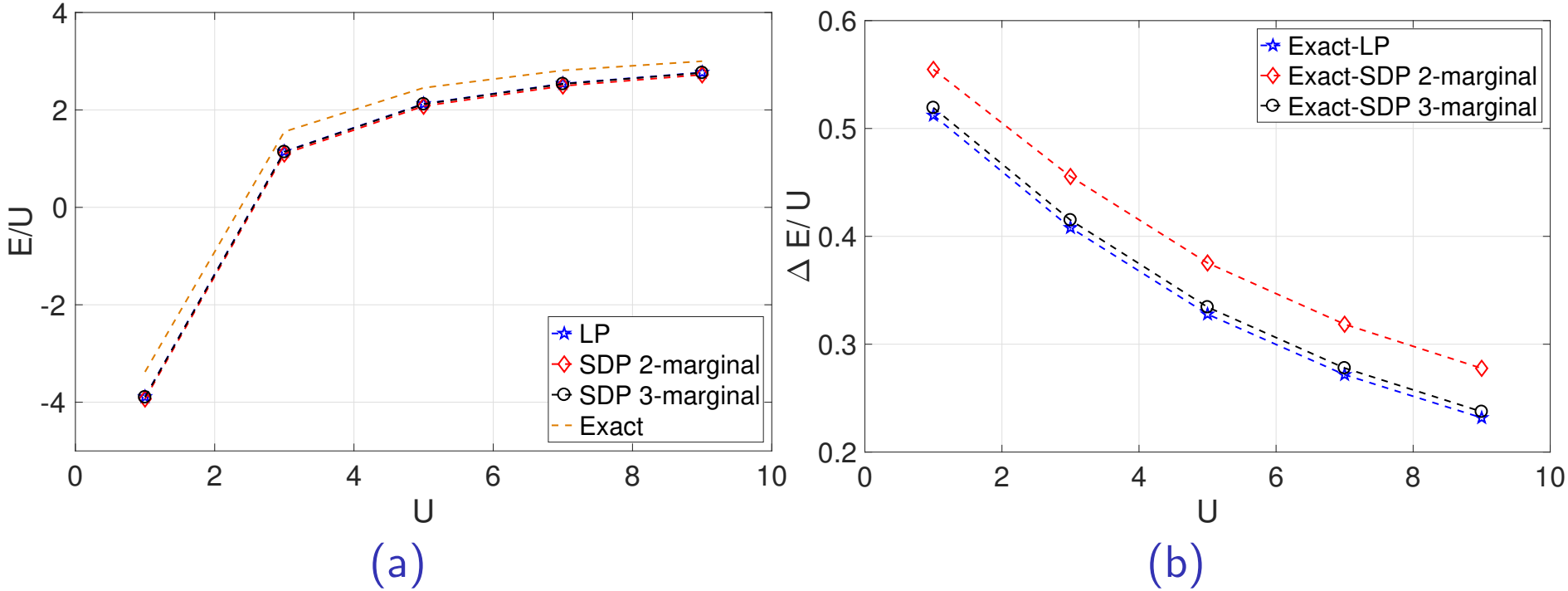
NNN interaction



(a) (b)

Figure: Spinless 1D fermionic lattice NNN model, $L = 14$, $N = 9$. (a) E/U as a function of U. (b) Difference between the exact energy and the Kohn-Sham SCE energies obtained from the unrelaxed LP and the SDP relaxations.

NNNN interaction



(a) (b)

Figure: Spinless 1D fermionic lattice NNNN model, $L = 14$, $N = 9$. (a) E/U as a function of U. (b) Difference between the exact energy and the Kohn-Sham SCE energies obtained from the unrelaxed LP and the SDP relaxations.

SCE potential

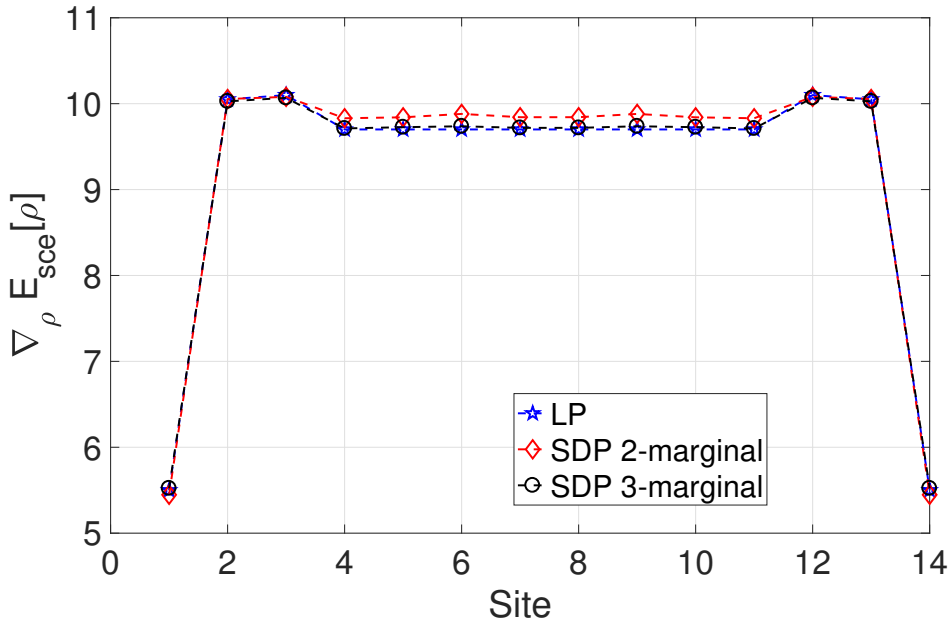
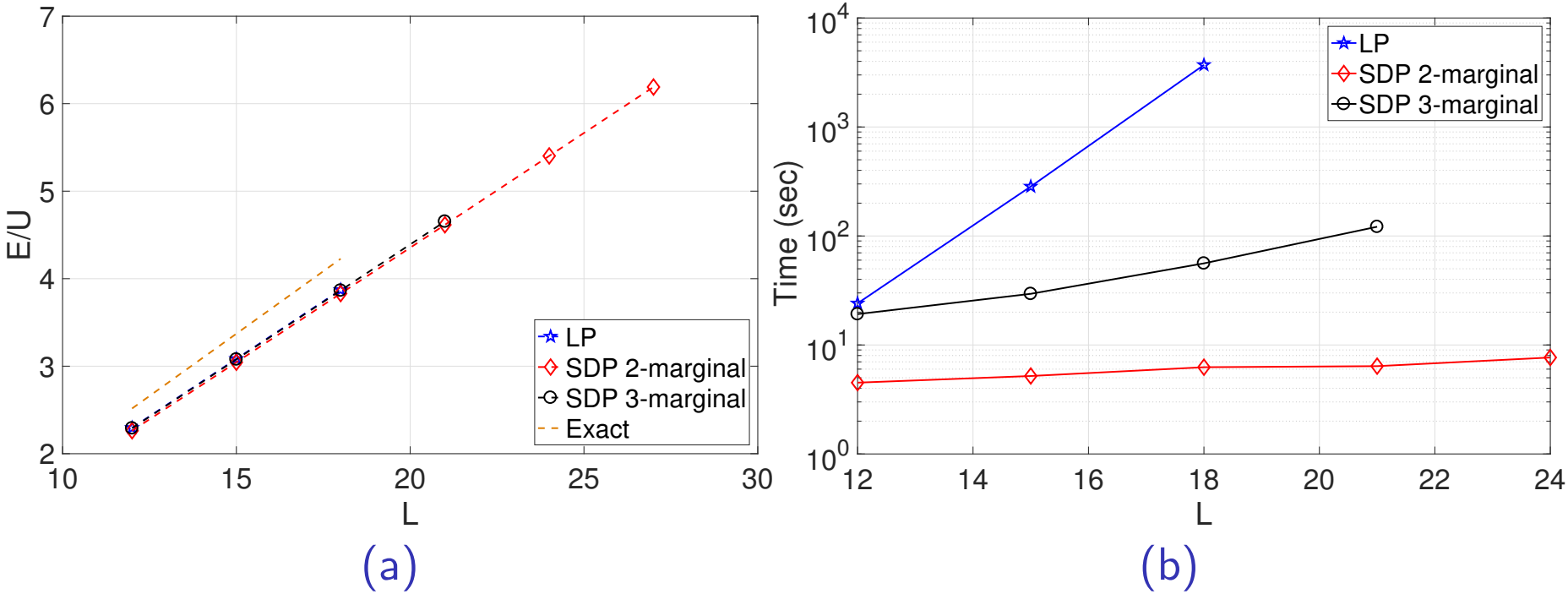


Figure: The effective potential for the spinless 1D fermionic lattice NNN model, $U = 5$, $L = 14$, $N = 9$. The relative ℓ^2 errors for the 2- and 3-marginal formulations (compared to the unrelaxed LP formulation) are 1.2×10^{-2} and 2.7×10^{-3}, respectively.

Energy and runtime scaling



(a) (b)

Figure: Spinless 1D fermionic lattice NNN model, $U = 5$, $N/L = 2/3$. (a) E/U as a function of L. (b) Running time as a function of L.

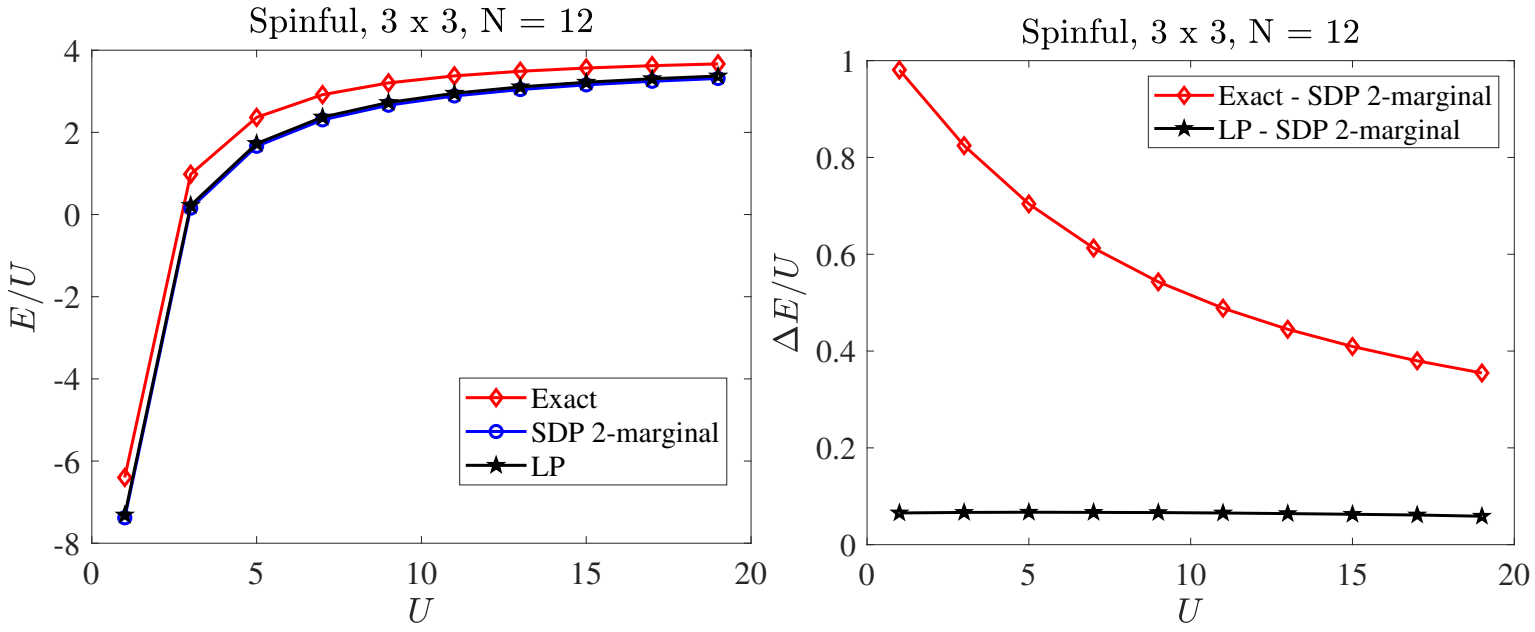
Generalized Hubbard model

We consider a 2D generalized Hubbard type model defined by the Hamiltonian

$$\hat{H} = -\sum_{i,j=1}^{L-1} \sum_{\sigma\in\{\uparrow,\downarrow\}} \left(\hat{a}^{\dagger}_{i+1,j;\sigma} \hat{a}_{i,j;\sigma} + \hat{a}^{\dagger}_{i,j+1;\sigma} \hat{a}_{i,j;\sigma} + \text{h.c.} \right)$$
$$+ U \sum_{i,j=1}^{L} \hat{n}_{i,j;\uparrow} \hat{n}_{i,j;\downarrow} + V \sum_{i,j=1}^{L-1} \left(\hat{n}_{i+1,j} \hat{n}_{i,j} + \hat{n}_{i,j+1} \hat{n}_{i,j} \right).$$

Here $\hat{n}_{i,j} := \hat{n}_{i,j;\uparrow} + \hat{n}_{i,j;\downarrow}$.

Generalized Hubbard model



(a) (b)

Figure: Spinful 3×3 Hubbard model with $N = 12$.

Extensions and conclusions

- As mentioned, possibility for fast algorithm for 2-marginal SDP
- Can get tighter relaxation by treating 3-marginals (and higher) as optimization variables
- Reference for this talk is arXiv:1905.08322
- Acknowledgments: my collaborators Yuehaw Khoo, Lin Lin, and Lexing Ying; NSF GRFP (DGE-1106400)