

Gauge constrained algorithm of variational discrete action theory at $\mathcal{N} = 3$ for the multiorbital Hubbard model

Zhengqian Cheng  and Chris A. Marianetti *Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York 10027, USA*

(Received 28 April 2023; revised 27 June 2023; accepted 28 June 2023; published 13 July 2023)

The recently developed variational discrete action theory (VDAT) provides a systematic variational approach to the ground state of the quantum many-body problem, where the quality of the solution is controlled by an integer \mathcal{N} , and increasing \mathcal{N} monotonically approaches the exact solution. VDAT can be exactly evaluated in the $d = \infty$ multiorbital Hubbard model using the self-consistent canonical discrete action theory (SCDA), which requires a self-consistency condition for the integer time Green's functions. Previous work demonstrates that $\mathcal{N} = 3$ accurately captures multiorbital Mott/Hund physics at a cost similar to the Gutzwiller approximation. Here we employ a gauge constraint to automatically satisfy the self-consistency condition of the SCDA at $\mathcal{N} = 3$, yielding an even more efficient algorithm with enhanced numerical stability. We derive closed form expressions of the gauge constrained algorithm for the multiorbital Hubbard model with general density-density interactions, allowing VDAT at $\mathcal{N} = 3$ to be straightforwardly applied to the seven-orbital Hubbard model. We present results and a performance analysis using $\mathcal{N} = 2$ and $\mathcal{N} = 3$ for the $SU(2N_{\text{orb}})$ Hubbard model in $d = \infty$ with $N_{\text{orb}} = 2-8$, and compare to numerically exact dynamical mean-field theory solutions where available. The developments in this work will greatly facilitate the application of VDAT at $\mathcal{N} = 3$ to strongly correlated electron materials.

DOI: [10.1103/PhysRevB.108.035127](https://doi.org/10.1103/PhysRevB.108.035127)

I. INTRODUCTION

The recently developed variational discrete action theory (VDAT) [1,2] has emerged as a powerful tool to study the ground state of the multiorbital Hubbard model [3], which can be considered as a minimal model for a wide class of strongly correlated electron materials [4,5]. VDAT consists of two central components: the sequential product density matrix (SPD) ansatz and the discrete action theory to evaluate observables under the SPD. The accuracy of the SPD is controlled by an integer \mathcal{N} , and the SPD monotonically approaches the exact solution for increasing \mathcal{N} . In the context of the Hubbard model, the SPD recovers most well-known variational wave functions [1]: $\mathcal{N} = 1$ recovers the Hartree-Fock wave function, $\mathcal{N} = 2$ recovers the Gutzwiller wave function [6–8], and $\mathcal{N} = 3$ recovers the Gutzwiller-Baeriswyl [9] and Baeriswyl-Gutzwiller wave functions [10]. The discrete action theory can be viewed as an integer time generalization of the imaginary time path integral, yielding an integer time generalization of the Green's function and Dyson equation [1]. For $d = \infty$, the SPD can be exactly evaluated using the self-consistent canonical discrete action (SCDA) [1,2]. VDAT within the SCDA offers a paradigm shift away from the dynamical mean-field theory (DMFT) [5,11,12], allowing the exact solution of the ground-state properties of the $d = \infty$ Hubbard model to be systematically approached within the wave function paradigm. The computational cost of VDAT grows with \mathcal{N} , at an exponential scaling for an exact evaluation and a polynomial scaling for a numerical evaluation using Monte Carlo, so rapid convergence with \mathcal{N} is important if VDAT is to be a practical alternative to DMFT. VDAT using $\mathcal{N} = 2, 3, 4$ has

been applied to the single-orbital Anderson impurity model on a ring [2], the $d = \infty$ single-orbital Hubbard model [2], and the $d = \infty$ two-orbital Hubbard model [3], and in all cases $\mathcal{N} = 3$ yields accurate results as compared to the numerically exact solutions. This success is particularly nontrivial in the two-orbital problem, where complex local interactions including the Hubbard U , Hund J , and crystal field Δ were studied over all parameter space. Therefore, VDAT within the SCDA at $\mathcal{N} = 3$ provides a minimal and accurate description of the two-orbital Hubbard model, but with a computational cost that is comparable to $\mathcal{N} = 2$, which recovers the Gutzwiller approximation [8] (GA) and slave boson mean-field theories [13–15]. The fact that VDAT within the SCDA at $\mathcal{N} = 3$ resolves all the limitations of the Gutzwiller approximation and the slave boson mean-field theories without substantially increasing the computational cost motivates a deeper understanding of how the SCDA works.

The SCDA provides a route for exactly evaluating the SPD in $d = \infty$, and the SCDA can be viewed as the integer time analog of DMFT [1,3]. While DMFT maps the Hubbard model to a self-consistently determined Anderson impurity model, the SCDA maps the SPD to a self-consistently determined canonical discrete action (CDA), parametrized by the corresponding noninteracting integer time Green's function \mathcal{G} , which implicitly depends on the variational parameters of the SPD. While the DMFT self-consistency condition only needs to be executed once, the SCDA self-consistency condition must be executed for every choice of variational parameters during the minimization. Previously, we proposed an approach to mitigate this issue by simultaneously minimizing the variational parameters and updating \mathcal{G} , and demonstrated

it to be efficient for the two-band Hubbard model [3]. However, for some regions of parameter space, such as in the large polarization regime, a very small step size is needed to maintain numerical stability. Such problems are partially due to inaccuracies of \mathcal{G} within the iteration process, given that \mathcal{G} is only highly precise when the fixed point is reached. Therefore, it would be advantageous if the self-consistency condition of the SCDA could be automatically satisfied.

Given that $\mathcal{N} = 2$ recovers the GA, it is interesting to recall how the GA automatically satisfies the SCDA self-consistency condition. As previously demonstrated [1], the GA has a prescribed form of \mathcal{G} , which is fully determined by imposing that the noninteracting and interacting local single-particle density matrices are identical, which we refer to as the Gutzwiller gauge. While the SCDA at $\mathcal{N} = 2$ can evaluate an SPD with arbitrary variational parameters, the GA is only valid when the SPD satisfies the Gutzwiller gauge. Therefore, the GA is a special case of the SCDA at $\mathcal{N} = 2$, but the Gutzwiller gauge does not limit the variational power of the SPD due to the gauge freedom of the SPD [3]. In summary, the GA provides an important lesson for numerically simplifying the SCDA at $\mathcal{N} = 2$ by exploiting the gauge freedom of the SPD, converting the problem of solving for \mathcal{G} into a constraint on the variational parameters of the SPD. In this paper, we demonstrate that the lessons of the GA can be generalized to $\mathcal{N} = 3$, which quantitatively captures the Mott and Hund physics of the multiorbital Hubbard model [3].

In order to demonstrate the power of the gauge constrained implementation of the SCDA at $\mathcal{N} = 3$, we study the $SU(2N_{\text{orb}})$ Hubbard model in $d = \infty$ with $N_{\text{orb}} = 2-8$. Our successful execution of these calculations demonstrates the viability of applying VDAT at $\mathcal{N} = 3$ to crystals bearing d or f electrons. Moreover, the $SU(2N_{\text{orb}})$ Hubbard model is interesting in its own right, given that experiments on ultracold atoms in an optical lattice can realize the $SU(2N_{\text{orb}})$ Hubbard model [16–21]. Therefore, VDAT should serve as a reliable tool for understanding and interpreting such experimental measurements.

The structure of this paper is as follows. Section II presents the gauge constrained algorithm of the SCDA at $\mathcal{N} = 3$, with Sec. II A providing a high-level overview of the entire algorithm, including all key equations, while the remaining sections provide detailed derivations. Section III presents results for the $SU(2N_{\text{orb}})$ Hubbard model in $d = \infty$ with $N_{\text{orb}} = 2-8$.

II. GAUGE CONSTRAINED ALGORITHM AT $\mathcal{N} = 3$

A. Overview

The goal of this section is to provide an overview of the gauge constrained algorithm for the SCDA at $\mathcal{N} = 3$, and Secs. II B, II C, and II D will derive all details of the procedure. We begin by highlighting how the SCDA exactly evaluates the SPD in $d = \infty$ [1–3]. Consider a fermionic lattice model having a Hamiltonian

$$\hat{H} = \hat{K} + \hat{H}_{\text{loc}} = \sum_{k\alpha\sigma} \epsilon_{k\alpha\sigma} \hat{n}_{k\alpha\sigma} + \sum_i \hat{H}_{\text{loc};i}, \quad (1)$$

where $i = 1, \dots, N_{\text{site}}$ enumerates over the lattice sites, $k = 1, \dots, N_{\text{site}}$ enumerates over the k points, $\alpha = 1, \dots, N_{\text{orb}}$

enumerates over the orbitals, and σ enumerates over spin. The G-type SPD for $\mathcal{N} = 3$ can be motivated from the following variational wave function:

$$\exp\left(\sum_{k\alpha\sigma} \gamma_{k\alpha\sigma} \hat{n}_{k\alpha\sigma}\right) \prod_i \hat{P}_i(u) |\Psi_0\rangle, \quad (2)$$

where $\{\gamma_{k\alpha\sigma}\}$ is the set of noninteracting variational parameters, $u = \{u_{i\Gamma}\}$ is the set of interacting variational parameters, $\hat{P}_i(u) = \sum_{\Gamma} u_{i\Gamma} \hat{P}_{i\Gamma}$, and $|\Psi_0\rangle$ is a noninteracting variational wave function; and both $\{\gamma_{k\alpha\sigma}\}$ and $\{u_{i\Gamma}\}$ are real numbers. The index Γ enumerates a set of many-body operators $\{\hat{P}_{i\Gamma}\}$ local to site i . The $\{\hat{P}_{i\Gamma}\}$ used for an $\hat{H}_{\text{loc};i}$ containing only density-density type interactions is given in Eq. (31), while the general case is given in Eq. (B2). It will be important to rewrite the above wave function as a density matrix, yielding the G-type SPD [1] for $\mathcal{N} = 3$

$$\hat{\varrho} = \hat{P}_1 \hat{P}_2 \hat{P}_3 = (\hat{K}_1 \hat{P}_1) (\hat{K}_2 \hat{P}_1) (\hat{K}_1), \quad (3)$$

where $\hat{K}_1 = \exp(\sum_{k\alpha\sigma} \gamma_{k\alpha\sigma} \hat{n}_{k\alpha\sigma})$, $\hat{P}_1 = \prod_i \hat{P}_i(u)$, and $\hat{K}_2 = |\Psi_0\rangle\langle\Psi_0|$. Here we have chosen \hat{K}_1 to be diagonal in $k\alpha\sigma$, while the most general case is addressed in Ref. [3].

Evaluating expectation values under the SPD is highly nontrivial, and we have developed the discrete action theory [1] to formalize the problem in a manner, which is amenable to systematic approximations. A key idea of the discrete action theory is the equivalence relation between an integer time correlation function and a corresponding expectation value in the compound space. An operator operator \hat{O} in the original space is promoted to the compound space with a given integer time index τ , denoted as $\hat{O}^{(\tau)}$ [1]. For the total energy with $\mathcal{N} = 3$, this equivalence is given as

$$\langle \hat{H} \rangle_{\hat{\varrho}} = \langle \hat{H}^{(\mathcal{N})} \rangle_{\hat{\varrho}}, \quad (4)$$

where $\hat{\varrho} = \hat{\varrho}_0 \prod_i \hat{P}_i$ is the discrete action of the SPD, $\hat{\varrho}_0 = \hat{Q} \hat{K}_1^{(1)} \hat{K}_2^{(2)} \hat{K}_1^{(3)}$ is the noninteracting discrete action, \hat{Q} is the integer time translation operator [1,3], and the interacting projector for site i is

$$\hat{P}_i = \hat{P}_i^{(1)}(u) \hat{P}_i^{(2)}(u) = \sum_{\Gamma\Gamma'} u_{i\Gamma} u_{i\Gamma'} \hat{P}_{i\Gamma}^{(1)} \hat{P}_{i\Gamma'}^{(2)}. \quad (5)$$

An important previous result is that the expectation value in the compound space can be exactly evaluated for $d = \infty$ using the self-consistent canonical discrete action theory (SCDA) [1,2].

Given the common scenario of translation symmetry, the SCDA can be presented in terms of two auxiliary effective discrete actions parameterized by $2N_{\text{orb}}\mathcal{N} \times 2N_{\text{orb}}\mathcal{N}$ matrices S_{loc} and \mathcal{G} , given as

$$\langle \hat{H}^{(\mathcal{N})} \rangle_{\hat{\varrho}} = \langle \hat{K}^{(\mathcal{N})} \rangle_{\hat{\varrho}_K} + N_{\text{site}} \langle \hat{H}_{\text{loc};i}^{(\mathcal{N})} \rangle_{\hat{\varrho}_{\text{loc};i}}, \quad (6)$$

$$\hat{\varrho}_K = \hat{\varrho}_0 \exp\left(-\sum_i \ln S_{\text{loc}}^T \cdot \hat{\mathbf{n}}_i\right), \quad (7)$$

$$\hat{\varrho}_{\text{loc};i} = \exp(-\ln(\mathcal{G}^{-1} - \mathbf{1})^T \cdot \hat{\mathbf{n}}_i) \hat{P}_i, \quad (8)$$

where $[\hat{\mathbf{n}}_i]_{\alpha\sigma\tau, \alpha'\sigma'\tau'} = \hat{a}_{i\alpha\sigma}^\dagger(\tau) \hat{a}_{i\alpha'\sigma'}(\tau')$, the dot product operation is defined as $\mathbf{a} \cdot \hat{\mathbf{b}} \equiv \sum_{\ell\ell'} [\mathbf{a}]_{\ell\ell'} [\hat{\mathbf{b}}]_{\ell\ell'}$, the discrete action $\hat{\varrho}_K$ is

used to compute all single-particle integer time Green's functions, and $\hat{\rho}_{\text{loc};i}$ is used to compute all N -particle integer time Green's functions local to site i . The \mathcal{G} and S_{loc} must satisfy the following two self-consistency conditions:

$$(\mathbf{g}_{\text{loc}}^{-1} - \mathbf{1}) = (\mathcal{G}^{-1} - \mathbf{1})S_{\text{loc}}, \quad (9)$$

$$\mathbf{g}_{\text{loc}} = \mathbf{g}'_{\text{loc}}, \quad (10)$$

where $\mathbf{g}_{\text{loc}} = \langle \hat{\mathbf{n}}_i \rangle_{\hat{\rho}_{\text{loc};i}}$ and $\mathbf{g}'_{\text{loc}} = \langle \hat{\mathbf{n}}_i \rangle_{\hat{\rho}_K}$.

One challenge posed by the SCDA is that the self-consistency condition must be satisfied for a given choice of variational parameters, which makes the minimization over the variational parameters nontrivial. An efficient algorithm for VDAT within the SCDA was proposed for the multiorbital Hubbard model for general \mathcal{N} , referred to as the decoupled minimization algorithm, and implemented in the two-orbital Hubbard model up to $\mathcal{N} = 4$ [3]. The decoupled minimization algorithm begins with an initial choice of variational parameters $\{\gamma_{k\alpha\sigma}\}$, u and an initial choice for \mathcal{G} , which determines the discrete action $\hat{\rho}_{\text{loc};i}$ [Eq. (8)], which yields \mathbf{g}_{loc} . Using the discrete Dyson equation [Eq. (9)], S_{loc} can be computed from \mathcal{G} and \mathbf{g}_{loc} . Then the discrete action $\hat{\rho}_K$ [Eq. (6)] can be used to compute \mathbf{g}'_{loc} . Using $\mathbf{g}_{\text{loc}} = \mathbf{g}'_{\text{loc}}$ in the discrete Dyson equation, a new \mathcal{G} can be obtained. During this self-consistency cycle, relevant first-order derivatives with regard to $\{\gamma_{k\alpha\sigma}\}$, u , and \mathcal{G} can be computed and two effective models can be constructed to update $\{\gamma_{k\alpha\sigma}\}$ and u . This entire procedure is iterated until $\{\gamma_{k\alpha\sigma}\}$, u , and \mathcal{G} are self-consistent. In a given iteration before reaching self-consistency, the energy and its gradients contain errors due to a deviation from the SCDA self-consistency condition, which can yield slow convergence in some regions of parameter space. Automatically satisfying the SCDA would yield a dramatic advantage when minimizing over the variational parameters.

In previous work [1], we demonstrated that the gauge freedom of the SPD can be used to automatically satisfy the SCDA self-consistency condition at $\mathcal{N} = 2$, which recovers the Gutzwiller approximation, and here we extend this line of reasoning to $\mathcal{N} = 3$. For simplicity, we use a restricted form of the SPD, where the kinetic projector is diagonal in k space and $\hat{P}_i(u)$ does not introduce off-diagonal terms at the level of the single-particle density matrix. Therefore, \mathcal{G} , S_{loc} , and \mathbf{g}_{loc} all have the form $[\mathbf{g}_{\text{loc}}]_{\alpha\sigma\tau,\alpha'\sigma'\tau'} = \delta_{\alpha\alpha'}\delta_{\sigma\sigma'}[\mathbf{g}_{\text{loc}}]_{\alpha\sigma\tau,\alpha\sigma\tau'}$, and the integer time Green's functions of each spin orbital are described by a 3×3 matrix. We begin by partitioning a local integer time 3×3 matrix $M_{\alpha\sigma}$ for a given spin orbital into submatrices as:

$$M_{\alpha\sigma} = \left(\begin{array}{cc|c} [M_{\alpha\sigma}]_{11} & [M_{\alpha\sigma}]_{12} & [M_{\alpha\sigma}]_{13} \\ [M_{\alpha\sigma}]_{21} & [M_{\alpha\sigma}]_{22} & [M_{\alpha\sigma}]_{23} \\ [M_{\alpha\sigma}]_{31} & [M_{\alpha\sigma}]_{32} & [M_{\alpha\sigma}]_{33} \end{array} \right) \quad (11)$$

$$= \begin{pmatrix} M_{\alpha\sigma;A} & M_{\alpha\sigma;B} \\ M_{\alpha\sigma;C} & M_{\alpha\sigma;D} \end{pmatrix}, \quad (12)$$

where M can be \mathcal{G} , S_{loc} , \mathbf{g}_{loc} , and \mathbf{g}'_{loc} . The main idea is to satisfy the self-consistency condition $\mathbf{g}_{\text{loc}} = \mathbf{g}'_{\text{loc}}$ in two stages: first for the A block and then for the B, C, and D blocks.

We proceed by outlining the logic and key equations of the first stage, which is treated in detail in Sec. II B and II C.

The first stage begins by considering $\mathcal{G}_{\alpha\sigma;A}$, a 2×2 matrix, which can be parametrized in terms of the single variable $\mathcal{G}_{\alpha\sigma;12}$ using the gauge freedom of the SPD, and $\mathcal{G}_{\alpha\sigma;12}$ should now be regarded as an independent variational parameter. The $\mathbf{g}_{\text{loc};A}$ and S_{loc} can be determined as a function of the sets $\mathcal{G}_{12} = \{\mathcal{G}_{\alpha\sigma;12}\}$ and $u = \{u_{i\Gamma}\}$, though we suppress the function arguments \mathcal{G}_{12} and u for brevity. The local density is also a function of \mathcal{G}_{12} and u , defined as $n_{\alpha\sigma}(\mathcal{G}_{12}, u) = [g_{\text{loc};\alpha\sigma}]_{22}$. For a given $\alpha\sigma$, the $S_{\text{loc};\alpha\sigma}$ can be parametrized using $S_{\alpha\sigma;11}$ and $S_{\alpha\sigma;12}$, and we can explicitly reparametrize the kinetic variational parameters as $n_{k\alpha\sigma;0}$ and $n_{k\alpha\sigma}$, where $n_{k\alpha\sigma;0} = 0, 1$ is the single-particle density matrix of \hat{K}_2 and $n_{k\alpha\sigma}$ is the single-particle density matrix of the SPD. It will be proven that $n_{k\alpha\sigma;0}$ determines the Fermi surface of both the interacting and noninteracting SPD, and therefore it will be useful to define two regions of momentum space, denoted as $<$ or $>$, where $<$ denotes the set of k points with $n_{k\alpha\sigma;0} = 1$ and $>$ indicates $n_{k\alpha\sigma;0} = 0$; and we assume $\int dk = 1$. For each region $X \in \{<, >\}$ of a given spin orbital $\alpha\sigma$, it will be useful to define the charge transfer $\Delta_{X\alpha\sigma}$ and charge fluctuation $\mathcal{A}_{X\alpha\sigma}$ as

$$\Delta_{X\alpha\sigma} = \int_X dk (n_{k\alpha\sigma;0} - n_{k\alpha\sigma}), \quad (13)$$

$$\mathcal{A}_{X\alpha\sigma} = \int_X dk \sqrt{n_{k\alpha\sigma}(1 - n_{k\alpha\sigma})}, \quad (14)$$

which measure the influence of the local interaction on the given spin orbital. Given that $\mathbf{g}'_{\text{loc};\alpha\sigma;A}$ is determined by $S_{\alpha\sigma;11}$, $S_{\alpha\sigma;12}$, $\{n_{k\alpha\sigma}\}$, and $\{n_{k\alpha\sigma;0}\}$, the self-consistency condition $\mathbf{g}_{\text{loc};A} = \mathbf{g}'_{\text{loc};A}$ becomes three linear constraints on $n_{k\alpha\sigma;0}$ and $n_{k\alpha\sigma}$, given as

$$\int n_{k\alpha\sigma;0} dk = n_{\alpha\sigma}(\mathcal{G}_{12}, u), \quad (15)$$

$$\int n_{k\alpha\sigma} dk = n_{\alpha\sigma}(\mathcal{G}_{12}, u), \quad (16)$$

$$\Delta_{<\alpha\sigma} = \Delta_{\alpha\sigma}(\mathcal{G}_{12}, u), \quad (17)$$

where

$$n_{\alpha\sigma}(\mathcal{G}_{12}, u) \equiv [g_{\text{loc};\alpha\sigma}]_{22}, \quad (18)$$

$$\Delta_{\alpha\sigma}(\mathcal{G}_{12}, u) \equiv \frac{S_{\alpha\sigma;12}}{S_{\alpha\sigma;11}} [g_{\text{loc};\alpha\sigma}]_{12}. \quad (19)$$

It should be noticed that the three constraints imply $\Delta_{<\alpha\sigma} = -\Delta_{>\alpha\sigma}$. The three constraints have a clear interpretation. Equation (15) indicates that the local density of the noninteracting reference system \hat{K}_2 is constrained to $[g_{\text{loc};\alpha\sigma}]_{22}$. Using Eq. (16), we see that $[g_{\text{loc};\alpha\sigma}]_{22} = [g_{\text{loc};\alpha\sigma}]_{33}$, dictating that the Fermi volume is equal to the local density obtained from the SPD. The result of these two constraints can be viewed as the wave function analog of the Luttinger theorem [22]. The third constraint, Eq. (17), reveals how the local interaction influences the density distribution. When the interacting projector is close to the identity, $S_{\alpha\sigma;12}$ approaches zero while the $S_{\alpha\sigma;11}$ and $[g_{\text{loc};\alpha\sigma}]_{12}$ approach a finite value, dictating that $\Delta_{\alpha\sigma}(\mathcal{G}_{12}, u)$ approaches zero and therefore $n_{k\alpha\sigma}$ approaches $n_{k\alpha\sigma;0}$. Alternatively, when the interacting projector deviates from the identity, $\Delta_{\alpha\sigma}(\mathcal{G}_{12}, u)$ increases and imposes a deviation of $n_{k\alpha\sigma}$ away from $n_{k\alpha\sigma;0}$. In summary, the first stage

enforces self-consistency for the A block, determining the kinetic energy.

In the second stage, $\mathbf{g}'_{\text{loc},\alpha\sigma}$ is fully determined by $S_{\alpha\sigma;11}$, $S_{\alpha\sigma;12}$, $n_{\alpha\sigma}(\mathcal{G}_{12}, u)$, $\Delta_{\alpha\sigma}(\mathcal{G}_{12}, u)$, $\mathcal{A}_{<\alpha\sigma}$, and $\mathcal{A}_{>\alpha\sigma}$, therefore $\mathbf{g}'_{\text{loc}} = \mathbf{g}'_{\text{loc}}$ can determine \mathcal{G} on the B, C, and D blocks, which is derived in Sec. IID. In summary, the self-consistency has been automatically satisfied and the local energy determined.

In conclusion, we have an explicit functional form for the total energy of the SPD parametrized by \mathcal{G}_{12} , u , and $n_{k\alpha\sigma}$, given as

$$\sum_{\alpha\sigma} \int dk \epsilon_{k\alpha\sigma} n_{k\alpha\sigma} + E_{\text{loc}}(\mathcal{G}_{12}, u, \mathcal{A}), \quad (20)$$

where $\mathcal{A} = \{\mathcal{A}_{<\alpha\sigma}, \mathcal{A}_{>\alpha\sigma}\}$ and $n_{k\alpha\sigma}$ is constrained by Eqs. (16) and (17) and the volume of Fermi sea is constrained by Eq. (15). The total energy has been expressed as a functional of \mathcal{G}_{12} , u , $\{n_{k\alpha\sigma}\}$, and $\{n_{k\alpha\sigma;0}\}$. This algorithm can be viewed as a nonlinear reparametrization of the original variational parameters $|\Psi_0\rangle$, u , and $\{\gamma_{k\alpha\sigma}\}$, where $\{n_{k\alpha\sigma;0}\}$ is a reparametrization of $|\Psi_0\rangle$, $\{n_{k\alpha\sigma}\}$ is a reparametrization of part of $\{\gamma_{k\alpha\sigma}\}$, and \mathcal{G}_{12} can be viewed as a set of variational parameters which reparametrizes the remaining part of $\{\gamma_{k\alpha\sigma}\}$ through condition (17).

It should be noted that $\{n_{k\alpha\sigma}\}$ only influences the local interaction energy through \mathcal{A} , and is constrained by $n_{\alpha\sigma}(\mathcal{G}_{12}, u)$ and $\Delta_{\alpha\sigma}(\mathcal{G}_{12}, u)$, and therefore to find an optimized $n_{k\alpha\sigma}$ in the region $X \in \{<, >\}$ of spin orbital $\alpha\sigma$, two Lagrange multipliers $a_{X\alpha\sigma}$ and $b_{X\alpha\sigma}$ can be introduced

$$F_{X\alpha\sigma} = \int_X dk (\epsilon_{k\alpha\sigma} n_{k\alpha\sigma} - a_{X\alpha\sigma} n_{k\alpha\sigma} - b_{X\alpha\sigma} \sqrt{n_{k\alpha\sigma}(1 - n_{k\alpha\sigma})}), \quad (21)$$

and we can solve for $n_{k\alpha\sigma}$ from $\frac{\delta F_{X\alpha\sigma}}{\delta n_{k\alpha\sigma}}|_{k \in X} = 0$, resulting in

$$n_{k\alpha\sigma}|_{k \in X} = \frac{1}{2} \left(1 + \frac{a_{X\alpha\sigma} - \epsilon_{k\alpha\sigma}}{\sqrt{(a_{X\alpha\sigma} - \epsilon_{k\alpha\sigma})^2 + b_{X\alpha\sigma}^2}} \right). \quad (22)$$

Therefore, the true independent variational parameters for the algorithm are \mathcal{G}_{12} , u , and $\mathbf{b} = \{b_{X\alpha\sigma}\}$, given that $\mathbf{a} = \{a_{X\alpha\sigma}\}$ can be determined as a function of \mathcal{G}_{12} , u , and \mathbf{b} through Eqs. (16) and (17). Finally, the ground-state energy can be determined as

$$\mathcal{E} = \min_{\mathcal{G}_{12}, u, \mathbf{b}} \left(\int dk \epsilon_{k\alpha\sigma} n_{k\alpha\sigma}(\mathbf{a}, \mathbf{b}) + E_{\text{loc}}(\mathcal{G}_{12}, u, \mathbf{b}) \right), \quad (23)$$

where the functional dependencies for $n_{k\alpha\sigma}(\mathbf{a}, \mathbf{b})$ are defined in Eq. (22) and $E_{\text{loc}}(\mathcal{G}_{12}, u, \mathbf{b})$ is detailed in the remaining sections. In this work, we used the Nelder-Mead algorithm [23] to perform the minimization in Eq. (23), which is a gradient-free algorithm. In some cases, it may be preferable to solve a Hamiltonian with fixed density $n_{\alpha\sigma}$, and this procedure is outlined in Appendix A.

It is useful to give some practical guidelines for the efficiency of the gauge constrained algorithm, which can roughly be broken down into two factors. First, there is the cost of evaluating expectation values under $\hat{\rho}_{\text{loc};i}$ [i.e., Eq. (29)], which will scale exponentially with the number of spin orbitals. Second, there is the number of independent variational parameters, which scales exponentially in the absence of symmetry.

The first factor is roughly independent of the symmetry of the Hamiltonian \hat{H} , which is being solved, while the second factor strongly depends on the symmetry. However, it is always possible to restrict the number of variational parameters in order to control the computational cost of the second factor, maintaining an upper bound for the total energy compared to the full variational minimization. Therefore, there are numerous avenues for engineering a minimal parametrization of the space of variational parameters. In the present paper, we study the $\text{SU}(2N_{\text{orb}})$ Hubbard model, where the high local symmetry results in a linear scaling for the number of variational parameters, and therefore the first factor completely dominates the computational cost.

B. Evaluating observables within the local A block

Here we will elucidate why the block structure introduced in Eq. (12) is the starting point for the gauge constrained algorithm. We begin by explaining why $\mathcal{G}_{\alpha\sigma;A}$ is the only block that needs to be considered when determining S_{loc} . Given that $\hat{\rho}$ only acts on the first and second integer time step, S_{loc} only has nontrivial elements on the A block, which are determined by $\mathcal{G}_{\alpha\sigma;A}$ and u (see Sec. VB in Ref. [1] for further background). Therefore, only the form of $\mathcal{G}_{\alpha\sigma;A}$ needs to be specified to initiate the algorithm.

We previously demonstrated that the gauge freedom of the SPD allows the following simple form [3]:

$$\mathcal{G}_{\alpha\sigma;A} = \begin{pmatrix} \frac{1}{2} & \mathcal{G}_{\alpha\sigma;12} \\ -\mathcal{G}_{\alpha\sigma;12} & \frac{1}{2} \end{pmatrix}, \quad (24)$$

where $\mathcal{G}_{\alpha\sigma;\tau\tau'} = [\mathcal{G}_{\alpha\sigma}]_{\tau\tau'}$ and $\mathcal{G}_{\alpha\sigma;12} \in [0, 1/2]$. Since the $\mathcal{G}_{\alpha\sigma;A}$ is completely determined, any observables within the local A block can now be explicitly determined. For any operator \hat{O} local to site i , the expectation value under $\hat{\rho}_{\text{loc};i}$ can be rewritten in terms of expectations values of the noninteracting part of $\hat{\rho}_{\text{loc};i}$ as

$$\langle \hat{O} \rangle_{\hat{\rho}_{\text{loc};i}} = \frac{\langle \hat{\rho}_i \hat{O} \rangle_{\hat{\rho}_{\text{loc};i,0}}}{\langle \hat{\rho}_i \rangle_{\hat{\rho}_{\text{loc};i,0}}}, \quad (25)$$

where

$$\hat{\rho}_{\text{loc};i,0} \equiv \exp(-\ln(\mathbf{G}^{-1} - \mathbf{1})^T \cdot \hat{\mathbf{n}}_i), \quad (26)$$

Using the form of $\hat{\rho}_i$ in Eq. (5), we have

$$\langle \hat{\rho}_i \hat{O} \rangle_{\hat{\rho}_{\text{loc};i,0}} = u^T (\hat{O})_u u, \quad (27)$$

where $u = (u_{i1}, \dots, u_{iN_\Gamma})^T$ is a N_Γ -element real vector, N_Γ is the number of local projectors, and $(\hat{O})_u$ is an $N_\Gamma \times N_\Gamma$ matrix with elements

$$[(\hat{O})_u]_{\Gamma\Gamma'} = \langle \hat{\rho}_{\Gamma\Gamma'}^{(1)} \hat{\rho}_{\Gamma\Gamma'}^{(2)} \hat{O} \rangle_{\hat{\rho}_{\text{loc};i,0}}. \quad (28)$$

It should be emphasized that the subscript in $(\hat{O})_u$ solely indicates that this matrix and the vector u are in the same representation, and the elements of $(\hat{O})_u$ defined in Eq. (28) are not dependent on the values of $u_{i\Gamma}$; a different representation, which is useful for constraining the density is presented in

Appendix A. The expectation value of \hat{O} under $\hat{\rho}_{\text{loc};i}$ is given as

$$\langle \hat{O} \rangle_{\hat{\rho}_{\text{loc};i}} = \frac{u^T (\hat{O})_u u}{u^T (\hat{1})_u u}. \quad (29)$$

For example, the local integer time Green's function can be computed as

$$[\mathbf{g}_{\text{loc};\alpha\sigma}]_{\tau\tau'} = \frac{u^T (\hat{a}_{\alpha\sigma}^{\dagger(\tau)} \hat{a}_{\alpha\sigma}^{(\tau')})_u u}{u^T (\hat{1})_u u}. \quad (30)$$

In the following, we present key formulas to evaluate Eq. (28). Given that we have restricted the SPD to be diagonal, the local projectors can be chosen as [3]

$$\hat{P}_{i\Gamma} = \prod_{\alpha\sigma} (\delta_{\Gamma_{\alpha\sigma},0} (1 - \hat{n}_{\alpha\sigma}) + \delta_{\Gamma_{\alpha\sigma},1} \hat{n}_{\alpha\sigma}), \quad (31)$$

where $\Gamma_{\alpha\sigma} \in \{0, 1\}$ and are determined from the binary relation $(\Gamma_{1\uparrow}\Gamma_{1\downarrow}\dots\Gamma_{N_{\text{orb}}\uparrow}\Gamma_{N_{\text{orb}}\downarrow})_2 = \Gamma - 1$. The matrix elements of $(\hat{1})_u$ are given as

$$[(\hat{1})_u]_{\Gamma\Gamma'} = \prod_{\alpha\sigma} p_{\alpha\sigma}(\Gamma_{\alpha\sigma}, \Gamma'_{\alpha\sigma}), \quad (32)$$

$$p_{\alpha\sigma}(\Gamma_{\alpha\sigma}, \Gamma'_{\alpha\sigma}) = \frac{1}{4} + (-1)^{\Gamma_{\alpha\sigma} + \Gamma'_{\alpha\sigma}} \mathcal{G}_{\alpha\sigma;12}^2. \quad (33)$$

Single-particle operators are evaluated as

$$\begin{aligned} [(\hat{a}_{\alpha\sigma}^{\dagger(\tau)} \hat{a}_{\alpha\sigma}^{(\tau')})_u]_{\Gamma\Gamma'} &= g_{\alpha\sigma}^{\tau\tau'}(\Gamma_{\alpha\sigma}, \Gamma'_{\alpha\sigma}) \\ &\times \prod_{\alpha'\sigma' \neq \alpha\sigma} p_{\alpha'\sigma'}(\Gamma_{\alpha'\sigma'}, \Gamma'_{\alpha'\sigma'}), \end{aligned} \quad (34)$$

where

$$\begin{aligned} g_{\alpha\sigma}^{\tau\tau'}(\Gamma_{\alpha\sigma}, \Gamma'_{\alpha\sigma}) &= p_{\alpha\sigma}(\Gamma_{\alpha\sigma}, \Gamma'_{\alpha\sigma}) \mathcal{G}_{\alpha\sigma;\tau\tau'} + (-1)^{\Gamma_{\alpha\sigma} + \Gamma'_{\alpha\sigma}} \\ &\times \left(\left(\frac{1}{2} (-1)^{\Gamma'_{\alpha\sigma} - 1} \mathcal{G}_{\alpha\sigma;1\tau'} \right. \right. \\ &\quad \left. \left. - \mathcal{G}_{\alpha\sigma;12} \mathcal{G}_{\alpha\sigma;2\tau'} \right) (\delta_{1,\tau} - \mathcal{G}_{\alpha\sigma;\tau 1}) \right. \\ &\quad \left. + \left(\frac{1}{2} (-1)^{\Gamma_{\alpha\sigma} - 1} \mathcal{G}_{\alpha\sigma;2\tau'} \right. \right. \\ &\quad \left. \left. + \mathcal{G}_{\alpha\sigma;12} \mathcal{G}_{\alpha\sigma;1\tau'} \right) (\delta_{2,\tau} - \mathcal{G}_{\alpha\sigma;\tau 2}) \right). \end{aligned} \quad (35)$$

Any two-particle correlation function of the below form are given as

$$\begin{aligned} &[(\hat{a}_{\alpha_1\sigma_1}^{\dagger(\tau_1)} \hat{a}_{\alpha_1\sigma_1}^{(\tau'_1)} \hat{a}_{\alpha_2\sigma_2}^{\dagger(\tau_2)} \hat{a}_{\alpha_2\sigma_2}^{(\tau'_2)})_u]_{\Gamma\Gamma'} \\ &= g_{\alpha_1\sigma_1}^{\tau_1\tau'_1}(\Gamma_{\alpha_1\sigma_1}, \Gamma'_{\alpha_1\sigma_1}) g_{\alpha_2\sigma_2}^{\tau_2\tau'_2}(\Gamma_{\alpha_2\sigma_2}, \Gamma'_{\alpha_2\sigma_2}) \\ &\times \prod_{\alpha'\sigma' \neq \alpha_1\sigma_1, \alpha_2\sigma_2} p_{\alpha'\sigma'}(\Gamma_{\alpha'\sigma'}, \Gamma'_{\alpha'\sigma'}), \end{aligned} \quad (36)$$

where $\alpha_1\sigma_1 \neq \alpha_2\sigma_2$. In Appendix B, we outline how to treat a general interacting projector. In summary, we have provided explicit formulas for evaluating local quantities up to the two-particle level, which is sufficient to execute the algorithm. It should be emphasized that these expressions for local observables are valid outside of the A block, but require complete knowledge of \mathcal{G} [e.g., see Eq. (35)].

Normally, evaluating expectation values under $\hat{\rho}_{\text{loc};i}$ [i.e., Eq. (29)] will be the rate limiting factor in the SCDA, and given that N_{Γ} scales exponentially with the number of spin orbitals, the overall computational cost will scale exponentially.

There are two possible routes to mitigate this exponential scaling. First, one could reduce the number of projectors, though this must be done carefully as it will limit the variational freedom. Second, one may use Monte Carlo to evaluate Eq. (29).

We now proceed to evaluate $\mathbf{S}_{\text{loc},\alpha\sigma}$. Given the choice of $\mathcal{G}_{\alpha\sigma;A}$ and using Eqs. (32) and (34), we find that $\mathbf{g}_{\alpha\sigma;A}$ has the following form:

$$\mathbf{g}_{\text{loc},\alpha\sigma;A} = \begin{pmatrix} n_{\alpha\sigma} & g_{\alpha\sigma;12} \\ -g_{\alpha\sigma;12} & n_{\alpha\sigma} \end{pmatrix}, \quad (37)$$

where $n_{\alpha\sigma}$ and $g_{\alpha\sigma;12}$ are functions of \mathcal{G}_{12} and u . Given that the local interacting projector only acts on the A block, the discrete Dyson equation simplifies to

$$(\mathbf{g}_{\text{loc},\alpha\sigma;A}^{-1} - \mathbf{1}) = (\mathcal{G}_{\alpha\sigma;A}^{-1} - \mathbf{1}) \mathbf{S}_{\text{loc},\alpha\sigma;A}, \quad (38)$$

which yields an integer time self-energy of the form

$$\mathbf{S}_{\text{loc},\alpha\sigma} = \begin{pmatrix} S_{\alpha\sigma;11} & S_{\alpha\sigma;12} & 0 \\ -S_{\alpha\sigma;12} & S_{\alpha\sigma;11} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (39)$$

where

$$\begin{aligned} S_{\alpha\sigma;11} &= \frac{1}{(4\mathcal{G}_{\alpha\sigma;12}^2 + 1)(g_{\alpha\sigma;12}^2 + n_{\alpha\sigma}^2)} \\ &\times (-g_{\alpha\sigma;12}^2 + 4g_{\alpha\sigma;12}\mathcal{G}_{\alpha\sigma;12} - n_{\alpha\sigma}^2 + n_{\alpha\sigma} \\ &\quad + 4\mathcal{G}_{\alpha\sigma;12}^2(g_{\alpha\sigma;12}^2 + (n_{\alpha\sigma} - 1)n_{\alpha\sigma})), \end{aligned} \quad (40)$$

$$\begin{aligned} S_{\alpha\sigma;12} &= \frac{1}{(4\mathcal{G}_{\alpha\sigma;12}^2 + 1)(g_{\alpha\sigma;12}^2 + n_{\alpha\sigma}^2)} \\ &\times (g_{\alpha\sigma;12}(4\mathcal{G}_{\alpha\sigma;12}(\mathcal{G}_{\alpha\sigma;12} - g_{\alpha\sigma;12}) - 1) \\ &\quad - 4(n_{\alpha\sigma} - 1)n_{\alpha\sigma}\mathcal{G}_{\alpha\sigma;12}). \end{aligned} \quad (41)$$

In summary, Eqs. (40) and (41) express the local integer time self-energy as a function of \mathcal{G}_{12} and u .

C. Parametrization of the integer time lattice Green's function and self-consistency of the A block

In the preceding section, we determined \mathbf{S}_{loc} , which completely determines $\hat{\rho}_K$ via Eq. (7), allowing the computation of $\mathbf{g}_{k\alpha\sigma} = \langle \hat{n}_{k\alpha\sigma} \rangle_{\hat{\rho}_K}$. We will demonstrate that $\mathbf{g}_{k\alpha\sigma}$ can be written analytically in terms of $\gamma_{k\alpha\sigma}$, the expectation value $n_{k\alpha\sigma;0} = \langle \hat{n}_{k\alpha\sigma} \rangle_{\hat{K}_2}$, and $\mathbf{S}_{\text{loc},\alpha\sigma}$. It is natural to reparametrize $\gamma_{k\alpha\sigma}$ using $\lambda_{k\alpha\sigma} = \langle \hat{n}_{k\alpha\sigma} \rangle_{\hat{K}_1} = [1 + \exp(-\gamma_{k\alpha\sigma})]^{-1} \in (0, 1)$ [1]. In general, \hat{K}_2 can be a mixed state, where $n_{k\alpha\sigma;0} \in [0, 1]$, and an analytic expression for $\mathbf{g}_{k\alpha\sigma}$ in terms of $\gamma_{k\alpha\sigma}$ and $n_{k\alpha\sigma;0}$ is given in the Appendix. At zero temperature in the metallic phase, \hat{K}_2 will be a pure state after minimization and $n_{k\alpha\sigma;0}$ is either zero or one. For the insulating phase at zero temperature, $\mathbf{g}_{k\alpha\sigma}$ does not depend on $n_{k\alpha\sigma;0}$, and therefore we are free to choose $n_{k\alpha\sigma;0} \in [0, 1]$, though for convenience we still choose zero or one. A general expression for $\mathbf{g}_{k\alpha\sigma}$ is presented in Eq. (S8) in Supplemental Material [24], which in the case of $n_{k\alpha\sigma;0} = 1$ reduces to

$$\mathbf{g}_{k\alpha\sigma} \Big|_{n_{k\alpha\sigma;0}=1} = C^{-1} \begin{pmatrix} C & 0 & 0 \\ -m_{21} & m_{22} & m_{23} \\ -m_{31} & -m_{32} & m_{22} \end{pmatrix}, \quad (42)$$

where

$$m_{21} = (\lambda_{k\alpha\sigma} - 1)^2 S_{\alpha\sigma;11}, \quad (43)$$

$$m_{22} = \lambda_{k\alpha\sigma}^2, \quad (44)$$

$$m_{23} = (1 - \lambda_{k\alpha\sigma}) \lambda_{k\alpha\sigma}, \quad (45)$$

$$m_{31} = (1 - \lambda_{k\alpha\sigma}) \lambda_{k\alpha\sigma} S_{\alpha\sigma;11}, \quad (46)$$

$$m_{32} = (1 - \lambda_{k\alpha\sigma}) \lambda_{k\alpha\sigma} S_{\alpha\sigma;12}, \quad (47)$$

$$C = \lambda_{k\alpha\sigma}^2 + (\lambda_{k\alpha\sigma} - 1)^2 S_{\alpha\sigma;12}. \quad (48)$$

Furthermore, it is natural to reparametrize $\lambda_{k\alpha\sigma}$ using $n_{k\alpha\sigma} = [\mathbf{g}_{k\alpha\sigma}]_{33} = \langle \hat{n}_{k\alpha\sigma}^{(3)} \rangle_{\hat{\rho}_k}$, which is the physical density distribution $\langle \hat{n}_{k\alpha\sigma} \rangle_{\hat{\rho}}$ within the SCDA, as

$$\lambda_{k\alpha\sigma} = \frac{n_{k\alpha\sigma} S_{\alpha\sigma;12}}{n_{k\alpha\sigma} S_{\alpha\sigma;12} + \sqrt{(1 - n_{k\alpha\sigma}) n_{k\alpha\sigma} S_{\alpha\sigma;12}}}, \quad (49)$$

resulting in

$$\mathbf{g}_{k\alpha\sigma} \Big|_{n_{k\alpha\sigma;0}=1} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{S_{\alpha\sigma;11}}{S_{\alpha\sigma;12}}(1 - n_{k\alpha\sigma}) & n_{k\alpha\sigma} & \frac{A}{\sqrt{S_{\alpha\sigma;12}}} \\ -\frac{S_{\alpha\sigma;11}}{\sqrt{S_{\alpha\sigma;12}}}A & -\sqrt{S_{\alpha\sigma;12}}A & n_{k\alpha\sigma} \end{pmatrix}, \quad (50)$$

where $A = \sqrt{(1 - n_{k\alpha\sigma}) n_{k\alpha\sigma}}$. For the case of $n_{k\alpha\sigma;0} = 0$, we have

$$\mathbf{g}_{k\alpha\sigma} \Big|_{n_{k\alpha\sigma;0}=0} = C^{-1} \begin{pmatrix} 0 & m_{12} & m_{13} \\ 0 & m_{22} & m_{23} \\ 0 & -m_{32} & m_{22} \end{pmatrix}, \quad (51)$$

where

$$m_{12} = \lambda_{k\alpha\sigma}^2 S_{\alpha\sigma;11}, \quad (52)$$

$$m_{13} = (1 - \lambda_{k\alpha\sigma}) \lambda_{k\alpha\sigma} S_{\alpha\sigma;11}, \quad (53)$$

$$m_{22} = \lambda_{k\alpha\sigma}^2 S_{\alpha\sigma;12}, \quad (54)$$

$$m_{23} = (1 - \lambda_{k\alpha\sigma}) \lambda_{k\alpha\sigma} S_{\alpha\sigma;12}, \quad (55)$$

$$m_{32} = (1 - \lambda_{k\alpha\sigma}) \lambda_{k\alpha\sigma} (S_{\alpha\sigma;11}^2 + S_{\alpha\sigma;12}^2), \quad (56)$$

$$C = (\lambda_{k\alpha\sigma} - 1)^2 (S_{\alpha\sigma;11}^2 + S_{\alpha\sigma;12}^2) + \lambda_{k\alpha\sigma}^2 S_{\alpha\sigma;12}. \quad (57)$$

We can similarly reparametrize $\lambda_{k\alpha\sigma}$ in terms of $n_{k\alpha\sigma}$ as

$$\lambda_{k\alpha\sigma} = \frac{\sqrt{n_{k\alpha\sigma}} S_{\alpha\sigma}}{\sqrt{n_{k\alpha\sigma}} S_{\alpha\sigma} + \sqrt{(1 - n_{k\alpha\sigma}) S_{\alpha\sigma;12}}}, \quad (58)$$

where $S_{\alpha\sigma} = \sqrt{S_{\alpha\sigma;11}^2 + S_{\alpha\sigma;12}^2}$, yielding

$$\mathbf{g}_{k\alpha\sigma} \Big|_{n_{k\alpha\sigma;0}=0} = \begin{pmatrix} 0 & \frac{S_{\alpha\sigma;11}}{S_{\alpha\sigma;12}} n_{k\alpha\sigma} & \frac{S_{\alpha\sigma;11}}{\sqrt{S_{\alpha\sigma;12} S_{\alpha\sigma}}} A \\ 0 & n_{k\alpha\sigma} & \frac{\sqrt{S_{\alpha\sigma;12}}}{S_{\alpha\sigma}} A \\ 0 & -\frac{S_{\alpha\sigma}}{\sqrt{S_{\alpha\sigma;12}}} A & n_{k\alpha\sigma} \end{pmatrix}, \quad (59)$$

where $A = \sqrt{(1 - n_{k\alpha\sigma}) n_{k\alpha\sigma}}$. The local integer time Green's function can now be constructed as an average over the Brillouin zone as

$$\mathbf{g}'_{\text{loc},\alpha\sigma} = \langle \hat{n}_{i\alpha\sigma} \rangle_{\hat{\rho}_K} = \int dk \mathbf{g}_{k\alpha\sigma}, \quad (60)$$

using the convention $\int dk = 1$.

Using the self-consistency condition on the A block,

$$\mathbf{g}'_{\text{loc},\alpha\sigma;A} = \mathbf{g}_{\text{loc},\alpha\sigma;A}, \quad (61)$$

we can determine the resulting constraints on $n_{k\alpha\sigma;0}$, $n_{k\alpha\sigma}$, u , and $\mathcal{G}_{\alpha\sigma;12}$. There are four constraining equations from the four corresponding entries of the A block, but only three of them are independent. The first constraint is $[\mathbf{g}'_{\text{loc},\alpha\sigma}]_{11} = [\mathbf{g}_{\text{loc},\alpha\sigma}]_{11}$, which yields

$$\int dk n_{k\alpha\sigma;0} = \int_{<} dk = n_{\alpha\sigma}(\mathcal{G}_{12}, u), \quad (62)$$

where $n_{\alpha\sigma}(\mathcal{G}_{12}, u) = [\mathbf{g}_{\text{loc},\alpha\sigma}]_{11} = [\mathbf{g}_{\text{loc},\alpha\sigma}]_{22}$, the symbol $<$ denotes the region where $n_{k\alpha\sigma;0} = 1$, while $>$ denotes the region where $n_{k\alpha\sigma;0} = 0$. The first constraint requires that $|\Psi_0\rangle$ has the same density as given by $n_{\alpha\sigma}(\mathcal{G}_{12}, u)$, and we refer to this as the Fermi volume constraint. The second constraint is $[\mathbf{g}'_{\text{loc},\alpha\sigma}]_{22} = [\mathbf{g}_{\text{loc},\alpha\sigma}]_{22}$, which yields the density constraint

$$\int dk n_{k\alpha\sigma} = \int_{<} dk n_{k\alpha\sigma} + \int_{>} dk n_{k\alpha\sigma} = n_{\alpha\sigma}(\mathcal{G}_{12}, u). \quad (63)$$

The third constraint is $[\mathbf{g}'_{\text{loc},\alpha\sigma}]_{12} = [\mathbf{g}_{\text{loc},\alpha\sigma}]_{12}$, which yields

$$\int_{>} dk n_{k\alpha\sigma} = \Delta_{\alpha\sigma}(\mathcal{G}_{12}, u), \quad (64)$$

where $\Delta_{\alpha\sigma}(\mathcal{G}_{12}, u) \equiv \frac{S_{\alpha\sigma;12}}{S_{\alpha\sigma;11}} [\mathbf{g}_{\text{loc},\alpha\sigma}]_{12}$. The fourth constraint is $[\mathbf{g}'_{\text{loc},\alpha\sigma}]_{21} = [\mathbf{g}_{\text{loc},\alpha\sigma}]_{21}$, which yields

$$\int_{<} dk (1 - n_{k\alpha\sigma}) = \Delta_{\alpha\sigma}(\mathcal{G}_{12}, u). \quad (65)$$

The third and fourth constraint are identical as long as the first and second constraint are satisfied,

$$\int_{<} dk (1 - n_{k\alpha\sigma}) = \int_{>} dk n_{k\alpha\sigma} = \Delta_{\alpha\sigma}(\mathcal{G}_{12}, u), \quad (66)$$

which we refer to as the charge transfer constraint.

We now discuss how to satisfy these three constraints, using constraints on $n_{k\alpha\sigma;0}$ and $n_{k\alpha\sigma}$. One can start with arbitrary u and $\mathcal{G}_{\alpha\sigma;12} \in [0, 1/2]$, which yields some $n_{\alpha\sigma}(\mathcal{G}_{12}, u)$ that determines the Fermi volume and $n_{k\alpha\sigma;0}$. Furthermore, one must choose $n_{k\alpha\sigma}$ such that Eqs. (63) and (66) are satisfied. To simplify the expression for \mathbf{g}'_{loc} , it is useful to define the following quantities:

$$\Delta_{<\alpha\sigma} = \int_{<} dk (n_{k\alpha\sigma;0} - n_{k\alpha\sigma}) = \int_{<} dk (1 - n_{k\alpha\sigma}), \quad (67)$$

$$\Delta_{>\alpha\sigma} = \int_{>} dk (n_{k\alpha\sigma;0} - n_{k\alpha\sigma}) = - \int_{>} dk n_{k\alpha\sigma}, \quad (68)$$

$$\mathcal{A}_{<\alpha\sigma} = \int_{<} dk \sqrt{n_{k\alpha\sigma} (1 - n_{k\alpha\sigma})}, \quad (69)$$

$$\mathcal{A}_{>\alpha\sigma} = \int_{>} dk \sqrt{n_{k\alpha\sigma} (1 - n_{k\alpha\sigma})}. \quad (70)$$

Using Eqs. (62) and (63), we have $\Delta_{<\alpha\sigma} = -\Delta_{>\alpha\sigma}$. Equations (62) and (63) are treated as independent conditions, and Eqs. (64) and (65) become the single condition given in Eq. (66). For convenience, we define

$$\Delta_{\alpha\sigma} \equiv \Delta_{<\alpha\sigma} = -\Delta_{>\alpha\sigma}, \quad (71)$$

$$n_{\alpha\sigma} \equiv \int_{<} dk = \int dk n_{k\alpha\sigma}, \quad (72)$$

which should not be confused with the corresponding quantities $\Delta_{\alpha\sigma}(\mathcal{G}_{12}, u)$ and $n_{\alpha\sigma}(\mathcal{G}_{12}, u)$ determined from the local discrete action. The quantity $\Delta_{\alpha\sigma}$ measures the total charge transfer generated by the projector $\hat{K}_1 \hat{P}_1$ across the Fermi surface determined by $|\Psi_0\rangle$, and is uniquely determined from $\hat{\rho}_{\text{loc};i}$. The noninteracting case yields $\Delta_{\alpha\sigma} = 0$, while the strong coupling limit of the Mott insulating phase yields $n_{k\alpha\sigma} = n_{\alpha\sigma}$ and $\Delta_{\alpha\sigma} = n_{\alpha\sigma}(1 - n_{\alpha\sigma})$. Once $n_{k\alpha\sigma}$ and $n_{k\alpha\sigma;0}$ have been constrained by Eqs. (62), (63), and (66), the total kinetic energy can be evaluated as

$$K = \sum_{\alpha\sigma} \int dk \epsilon_{k\alpha\sigma} n_{k\alpha\sigma}. \quad (73)$$

We now discuss the properties of $\mathcal{A}_{<\alpha\sigma}$ and $\mathcal{A}_{>\alpha\sigma}$, which are relevant for evaluating the B, C, and D blocks of $\mathbf{g}'_{\text{loc};\alpha\sigma}$. The noninteracting case yields $\mathcal{A}_{<\alpha\sigma} = \mathcal{A}_{>\alpha\sigma} = 0$, while for a given $n_{\alpha\sigma}$ and $\Delta_{\alpha\sigma}$ the maximum value of $\mathcal{A}_{<\alpha\sigma}$ is reached when $n_{k\alpha\sigma}|_{k \in <} = 1 - \Delta_{\alpha\sigma}/n_{\alpha\sigma}$, and therefore $\mathcal{A}_{<\alpha\sigma} \in [0, \sqrt{(n_{\alpha\sigma} - \Delta_{\alpha\sigma})\Delta_{\alpha\sigma}}]$. Similarly, the maximum of $\mathcal{A}_{>\alpha\sigma}$ is reached when $n_{k\alpha\sigma}|_{k \in >} = \Delta_{\alpha\sigma}/(1 - n_{\alpha\sigma})$ and therefore $\mathcal{A}_{>\alpha\sigma} \in [0, \sqrt{(1 - n_{\alpha\sigma} - \Delta_{\alpha\sigma})\Delta_{\alpha\sigma}}]$. Using these definitions, we have

$$\int_{<} dk \mathbf{g}_{k\alpha\sigma} = \begin{pmatrix} n_{\alpha\sigma} & 0 & 0 \\ -\frac{S_{\alpha\sigma;11}}{S_{\alpha\sigma;12}} \Delta_{\alpha\sigma} & n_{\alpha\sigma} - \Delta_{\alpha\sigma} & \frac{\mathcal{A}_{<\alpha\sigma}}{\sqrt{S_{\alpha\sigma;12}}} \\ -\frac{S_{\alpha\sigma;11}}{\sqrt{S_{\alpha\sigma;12}}} \mathcal{A}_{<\alpha\sigma} & -\sqrt{S_{\alpha\sigma;12}} \mathcal{A}_{<\alpha\sigma} & n_{\alpha\sigma} - \Delta_{\alpha\sigma} \end{pmatrix}, \quad (74)$$

and

$$\int_{>} dk \mathbf{g}_{k\alpha\sigma} = \begin{pmatrix} 0 & \frac{S_{\alpha\sigma;11}}{S_{\alpha\sigma;12}} \Delta_{\alpha\sigma} & \frac{S_{\alpha\sigma;11}}{\sqrt{S_{\alpha\sigma;12} S_{\alpha\sigma}}} \mathcal{A}_{>\alpha\sigma} \\ 0 & \Delta_{\alpha\sigma} & \frac{\sqrt{S_{\alpha\sigma;12}}}{S_{\alpha\sigma}} \mathcal{A}_{>\alpha\sigma} \\ 0 & -\frac{S_{\alpha\sigma}}{\sqrt{S_{\alpha\sigma;12}}} \mathcal{A}_{>\alpha\sigma} & \Delta_{\alpha\sigma} \end{pmatrix}, \quad (75)$$

yielding the local lattice Green's function

$$\mathbf{g}'_{\text{loc};\alpha\sigma} = \begin{pmatrix} n_{\alpha\sigma} & \frac{S_{\alpha\sigma;11}}{S_{\alpha\sigma;12}} \Delta_{\alpha\sigma} & g_{13} \\ -\frac{S_{\alpha\sigma;11}}{S_{\alpha\sigma;12}} \Delta_{\alpha\sigma} & n_{\alpha\sigma} & g_{23} \\ g_{31} & g_{32} & n_{\alpha\sigma} \end{pmatrix}, \quad (76)$$

where

$$g_{13} = \frac{S_{\alpha\sigma;11}}{\sqrt{S_{\alpha\sigma;12} S_{\alpha\sigma}}} \mathcal{A}_{>\alpha\sigma}, \quad (77)$$

$$g_{23} = \frac{1}{\sqrt{S_{\alpha\sigma;12}}} \mathcal{A}_{<\alpha\sigma} + \frac{\sqrt{S_{\alpha\sigma;12}}}{S_{\alpha\sigma}} \mathcal{A}_{>\alpha\sigma}, \quad (78)$$

$$g_{31} = -\frac{S_{\alpha\sigma;11}}{\sqrt{S_{\alpha\sigma;12}}} \mathcal{A}_{<\alpha\sigma}, \quad (79)$$

$$g_{32} = -\sqrt{S_{\alpha\sigma;12}} \mathcal{A}_{<\alpha\sigma} - \frac{S_{\alpha\sigma}}{\sqrt{S_{\alpha\sigma;12}}} \mathcal{A}_{>\alpha\sigma}. \quad (80)$$

The above equations provide explicit expressions for all blocks of $\mathbf{g}'_{\text{loc};\alpha\sigma}$.

D. Determining the B, C, and D blocks of \mathcal{G} and evaluating the total energy

Assuming that \mathcal{S}_{loc} and \mathbf{g}'_{loc} have been completely determined, the self-consistency condition $\mathbf{g}_{\text{loc}} = \mathbf{g}'_{\text{loc}}$ can be used to determine the remaining blocks of \mathcal{G} via the discrete Dyson equation as

$$\mathcal{G} = (\mathbf{1} + (\mathbf{g}'_{\text{loc}} - \mathbf{1})\mathcal{S}_{\text{loc}}^{-1})^{-1}, \quad (81)$$

which yields

$$\mathcal{G}_B = \mathcal{G} \mathbf{g}'_{\text{loc};A} \mathcal{S}_{\text{loc};B}^{-1}, \quad (82)$$

$$\mathcal{G}_C = \mathbf{g}_{\text{loc};C} (\mathbf{1} - \mathbf{g}_{\text{loc};A})^{-1} (\mathbf{1} - \mathcal{G}_A), \quad (83)$$

$$\mathcal{G}_D = \mathbf{g}_{\text{loc};D} + \mathbf{g}_{\text{loc};C} (\mathbf{1} - \mathbf{g}_{\text{loc};A})^{-1} (\mathbf{g}_{\text{loc};A} - \mathcal{G}_A) \mathbf{g}'_{\text{loc};A} \mathcal{S}_{\text{loc};B}^{-1}. \quad (84)$$

The individual matrix elements of \mathcal{G} are provided in Supplemental Material [24].

We now proceed to compute the total energy. Having completely determined \mathcal{G} , the local interaction energy can be computed using Eq. (29) as

$$E_{\text{loc}} = \frac{u^T (\hat{H}_{\text{loc}}^{(3)})_u u}{u^T (\hat{\mathbf{1}})_u u}, \quad (85)$$

where the matrix $(\hat{H}_{\text{loc}}^{(3)})_u$ depends on \mathcal{G} and the matrix $(\hat{\mathbf{1}})_u$ depends on \mathcal{G}_A . It should be emphasized that the density distribution $n_{k\alpha\sigma}$ will influence E_{loc} through \mathcal{G}_B and \mathcal{G}_C .

III. RESULTS FOR SU(2N_{orb}) HUBBARD MODEL

We now proceed to illustrate the gauge constrained implementation of VDAT within the SCDA at $\mathcal{N} = 3$. We choose to study the SU(2N_{orb}) Hubbard model, where $\hat{H}_{\text{loc};i} = U \sum_{\ell < \ell'} \hat{n}_{i\ell} \hat{n}_{i\ell'}$ and $\ell = 1, \dots, 2N_{\text{orb}}$, in order to showcase the advantages of VDAT over DMFT for obtaining zero-temperature results at large N_{orb}. It should be noted that the SU(2N_{orb}) symmetry can be exploited when evaluating the expectation values of local observables [i.e., Eq. (29)], but for purposes of benchmarking the computational cost, we utilized the general algorithm, which does not exploit the SU(2N_{orb}) symmetry. Therefore, for a single evaluation of the SPD, the computational cost is the same as the general case. To provide a rough idea of the computational cost on a typical single processor core, when N_{orb} ≤ 3 the cost of a single evaluation of

the SPD is 10^{-4} seconds. For $N_{\text{orb}} > 3$, the computational cost scales exponentially, requiring 10^{-3} , 0.02, 0.1, and 3 s for N_{orb} of 4, 5, 6, and 7, respectively. The timing difference between $\mathcal{N} = 2$ and $\mathcal{N} = 3$ is negligible in all the aforementioned cases. We study the $SU(2N_{\text{orb}})$ Hubbard model on the $d = \infty$ Bethe lattice for $N_{\text{orb}} = 2-8$ at half-filling and $2N_{\text{orb}} = 5$ for all fillings. At half-filling for $N_{\text{orb}} = 2, 4$ we compare to a zero-temperature extrapolation of published DMFT results using a quantum Monte Carlo (QMC) impurity solver [25], while for $2N_{\text{orb}} = 5$ we compare to published DMFT results using the numerical renormalization group (NRG) impurity solver [26]. All VDAT results in this paper are generated by a Julia implementation of the gauge constrained algorithm, which can treat general density-density interactions, and is available at Ref. [27].

We begin by considering the case of half-filling for $N_{\text{orb}} = 2-8$, exceeding the number of orbitals that would be encountered for a d or f electron manifold relevant to strongly correlated electron materials. The basic aspects of the $SU(2N_{\text{orb}})$ Hubbard model in $d = \infty$ are well understood: there is a metal-insulator transition (MIT) at a critical value of U , and this transition value increases with N_{orb} . The signature of the MIT can be seen in the variational parameters of the SPD, and therefore it is instructive to examine $\mathcal{G}_{\alpha\sigma;12}$, $b_{\alpha\sigma<}$, and the local variational parameters $u = \{u_{\Gamma}\}$ (see Fig. 1). It is also useful to explore the intermediate parameters $a_{\alpha\sigma<}$, which can be determined from the variational parameters. Because of the $SU(2N_{\text{orb}})$ symmetry, the $\mathcal{G}_{\alpha\sigma;12}$, $b_{\alpha\sigma<}$, and $a_{\alpha\sigma<}$ are independent of $\alpha\sigma$, while u_{Γ} only depends on the particle number of Γ . Furthermore, particle-hole symmetry at half-filling dictates that $b_{\alpha\sigma<} = b_{\alpha\sigma>}$, $a_{\alpha\sigma<} = -a_{\alpha\sigma>}$, and $u_{\Gamma} = u_{\Gamma'}$ if the sum of the particle numbers of Γ and Γ' is $2N_{\text{orb}}$. Considering $\mathcal{G}_{\alpha\sigma;12}$, it begins at $U = 0$ in the metallic phase with a value of roughly 0.37 and monotonically increases to 0.5 at the MIT, and is fixed at 0.5 in the insulating regime [see Fig. 1(a)]. Increasing the number of orbitals from $N_{\text{orb}} = 2-8$ has several effects. In the small U regime, $\mathcal{G}_{\alpha\sigma;12}$ increases with N_{orb} , enhancing the electronic correlations for larger N_{orb} . Alternatively, increasing N_{orb} will increase the critical value of U for the metal-insulator transition. The variational parameter $b_{\alpha\sigma<}$ turns out to be approximately equal to U , and therefore we also plot $b_{\alpha\sigma<} - U$ [see Fig. 1(b)]. The intermediate parameter $a_{\alpha\sigma<}$, which can be computed from the variational parameters, has a value of $2t$ for $U = 0$, independent of N_{orb} , and goes to zero in the insulating phase [see Fig. 1(c)]. The latter can be appreciated from Eq. (22), which dictates that the quasiparticle weight becomes zero when $a_{\alpha\sigma<} = a_{\alpha\sigma>}$ and $b_{\alpha\sigma<} = b_{\alpha\sigma>}$. For a given N_{orb} , there is an arbitrary coefficient for the interacting projector, which can be fixed by choosing $u_{\Gamma} = 1$ when the particle number of Γ is N_{orb} . Therefore, there are N_{orb} independent local variational parameters u_{Γ} for Γ with particle number equal to $0, \dots, N_{\text{orb}} - 1$. For a given N_{orb} , we plot the u_{Γ} where the particle number for Γ is $N_{\text{orb}} - 1$ [see Fig. 1(d)], in addition to all u_{Γ} for $N_{\text{orb}} = 5$. The u_{Γ} for Γ with particle number $N_{\text{orb}} - 1$ goes to zero in the insulating phase.

We now consider the total energy, where we explore both $\mathcal{N} = 2$ and $\mathcal{N} = 3$ [see Fig. 2(a)]. For clarity, we plot $\Delta E = E(t, U) - E(0, U)$ divided by the number of spin orbitals.

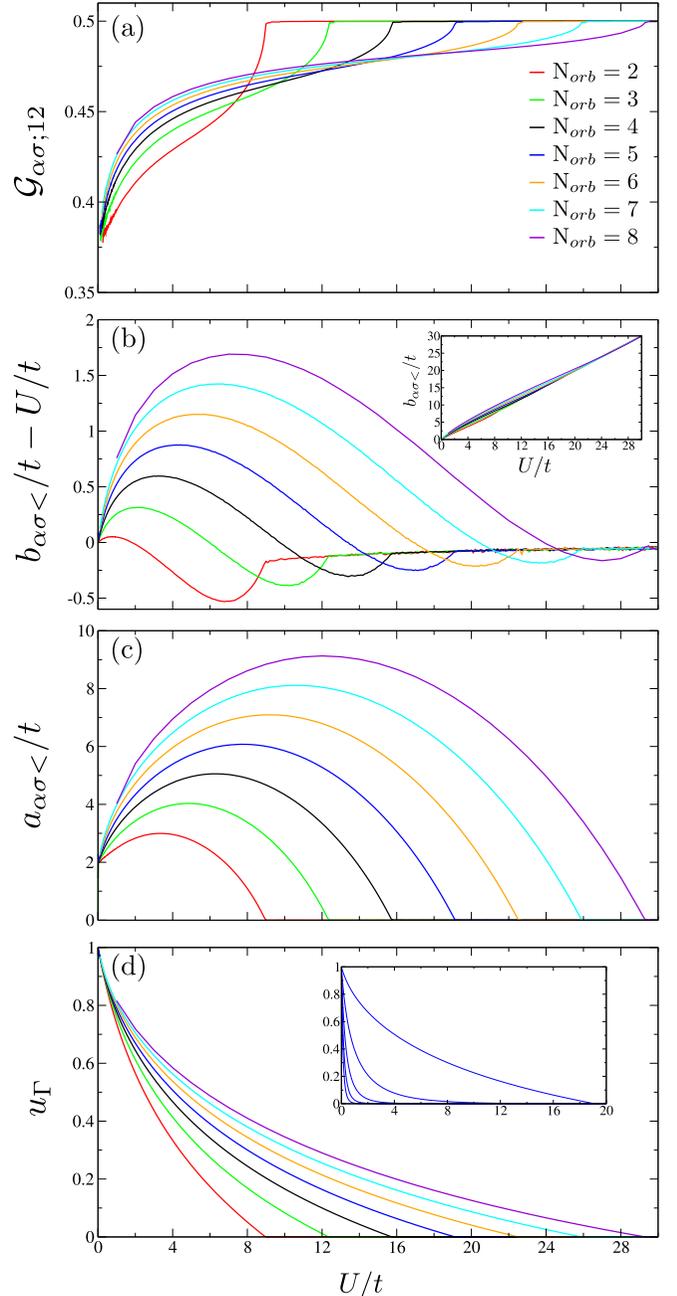


FIG. 1. Optimized variational parameters of VDAT within the SCDA at $\mathcal{N} = 3$ for the $SU(2N_{\text{orb}})$ Hubbard model on the $d = \infty$ Bethe lattice at half-filling and $T = 0$, with $N_{\text{orb}} = 2-8$. The variational parameters $\mathcal{G}_{\alpha\sigma;12}$, $b_{\alpha\sigma<}$, and u_{Γ} for Γ with particle number $N_{\text{orb}} - 1$ are shown in (a), (b), and (d), respectively, while the dependent parameters $a_{\alpha\sigma<}$ are shown in (c). The inset of (d) shows all u_{Γ} for $N_{\text{orb}} = 5$, and the value monotonically decreases as the particle number of Γ decreases from four to zero.

The $\mathcal{N} = 3$ results are always lower in energy than the $\mathcal{N} = 2$ results, as is required by the variational principle. Furthermore, the $\mathcal{N} = 3$ results resolve the well-known deficiency of the $\mathcal{N} = 2$ results in the insulating regime. Interestingly, in the small and large U regimes, the quantity $\Delta E/(2N_{\text{orb}})$ is largely independent of N_{orb} , while in the intermediate regime it decreases with N_{orb} . Similar behavior is observed in the

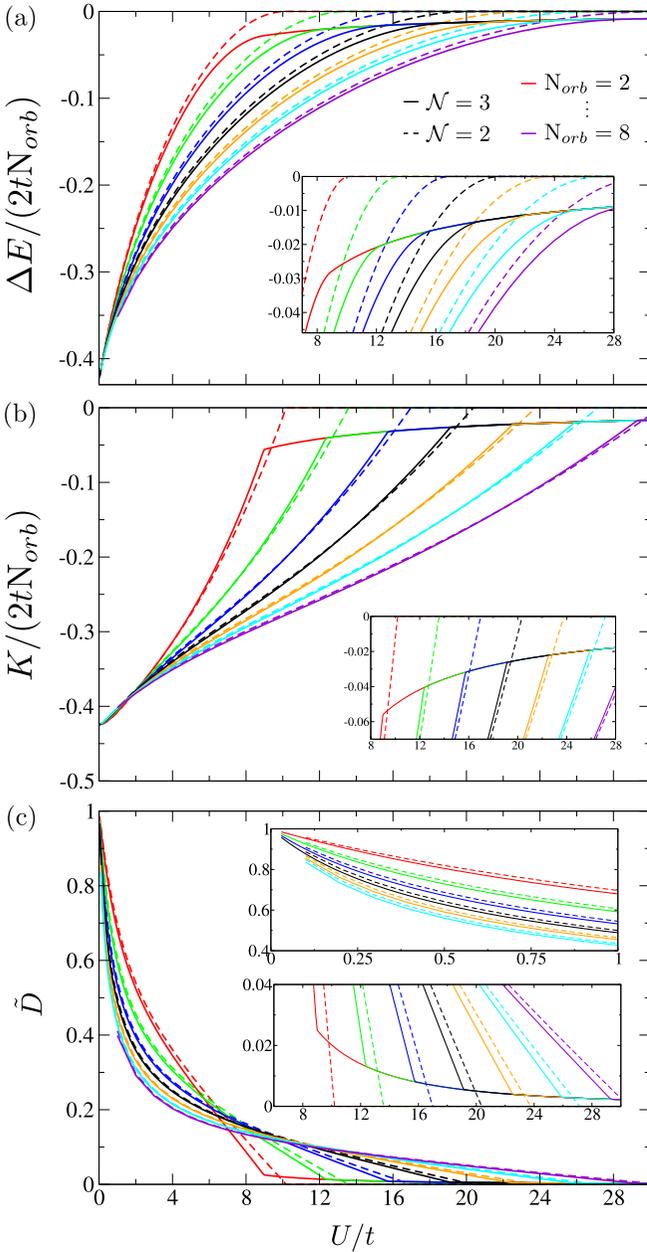


FIG. 2. Static observables computed from VDAT within the SCDA at $\mathcal{N} = 2$ (dashed lines) and $\mathcal{N} = 3$ (solid lines) for the $SU(2N_{\text{orb}})$ Hubbard model on the $d = \infty$ Bethe lattice at half-filling and $T = 0$, with $N_{\text{orb}} = 2-8$. The legend colors are identical to Fig. 1. (a) Total energy difference per spin orbital vs. U/t , where $\Delta E = E(t, U) - E(0, U)$. (b) Kinetic energy per spin orbital vs. U/t . (c) Scaled double occupancy vs. U/t , where $\tilde{D} = (D - D_{\text{at}})/(D_0 - D_{\text{at}})$; D , D_0 , and D_{at} are the double occupancies at the given U/t , the noninteracting value, and the atomic value, respectively.

kinetic energy per orbital [see Fig. 2(b)]. The double occupancy D determines the interaction energy, and for convenience we plot the scaled double occupancy $\tilde{D} = (D - D_{\text{at}})/(D_0 - D_{\text{at}})$, where D_0 and D_{at} are the noninteracting and atomic double occupancy, respectively [see Fig. 2(c)]. In the small U regime, the scaled double occupancy decreases with N_{orb} , while in the large U regime it is independent of N_{orb} .

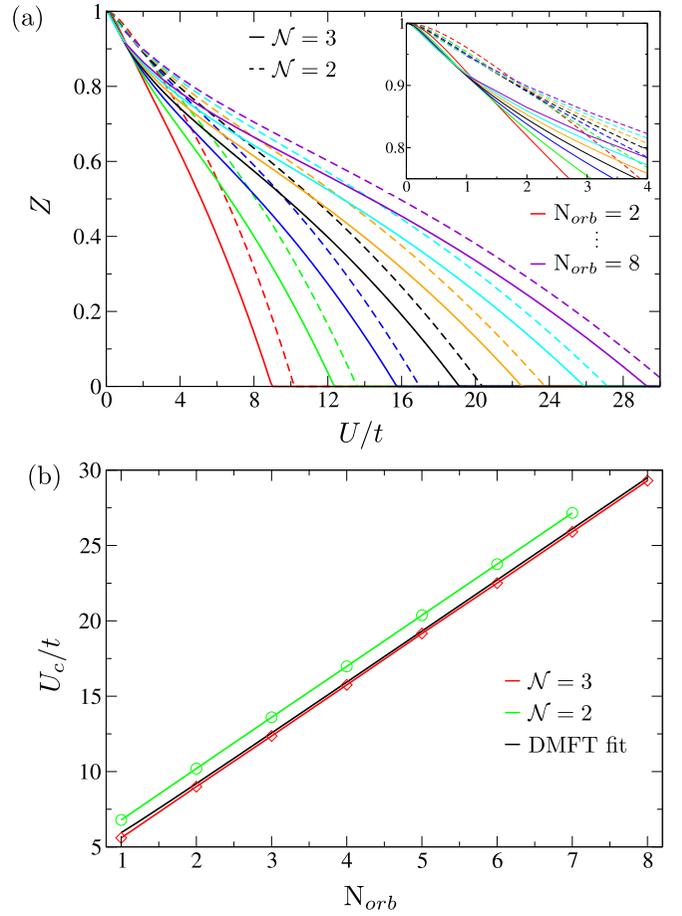


FIG. 3. Results for VDAT within the SCDA at $\mathcal{N} = 2$ and $\mathcal{N} = 3$ for the $SU(2N_{\text{orb}})$ Hubbard model on the $d = \infty$ Bethe lattice at half-filling and $T = 0$, with $N_{\text{orb}} = 2-8$. (a) Quasiparticle weight. The legend is identical to Fig. 1. (b) The critical value of U for the MIT, denoted U_c , as a function of N_{orb} . The DMFT curve is a plot of a previously published fit to DMFT results [25], given as $U_c/t = 1.7(2N_{\text{orb}} + 1)(1 + 0.166N_{\text{orb}}^{-1})$.

It is also interesting to evaluate the quasiparticle weight as a function of U/t , which determines the critical value of U for the MIT [see Fig. 3(a)]. The $\mathcal{N} = 2$ result always produces a large quasiparticle weight than $\mathcal{N} = 3$, and therefore produces a larger critical value of U for the MIT. Interestingly, for $\mathcal{N} = 3$ at approximately $U/t = 1$, the quasiparticle weight is insensitive to N_{orb} , and a similar effect is observed for $\mathcal{N} = 2$ at a slightly large value of U/t . We also examine the transition value U_c/t as a function of N_{orb} and compare to the previously published scaling relation [25] $U_c/t = 1.7(2N_{\text{orb}} + 1)(1 + 0.166N_{\text{orb}}^{-1})$ extracted from DMFT calculations, which use QMC as the impurity solver [see Fig. 3(b)]. The $\mathcal{N} = 2$ case recovers the previously published results of the Gutzwiller approximation [28,29], yielding a linear relation. Interestingly, $\mathcal{N} = 3$ also produces a nearly linear result, but the result is shifted downward, nearly coinciding with the DMFT extrapolation.

We now proceed to compare the total energy at half-filling and zero temperature from VDAT and DMFT. Given that the previously published DMFT results are at a finite temperature

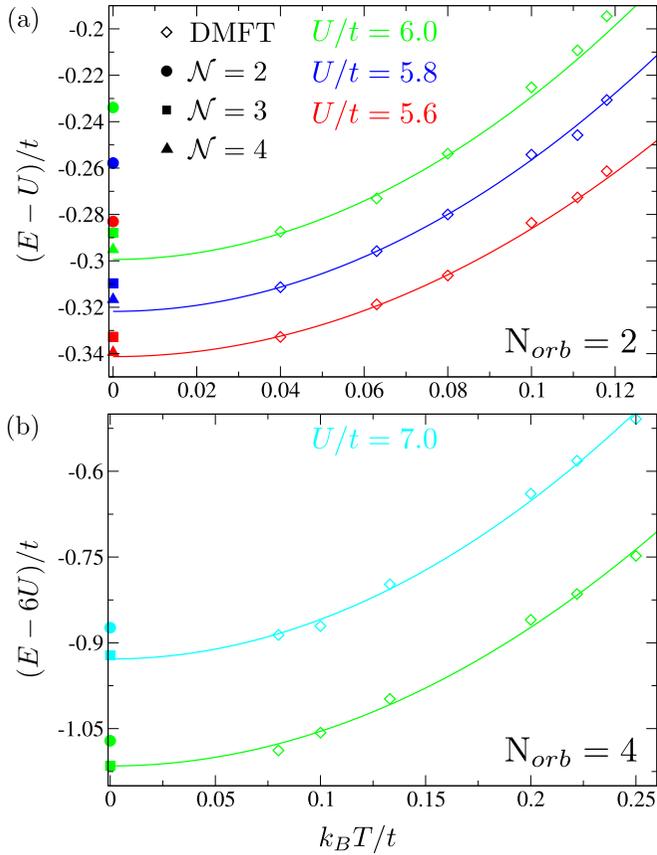


FIG. 4. The total energy of the $SU(2N_{\text{orb}})$ Hubbard model on the $d = \infty$ Bethe lattice at half-filling from published DMFT calculations solved using QMC at a finite temperature [25] and VDAT within the SCDA at zero temperature using $\mathcal{N} = 2-4$, with $N_{\text{orb}} = 2$ in (a) and $N_{\text{orb}} = 4$ in (b). The lines are a fit to the DMFT results assuming $E(T) - E(0) \propto T^2$.

[25], a quadratic fit assuming $E(T) - E(0) \propto T^2$ was used to extrapolate to zero temperature (see Fig. 4). We were only able to examine select cases where the QMC data sufficiently resembled a quadratic. For $N_{\text{orb}} = 2$, we present VDAT results for $\mathcal{N} = 2-4$, which were previously published in Ref. [3], while for $N_{\text{orb}} = 4$ we present VDAT results for $\mathcal{N} = 2-3$. As required by the variational principle, the energy within VDAT monotonically decreases with increasing \mathcal{N} . The dramatic improvement of $\mathcal{N} = 3$ over $\mathcal{N} = 2$ is clearly illustrated, and it should be recalled that these two cases have a similar computational cost.

Finally, we evaluate the doping dependence of the total energy, in addition to the corresponding first and second derivatives, using VDAT with $\mathcal{N} = 2$ and $\mathcal{N} = 3$. We compare to a previously published DMFT study which used the NRG impurity solver [26] to study $2N_{\text{orb}} = 5$ for $U/t = 6$ and $U/t = 14$ (see Fig. 5), where $U/t = 6$ is a metal at all densities and $U/t = 14$ undergoes an MIT at the integer fillings of $n = 0.2$ and $n = 0.4$. We first compare the total energy, where the variational principle guarantees that the energy monotonically decreases from $\mathcal{N} = 2$ to $\mathcal{N} = 3$ to the numerically exact solution given by DMFT solved within NRG [see Fig. 5(a)]. It should be noted the DMFT study

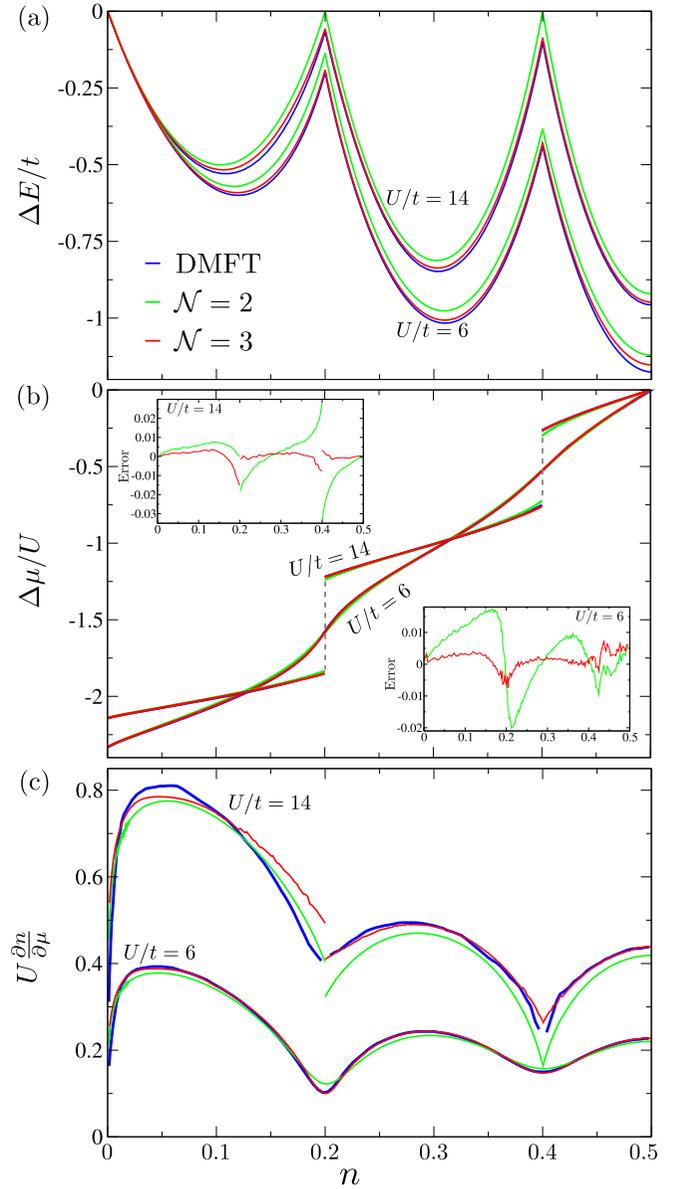


FIG. 5. Doping dependent results of the $SU(2N_{\text{orb}})$ Hubbard model on the $d = \infty$ Bethe lattice at zero temperature from published DMFT calculations solved using NRG [26] and VDAT within the SCDA using $\mathcal{N} = 2$ and $\mathcal{N} = 3$, with $2N_{\text{orb}} = 5$. (a) Total energy difference ΔE vs. density, where $\Delta E = E(t, U) - E(0, U)$. The DMFT curve is obtained by integrating the chemical potential over density. (b) $\Delta\mu/U$ vs. density, where $\Delta\mu = \mu - 2U$. The insets plot the absolute error in $\Delta\mu/U$ vs. density. (c) The derivative $\partial n/\partial\mu$ times U vs. the density.

did not provide the total energy, and we obtained it by numerically integrating the chemical potential over the density. For clarity, we plot $\Delta E/t$ where $\Delta E = E(t, U) - E(0, U)$ and $E(t, U)$ is the total energy evaluated at a given density. Overall, $\mathcal{N} = 3$ yields a dramatic improvement over $\mathcal{N} = 2$, especially at integer fillings. Interestingly, the trends in the absolute error of the total energy are notably different for $\mathcal{N} = 2$ and $\mathcal{N} = 3$, where the former has the largest absolute error at integer filling while the latter has the largest absolute error midway between integer fillings. The trend for $\mathcal{N} = 3$

might be attributed to the efficacy of the kinetic projector for capturing superexchange at integer filling. We proceed to compare the chemical potential, which is the derivative of the energy with respect to the density, as a function of density [see Fig. 5(b)]. Within VDAT, the chemical potential was obtained using finite difference to take the first derivative of the total energy with respect to the density, and a grid spacing of 0.002 was used for $0.02 < n_{\alpha\sigma} \leq 0.5$ while 0.0001 was used for $n_{\alpha\sigma} < 0.02$. For clarity, we plot $\Delta\mu/U$ where $\Delta\mu = \mu - 2U$. The Mott transition can clearly be identified as a discontinuity in the chemical potential at integer fillings. While $\mathcal{N} = 2$ is reasonable overall, there are clear discrepancies near the MIT (see insets for absolute error in $\Delta\mu/U$), and $\mathcal{N} = 3$ largely resolves these issues. However, it should be recalled that unlike the total energy, the convergence of the chemical potential is not necessarily monotonic in \mathcal{N} . For example, in the case of $U/t = 14$ and $n \rightarrow 0.2^-$, the total energy for $\mathcal{N} = 2$ is substantially larger than $\mathcal{N} = 3$, yet the chemical potential for $\mathcal{N} = 2$ is much closer to the exact solution. To further scrutinize these same results from another viewpoint, we examine $U \partial n / \partial \mu$ as a function of density [see Fig. 5(c)]. Within VDAT, $\partial \mu / \partial n$ was obtained using finite difference to take the second derivative of the total energy, and the same grid spacing was used as in the case of the chemical potential. For the low-density region $n_{\alpha\sigma} < 0.1$, the $\mathcal{N} = 3$ results were smoothed using a spline interpolation. Similar to the chemical potential, the convergence of this quantity is not necessarily monotonic in \mathcal{N} , and the same conclusions can be drawn.

IV. CONCLUSIONS

In this paper, we proposed a gauge constrained algorithm to evaluate the SPD ansatz at $\mathcal{N} = 3$ within the SCDA for the multiorbital Hubbard model. The key feature of this algorithm is that it automatically satisfies the self-consistency condition of the SCDA using the gauge freedom of the SPD, greatly facilitating the minimization over the variational parameters. Interestingly, the gauge constrained algorithm yields a simple analytical form of the single-particle density matrix of the optimized SPD ansatz. The convenient mathematical form of the gauge constrained algorithm greatly simplifies the implementation of VDAT within the SCDA when treating a large number of orbitals. In order to showcase the power of the gauge constrained algorithm, we studied the $SU(2N_{\text{orb}})$ Hubbard model at zero temperature on the Bethe lattice in $d = \infty$ for $N_{\text{orb}} = 2-8$, and compare to numerically exact DMFT results where available. While the symmetry of the $SU(2N_{\text{orb}})$ Hubbard model greatly reduces the computational cost for solving the DMFT impurity model, computational limitations still restrict most studies to relatively small values of N_{orb} . A DMFT study using a numerical renormalization group impurity solver presented results up to $2N_{\text{orb}} = 5$ [26], while a study using a quantum Monte Carlo impurity solver presented finite-temperature results up to $N_{\text{orb}} = 8$ [25]; though the latter results were at relatively high temperatures and appear to have nontrivial stochastic error. Therefore, our successful execution of $N_{\text{orb}} = 8$ showcases the utility of the gauge constrained algorithm for executing VDAT within the SCDA at $\mathcal{N} = 3$. At half-filling, we evaluated the kinetic energy, interaction energy, and quasiparticle weight as a function of

U/t . For the doped case, we evaluated the density as a function of the chemical potential, in addition to the derivative, at various values of U/t . As expected, $\mathcal{N} = 3$ yields a dramatic improvement over $\mathcal{N} = 2$, at a similar computational cost. The successful computation of the ground-state energy for the $N_{\text{orb}} = 7$ Hubbard model on a single processor core in under one hour demonstrates the viability of VDAT to study realistic f -electron systems. The technical developments in this work are a key step forward towards studying realistic Hamiltonians of complex strongly correlated electron materials.

ACKNOWLEDGMENTS

We thank Shuxiang Yang for useful discussions about the manuscript. This work was supported by the Columbia Center for Computational Electrochemistry. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

APPENDIX A: SOLVING MULTIORBITAL HUBBARD MODEL WITH A DENSITY CONSTRAINT

It is often desirable to solve a Hamiltonian with fixed densities for the spin orbitals, which can be efficiently executed by reparametrizing the variational parameters u . We begin by realizing that the vector space associated with u can be constructed as a direct product of two-dimensional vector spaces associated with each spin orbital. For an operator in the compound space $\hat{O} = \prod_{\alpha\sigma} \hat{O}_{\alpha\sigma}$, the representation in the u basis can be constructed as

$$(\hat{O})_u = (\hat{O}_{1\uparrow})_{u;1\uparrow} \otimes \cdots \otimes (\hat{O}_{N_{\text{orb}}\downarrow})_{u;N_{\text{orb}}\downarrow}. \quad (\text{A1})$$

Using this relation, Eqs. (32) and (34) are recast as

$$(\hat{1})_u = (\hat{1})_{u;1\uparrow} \otimes \cdots \otimes (\hat{1})_{u;N_{\text{orb}}\downarrow}, \quad (\text{A2})$$

$$\begin{aligned} (\hat{a}_{\alpha\sigma}^{\dagger(\tau)} \hat{a}_{\alpha\sigma}^{(\tau')})_u &= (\hat{1})_{u;1\uparrow} \otimes \cdots \\ &\times \otimes (\hat{a}_{\alpha\sigma}^{\dagger(\tau)} \hat{a}_{\alpha\sigma}^{(\tau')})_{u;\alpha\sigma} \cdots \otimes (\hat{1})_{u;N_{\text{orb}}\downarrow}, \end{aligned} \quad (\text{A3})$$

where

$$[(\hat{1})_{u;\alpha\sigma}]_{\Gamma_{\alpha\sigma} \Gamma'_{\alpha\sigma}} = p_{\alpha\sigma}(\Gamma_{\alpha\sigma}, \Gamma'_{\alpha\sigma}), \quad (\text{A4})$$

$$[(\hat{a}_{\alpha\sigma}^{\dagger(\tau)} \hat{a}_{\alpha\sigma}^{(\tau')})_{u;\alpha\sigma}]_{\Gamma_{\alpha\sigma} \Gamma'_{\alpha\sigma}} = g_{\alpha\sigma}^{\tau\tau'}(\Gamma_{\alpha\sigma}, \Gamma'_{\alpha\sigma}), \quad (\text{A5})$$

where $p_{\alpha\sigma}(\Gamma_{\alpha\sigma}, \Gamma'_{\alpha\sigma})$ and $g_{\alpha\sigma}^{\tau\tau'}(\Gamma_{\alpha\sigma}, \Gamma'_{\alpha\sigma})$ are defined in Eqs. (33) and (35), respectively. The relevant matrices, which will be needed to constrain the orbital density are

$$(\hat{1})_{u;\alpha\sigma} = \begin{pmatrix} \mathcal{G}_{\alpha\sigma,12}^2 + \frac{1}{4} & \frac{1}{4} - \mathcal{G}_{\alpha\sigma,12}^2 \\ \frac{1}{4} - \mathcal{G}_{\alpha\sigma,12}^2 & \mathcal{G}_{\alpha\sigma,12}^2 + \frac{1}{4} \end{pmatrix}, \quad (\text{A6})$$

$$(\hat{a}_{\alpha\sigma}^{\dagger(1)} \hat{a}_{\alpha\sigma}^{(1)})_{u;\alpha\sigma} = \begin{pmatrix} 0 & 0 \\ \frac{1}{4} - \mathcal{G}_{\alpha\sigma,12}^2 & \mathcal{G}_{\alpha\sigma,12}^2 + \frac{1}{4} \end{pmatrix}, \quad (\text{A7})$$

$$(\hat{a}_{\alpha\sigma}^{\dagger(2)} \hat{a}_{\alpha\sigma}^{(2)})_{u;\alpha\sigma} = \begin{pmatrix} 0 & \frac{1}{4} - \mathcal{G}_{\alpha\sigma,12}^2 \\ 0 & \mathcal{G}_{\alpha\sigma,12}^2 + \frac{1}{4} \end{pmatrix}. \quad (\text{A8})$$

In summary, Eq. (A1) provides a simple mathematical structure to construct $(\hat{O})_u$.

We now proceed to reparametrize the variational parameters u . In general, one can introduce a linear transformation over the variational parameters as $u = Vw$, and by requiring $w^T (\hat{O})_w w = u^T (\hat{O})_u u$, a new matrix form is obtained as

$$(\hat{O})_w \equiv V^T (\hat{O})_u V. \quad (\text{A9})$$

In order to preserve the direct product structure of $(\hat{O})_w$, the transformation is constructed as $V = V_{1\uparrow} \otimes \cdots \otimes V_{N_{\text{orb}}\downarrow}$, resulting in

$$(\hat{O})_w = (\hat{O}_{1\uparrow})_{w;1\uparrow} \otimes \cdots \otimes (\hat{O}_{N_{\text{orb}}\downarrow})_{w;N_{\text{orb}}\downarrow}, \quad (\text{A10})$$

$$(\hat{O}_{\alpha\sigma})_{w;\alpha\sigma} \equiv V_{\alpha\sigma}^T (\hat{O}_{\alpha\sigma})_{u;\alpha\sigma} V_{\alpha\sigma}. \quad (\text{A11})$$

In order to ensure that $(\hat{1})_{w;\alpha\sigma}$ is the identity matrix and the symmetric part of $(\hat{a}_{\alpha\sigma}^{\dagger(1)} \hat{a}_{\alpha\sigma}^{(1)})_{w;\alpha\sigma}$ is diagonal, we have

$$V_{\alpha\sigma} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{2\mathcal{G}_{\alpha\sigma,12}} + 1 & 1 - \frac{1}{2\mathcal{G}_{\alpha\sigma,12}} \\ 1 - \frac{1}{2\mathcal{G}_{\alpha\sigma,12}} & \frac{1}{2\mathcal{G}_{\alpha\sigma,12}} + 1 \end{pmatrix}, \quad (\text{A12})$$

thus completely defining the reparametrization. One of the necessary reparametrized matrix elements is

$$\begin{aligned} n_{w;\alpha\sigma} &\equiv \frac{1}{2} \left((\hat{a}_{\alpha\sigma}^{\dagger(1)} \hat{a}_{\alpha\sigma}^{(1)})_{w;\alpha\sigma} + (\hat{a}_{\alpha\sigma}^{\dagger(1)} \hat{a}_{\alpha\sigma}^{(1)})_{w;\alpha\sigma}^T \right) \\ &= \begin{pmatrix} -\frac{(1-2\mathcal{G}_{\alpha\sigma,12})^2}{8\mathcal{G}_{\alpha\sigma,12}} & 0 \\ 0 & \frac{(2\mathcal{G}_{\alpha\sigma,12}+1)^2}{8\mathcal{G}_{\alpha\sigma,12}} \end{pmatrix}, \end{aligned} \quad (\text{A13})$$

and the others are provided in Supplemental Material [24].

We now proceed to constrain the density for each spin orbital, and we begin by considering the case where the interacting projector is a noninteracting density matrix, which can be written as

$$w_{\Gamma;0}^2 = \prod_{\alpha\sigma} ((1 - n_{\alpha\sigma;\text{eff}})(1 - \Gamma_{\alpha\sigma}) + n_{\alpha\sigma;\text{eff}}\Gamma_{\alpha\sigma}), \quad (\text{A14})$$

where $n_{\alpha\sigma;\text{eff}}$ can be determined from

$$\text{Tr} \left(n_{w;\alpha\sigma} \begin{pmatrix} 1 - n_{\alpha\sigma;\text{eff}} & 0 \\ 0 & n_{\alpha\sigma;\text{eff}} \end{pmatrix} \right) = n_{\alpha\sigma}, \quad (\text{A15})$$

which can be solved as

$$n_{\alpha\sigma;\text{eff}} = \frac{2(2n_{\alpha\sigma} - 1)\mathcal{G}_{\alpha\sigma,12}}{4\mathcal{G}_{\alpha\sigma,12}^2 + 1} + \frac{1}{2}. \quad (\text{A16})$$

Subsequently, $2^{2N_{\text{orb}}} - (1 + 2N_{\text{orb}})$ variational parameters x_η can be introduced to describe the deviations from $w_{\Gamma;0}^2$, which do not change the density or the normalization. It is then useful to define a $2^{2N_{\text{orb}}} \times 2^{2N_{\text{orb}}}$ matrix $V_{\Gamma\eta}$ as

$$V_{\Gamma i(\{\alpha_1\sigma_1, \dots, \alpha_n\sigma_n\})} = \prod_{j=1}^n \left(\Gamma_{\alpha_j\sigma_j} - \frac{1}{2} \right), \quad (\text{A17})$$

where $i(\{\alpha_1\sigma_1, \dots, \alpha_n\sigma_n\}) = 1, \dots, 2^{2N_{\text{orb}}}$ is a convention for indexing all subsets of $\{1\uparrow, 1\downarrow, \dots, N_{\text{orb}}\uparrow, N_{\text{orb}}\downarrow\}$, and $n = 0, \dots, 2N_{\text{orb}}$ denotes the number of spin orbitals contained in a given subset. A convenient convention for sorting the subsets is first sorting by increasing cardinality and then by the binary interpretation of the subset. The subsets with cardinality greater than one form $2^{2N_{\text{orb}}} - (1 + 2N_{\text{orb}})$ orthogonal vectors that do not change the normalization or the orbital

occupation. A similar approach has been used to represent the Bernoulli distribution [30]. A general interacting projector that is constrained to the given orbital occupation can be parameterized as

$$w_\Gamma^2 = w_{\Gamma;0}^2 + \sum_{\eta=2+2N_{\text{orb}}}^{2^{2N_{\text{orb}}}} V_{\Gamma\eta} x_\eta, \quad (\text{A18})$$

where x_η are real numbers that are constrained by the condition that $\omega_\Gamma^2 \geq 0$. For example, $N_{\text{orb}} = 1$ results in one independent variational parameter x , yielding

$$w_1^2 = (1 - n_{\uparrow;\text{eff}})(1 - n_{\downarrow;\text{eff}}) + \frac{1}{4}x, \quad (\text{A19})$$

$$w_2^2 = (1 - n_{\uparrow;\text{eff}})n_{\downarrow;\text{eff}} - \frac{1}{4}x, \quad (\text{A20})$$

$$w_3^2 = n_{\uparrow;\text{eff}}(1 - n_{\downarrow;\text{eff}}) - \frac{1}{4}x, \quad (\text{A21})$$

$$w_4^2 = n_{\uparrow;\text{eff}}n_{\downarrow;\text{eff}} + \frac{1}{4}x, \quad (\text{A22})$$

where $x \in [x_{\text{min}}, x_{\text{max}}]$ and

$$x_{\text{min}} = -4 \min((1 - n_{\uparrow;\text{eff}})(1 - n_{\downarrow;\text{eff}}), n_{\uparrow;\text{eff}}n_{\downarrow;\text{eff}}), \quad (\text{A23})$$

$$x_{\text{max}} = 4 \min((1 - n_{\uparrow;\text{eff}})n_{\downarrow;\text{eff}}, n_{\uparrow;\text{eff}}(1 - n_{\downarrow;\text{eff}})). \quad (\text{A24})$$

Finally, the ground-state energy can be obtained as

$$E = \min_{\mathcal{G}_{12}, x, \mathbf{b}} \left(\int dk \epsilon_{k\alpha\sigma} n_{k\alpha\sigma}(\mathbf{a}, \mathbf{b}) + E_{\text{loc}}(\mathcal{G}_{12}, x, \mathbf{b}) \right), \quad (\text{A25})$$

where $x = \{x_\eta\}$ and \mathbf{a} is determined from $\{n_{\alpha\sigma}\}$, \mathcal{G}_{12} , x , and \mathbf{b} .

APPENDIX B: GAUGE CONSTRAINED ALGORITHM USING GENERAL LOCAL PROJECTORS

In this paper, we have assumed that the interacting projector can be written as a linear combination of diagonal Hubbard operators in the basis $\alpha\sigma$, and that \mathcal{G} is diagonal in basis $\alpha\sigma$. Here we outline how to treat the general case, starting with the first assumption. A general local interacting projector can be an arbitrary linear combination of all possible Hubbard operators, including off-diagonal Hubbard operators. A general Hubbard operator can be constructed as

$$\begin{aligned} \hat{P}_{i\Gamma} &= \prod_{\alpha\sigma} (\delta_{\Gamma_{\alpha\sigma}, 0}(1 - \hat{n}_{i\alpha\sigma}) + \delta_{\Gamma_{\alpha\sigma}, 1}\hat{n}_{i\alpha\sigma} \\ &\quad + \delta_{\Gamma_{\alpha\sigma}, 2}\hat{a}_{i\alpha\sigma}^\dagger + \delta_{\Gamma_{\alpha\sigma}, 3}\hat{a}_{i\alpha\sigma}), \end{aligned} \quad (\text{B1})$$

$$= \prod_{\alpha\sigma} \hat{P}_{i\alpha\sigma; \Gamma_{\alpha\sigma}}, \quad (\text{B2})$$

where $\Gamma - 1 = (\Gamma_{1\uparrow} \dots \Gamma_{N_{\text{orb}}\downarrow})_4$ and $\Gamma = 1, \dots, 4^{2N_{\text{orb}}}$. The most general interacting projector can be constructed as $\hat{P}_i(u) = \sum_\Gamma u_\Gamma \hat{P}_{i\Gamma}$. However, given that we require $\hat{P}_i(u)$ to obey certain symmetries and conservation relations, some u_Γ may be zero. In order to evaluate $\langle \hat{P}_i^{(1)} \hat{P}_i^{(2)} \hat{O} \rangle_{\hat{\rho}_{\text{loc};i,0}}$, we first consider

$$\hat{O} = \prod_{\alpha\sigma} \hat{O}_{\alpha\sigma}, \quad (\text{B3})$$

where $\hat{\mathcal{O}}_{\alpha\sigma}$ is a single product in terms of $\hat{a}_{i\alpha\sigma}^{\dagger(\tau)}$ and $\hat{a}_{i\alpha\sigma}^{(\tau)}$, yielding

$$\left\langle \hat{P}_i^{(1)} \hat{P}_i^{(2)} \hat{\mathcal{O}} \right\rangle_{\hat{P}_{\text{loc};i,0}} = \sum_{\Gamma\Gamma'} u_{\Gamma} u_{\Gamma'} \left\langle \hat{P}_{i\Gamma}^{(1)} \hat{P}_{i\Gamma'}^{(2)} \hat{\mathcal{O}} \right\rangle_{\hat{P}_{\text{loc};i,0}}, \quad (\text{B4})$$

where

$$\begin{aligned} \left\langle \hat{P}_{i\Gamma}^{(1)} \hat{P}_{i\Gamma'}^{(2)} \hat{\mathcal{O}} \right\rangle_{\hat{P}_{\text{loc};i,0}} &= \left\langle \prod_{\alpha\sigma} \hat{P}_{i\alpha\sigma;\Gamma}^{(1)} \prod_{\alpha\sigma} \hat{P}_{i\alpha\sigma;\Gamma'}^{(2)} \prod_{\alpha\sigma} \hat{\mathcal{O}}_{\alpha\sigma} \right\rangle_{\hat{P}_{\text{loc};i,0}} \quad (\text{B5}) \\ &= \theta(\hat{\mathcal{O}}, \Gamma, \Gamma') \prod_{\alpha\sigma} \left\langle \hat{P}_{i\alpha\sigma;\Gamma}^{(1)} \hat{P}_{i\alpha\sigma;\Gamma'}^{(2)} \hat{\mathcal{O}}_{\alpha\sigma} \right\rangle_{\hat{P}_{\text{loc};i,0}}, \quad (\text{B6}) \end{aligned}$$

where $\theta(\hat{\mathcal{O}}, \Gamma, \Gamma') = \pm 1$ and is determined by tracking the sign when ordering the expression from Eq. (B5) to Eq. (B6). For a general operator $\hat{\mathcal{O}}$, one can always decompose it into a sum of operators, which has the form of Eq. (B3) and apply the above formulas.

In order to treat a general \mathcal{G} and a general operator $\hat{\mathcal{O}}$, one must straightforwardly apply Wick's theorem to evaluate expectation values [1], though the resulting gauge constrained algorithm will be more complicated. For example, simple closed form equations such as Eq. (22) may not be obtained, requiring a numerical minimization to obtain the density distribution.

APPENDIX C: UNDERSTANDING HOW THE $\mathcal{N} = 3$ GAUGE CONSTRAINED ALGORITHM WITH A RESTRICTED KINETIC PROJECTOR REDUCES TO $\mathcal{N} = 2$

In Ref. [1], we illustrated how the SCDA at $\mathcal{N} = 2$ using the Gutzwiller gauge recovers the Gutzwiller approximation. In the present work where we address $\mathcal{N} = 3$, the gauge constrained algorithm uses a different type of gauge. Therefore, it is interesting to see how the $\mathcal{N} = 3$ gauge constrained algorithm with a restricted kinetic projector can recover the Gutzwiller approximation. In particular, the restricted kinetic projector will force the density distribution to be flat both above and below the fermi surface. We begin by assuming

$$\mathcal{G}_{\alpha\sigma} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad (\text{C1})$$

which is motivated by the Gutzwiller gauge. The canonical discrete action of $\mathcal{G}_{\alpha\sigma}$ corresponds to an SPD [1], which is the product of three identity operators, and we can write the A block for interacting Green's function as

$$\mathbf{g}_{\text{loc},\alpha\sigma;A} = \begin{pmatrix} n_{\alpha\sigma} & a_{\alpha\sigma} r_{\alpha\sigma} \\ -a_{\alpha\sigma} r_{\alpha\sigma} & n_{\alpha\sigma} \end{pmatrix}, \quad (\text{C2})$$

where $a_{\alpha\sigma} = \sqrt{(1 - n_{\alpha\sigma})n_{\alpha\sigma}}$ and $r_{\alpha\sigma}$ denotes the renormalization for the off-diagonal elements of the A block compared to the reference interacting Green's function

$$\mathbf{g}_{\text{loc},\alpha\sigma;A;\text{ref}} = \begin{pmatrix} n_{\alpha\sigma} & a_{\alpha\sigma} \\ -a_{\alpha\sigma} & n_{\alpha\sigma} \end{pmatrix}, \quad (\text{C3})$$

which corresponds to the case where \hat{P} is noninteracting, denoted as \hat{P}_0 . The \hat{P}_0 is chosen such that

$$\frac{\text{Tr}(\hat{P}_0^2 \hat{n}_{\alpha\sigma})}{\text{Tr}(\hat{P}_0^2)} = \frac{\text{Tr}(\hat{P}^2 \hat{n}_{\alpha\sigma})}{\text{Tr}(\hat{P}^2)} = n_{\alpha\sigma}, \quad (\text{C4})$$

and $r_{\alpha\sigma}$ is given as

$$r_{\alpha\sigma} = \frac{\text{Tr}(\hat{P} \hat{a}_{\alpha\sigma}^{\dagger} \hat{P} \hat{a}_{\alpha\sigma})}{\text{Tr}(\hat{P}^2)} \bigg/ \frac{\text{Tr}(\hat{P}_0 \hat{a}_{\alpha\sigma}^{\dagger} \hat{P}_0 \hat{a}_{\alpha\sigma})}{\text{Tr}(\hat{P}_0^2)}, \quad (\text{C5})$$

The point of introducing the reference \hat{P}_0 is to allow comparison with the Gutzwiller approximation. Given that the local energy can be computed as

$$E_{\text{loc}} = \frac{\text{Tr}(\hat{P}^2 \hat{H}_{\text{loc}})}{\text{Tr}(\hat{P}^2)}, \quad (\text{C6})$$

which is the same as in the Gutzwiller approximation, the remaining task is to confirm that the kinetic energy recovers the Gutzwiller approximation when restricting the density distribution to be flat, and confirm that the self-consistency of the SCDA is maintained.

We begin by computing the A block of the local integer time self-energy as

$$\mathbf{S}_{\text{loc},\alpha\sigma;A} = (\mathcal{G}_{\alpha\sigma;A}^{-1} - \mathbf{1})^{-1} (\mathbf{g}_{\text{loc},\alpha\sigma;A}^{-1} - \mathbf{1}), \quad (\text{C7})$$

which yields the integer time self-energy as

$$\mathbf{S}_{\text{loc},\alpha\sigma} = \begin{pmatrix} S_{\alpha\sigma;11} & S_{\alpha\sigma;12} & 0 \\ -S_{\alpha\sigma;12} & S_{\alpha\sigma;11} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where

$$\begin{aligned} S_{\alpha\sigma;11} &= \frac{a_{\alpha\sigma} r_{\alpha\sigma}}{n_{\alpha\sigma} (n_{\alpha\sigma} - (n_{\alpha\sigma} - 1) r_{\alpha\sigma}^2)}, \\ S_{\alpha\sigma;12} &= \frac{1}{n_{\alpha\sigma} - (n_{\alpha\sigma} - 1) r_{\alpha\sigma}^2} - 1. \end{aligned}$$

Assuming the kinetic projector is the identity, corresponding to $\lambda_{k\alpha\sigma,1} = 1/2$, we get a flat distribution in both the $<$ and $>$ region, yielding

$$n_{k\alpha\sigma} |_{k \in <} = n_{\alpha\sigma} (1 - r_{\alpha\sigma}^2) + r_{\alpha\sigma}^2, \quad (\text{C8})$$

$$n_{k\alpha\sigma} |_{k \in >} = n_{\alpha\sigma} (1 - r_{\alpha\sigma}^2). \quad (\text{C9})$$

One can verify that the integral of the density distribution yields the corresponding local density as

$$\int_{<} n_{k\alpha\sigma} dk + \int_{>} n_{k\alpha\sigma} dk = n_{\alpha\sigma}, \quad (\text{C10})$$

and the real-space renormalization of the hopping parameter, $Z_{\alpha\sigma}$, can be computed as

$$Z_{\alpha\sigma} = n_{k\alpha\sigma} |_{k \in <} - n_{k\alpha\sigma} |_{k \in >} = r_{\alpha\sigma}^2, \quad (\text{C11})$$

which recovers the Gutzwiller approximation.

We now confirm that the self-consistency condition of the SCDA is maintained. Computing the local integer time Green's function yields

$$\mathbf{g}'_{\text{loc}} = \begin{pmatrix} n_{\alpha\sigma} & a_{\alpha\sigma} r_{\alpha\sigma} & a_{\alpha\sigma} r_{\alpha\sigma} \\ -a_{\alpha\sigma} r_{\alpha\sigma} & n_{\alpha\sigma} & n_{\alpha\sigma} \\ -a_{\alpha\sigma} r_{\alpha\sigma} & n_{\alpha\sigma} - 1 & n_{\alpha\sigma} \end{pmatrix}, \quad (\text{C12})$$

which yields

$$\mathbf{g}' = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad (\text{C13})$$

which indicates that the SCDA self-consistency is satisfied.

- [1] Z. Q. Cheng and C. A. Marianetti, Foundations of variational discrete action theory, *Phys. Rev. B* **103**, 195138 (2021).
- [2] Z. Q. Cheng and C. A. Marianetti, Variational Discrete Action Theory, *Phys. Rev. Lett.* **126**, 206402 (2021).
- [3] Z. Cheng and C. A. Marianetti, Precise ground state of multi-orbital Mott systems via the variational discrete action theory, *Phys. Rev. B* **106**, 205129 (2022).
- [4] M. Imada, A. Fujimori, and Y. Tokura, Metal-insulator transitions, *Rev. Mod. Phys.* **70**, 1039 (1998).
- [5] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, Electronic structure calculations with dynamical mean-field theory, *Rev. Mod. Phys.* **78**, 865 (2006).
- [6] M. C. Gutzwiller, Effect of Correlation on the Ferromagnetism of Transition Metals, *Phys. Rev. Lett.* **10**, 159 (1963).
- [7] M. C. Gutzwiller, Effect of correlation on ferromagnetism of transition metals, *Phys. Rev.* **134**, A923 (1964).
- [8] M. C. Gutzwiller, Correlation of electrons in a narrow s band, *Phys. Rev.* **137**, A1726 (1965).
- [9] H. Otsuka, Variational Monte-Carlo studies of the Hubbard-model in one-dimension and 2-dimension - off-diagonal intersite correlation-effects, *J. Phys. Soc. Jpn.* **61**, 1645 (1992).
- [10] M. Dzierzawa, D. Baeriswyl, and M. Distasio, Variational wave-functions for the Mott transition - the $1/r$ -Hubbard chain, *Phys. Rev. B* **51**, 1993 (1995).
- [11] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions, *Rev. Mod. Phys.* **68**, 13 (1996).
- [12] D. Vollhardt, Dynamical mean-field theory for correlated electrons, *Ann. Phys. (NY)* **524**, 1 (2012).
- [13] G. Kotliar and A. E. Ruckenstein, New Functional Integral Approach to Strongly Correlated Fermi Systems: The Gutzwiller Approximation as a Saddle Point, *Phys. Rev. Lett.* **57**, 1362 (1986).
- [14] J. Bunemann and F. Gebhard, Equivalence of Gutzwiller and slave-boson mean-field theories for multiband Hubbard models, *Phys. Rev. B* **76**, 193104 (2007).
- [15] C. Piefke and F. Lechermann, Rigorous symmetry adaptation of multiorbital rotationally invariant slave-boson theory with application to Hund's rules physics, *Phys. Rev. B* **97**, 125154 (2018).
- [16] C. He, Z. Ren, B. Song, E. Zhao, J. Lee, Y. C. Zhang, S. Zhang, and G. B. Jo, Collective excitations in two-dimensional $SU(N)$ Fermi gases with tunable spin, *Phys. Rev. Res.* **2**, 012028(R) (2020).
- [17] H. Ozawa, S. Taie, Y. Takasu, and Y. Takahashi, Antiferromagnetic Spin Correlation of $SU(N)$ Fermi Gas in an Optical Superlattice, *Phys. Rev. Lett.* **121**, 225303 (2018).
- [18] F. Schafer, T. Fukuhara, S. Sugawa, Y. Takasu, and Y. Takahashi, Tools for quantum simulation with ultracold atoms in optical lattices, *Nature Rev. Phys.* **2**, 411 (2020).
- [19] L. Sonderhouse, C. Sanner, R. B. Hutson, A. Goban, T. Bilitewski, L. F. Yan, W. R. Milner, A. M. Rey, and J. Ye, Thermodynamics of a deeply degenerate $SU(N)$ -symmetric Fermi gas, *Nature Phys.* **16**, 1216 (2020).
- [20] S. Taie, R. Yamazaki, S. Sugawa, and Y. Takahashi, An $SU(6)$ Mott insulator of an atomic Fermi gas realized by large-spin Pomeranchuk cooling, *Nature Phys.* **8**, 825 (2012).
- [21] R. Zhang, Y. T. Cheng, P. Zhang, and H. Zhai, Controlling the interaction of ultracold alkaline-earth atoms, *Nature Rev. Phys.* **2**, 213 (2020).
- [22] J. M. Luttinger, Fermi surface and some simple equilibrium properties of a system of interacting fermions, *Phys. Rev.* **119**, 1153 (1960).
- [23] J. A. Nelder and R. Mead, A simplex-method for function minimization, *Comput. J.* **7**, 308 (1965).
- [24] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevB.108.035127> for detailed formulas pertaining to integer time Green's functions.
- [25] N. Blumer and E. V. Gorelik, Mott transitions in the half-filled $SU(2M)$ symmetric Hubbard model, *Phys. Rev. B* **87**, 085115 (2013).
- [26] S. Lee, J. Von delft, and A. Weichselbaum, Filling-driven Mott transition in $SU(N)$ Hubbard models, *Phys. Rev. B* **97**, 165143 (2018).
- [27] Z. Cheng, VDATN3multi.jl, 3 2023, <https://github.com/chengzhengqian/VDATN3multi.jl>.
- [28] S. Florens, A. Georges, G. Kotliar, and O. Parcollet, Mott transition at large orbital degeneracy: Dynamical mean-field theory, *Phys. Rev. B* **66**, 205102 (2002).
- [29] J. P. Lu, Metal-insulator transitions in degenerate Hubbard models and A_3C_{60} , *Phys. Rev. B* **49**, 5687 (1994).
- [30] J. L. Teugels, Some representations of the multivariate Bernoulli and binomial distributions, *J. Multivariate Anal.* **32**, 256 (1990).