5 Variational Wave Functions for Strongly Correlated Fermionic Systems

Federico Becca
Dipartimento di Fisica, Università di Trieste
Strada Costiera 11, I-34151 Trieste, Italy

Contents

E. Pavarini, E. Koch, and S. Zhang (eds.)
Many-Body Methods for Real Materials
Modeling and Simulation Vol. 9
Forschungszentrum Jülich, 2019, ISBN 978-3-95806-400-3
http://www.cond-mat.de/events/correl19

1 Introduction

Ordinary metals are characterized by wide electronic bands and large screening effects, thus implying that electron-electron interactions may be considered perturbatively or even neglected. The situation changes completely in transition-metal oxides, where the bands close to the Fermi energy are relatively narrow and the electronic interactions play a predominant role in determining low-energy properties. Indeed, while the kinetic energy favors electron delocalization, the Coulomb repulsion drives the system towards localization, whose ultimate effect is the stabilization of an insulator. This state, which is established by the strong electron-electron correlation, goes under the name of *Mott insulator* [1]. In addition, materials characterized by strong correlations possess unusual properties in the metallic phase, as well as the possibility to show unconventional superconductivity [2, 3]. The lack of a consistent microscopic description of these phenomena clearly implies that a better understanding of correlation effects is needed. Since the early pioneering works on transition-metal oxides, the theoretical approach to Mott insulators has focused on the Hubbard model, which has been independently conceived by Hubbard [4], Gutzwiller [5], and Kanamori [6]. Here, electrons on a lattice interact among each others through a simplified "Coulomb" potential that includes only the on-site term

$$\mathcal{H} = \sum_{i,j,\sigma} t_{i,j} c^{\dagger}_{i,\sigma} c_{j,\sigma} + \text{h.c.} + U \sum_{i} n_{i,\uparrow} n_{i,\downarrow} \,, \tag{1}$$

where $t_{i,j}$ is the hopping amplitude in a d-dimensional lattice with L sites (in the simplest case, $t_{i,j}$ is non-zero only for nearest-neighbor sites and is denoted by t) and U is the local electron-electron repulsion; $c^{\dagger}_{j,\sigma}$ ($c_{j,\sigma}$) creates (destroys) one electron with spin σ on a Wannier orbital residing on the site j

$$\Xi_j(\mathbf{r}) = \frac{1}{\sqrt{L}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}_j} \, \Psi_k(\mathbf{r}) \,, \tag{2}$$

where $\Psi_k(\mathbf{r})$ are Bloch states constructed with the orbitals $\phi(\mathbf{r}-\mathbf{R}_i)$ centered around each site i. The operators at different sites create orthogonal states, thus satisfying the anti-commutation relations

$$\left\{ c_{i,\sigma}, c^{\dagger}_{j,\tau} \right\} = \delta_{i,j} \, \delta_{\sigma,\tau}, \tag{3}$$

$$\left\{ c^{\dagger}_{i,\sigma}, c^{\dagger}_{j,\tau} \right\} = 0. \tag{4}$$

The Hubbard model is defined in the Hilbert space where each site can be empty, singly occupied (with either spin up or down), or doubly occupied. Moreover, the Hamiltonian (1) commutes with the total number of particles with up or down spin (i.e., $N_{\uparrow}$ and $N_{\downarrow}$, $N_e = N_{\uparrow} + N_{\downarrow}$ being the total number of electrons), thus allowing us to consider sectors with different number of particles separately. Although very simple in its formulation, the Hubbard model is generally not solvable with the available analytical techniques, apart from the one-dimensional case [7]. Therefore, several approximate schemes have been introduced, with the support of numerical calculations.

Within the standard approaches, based upon the independent-electron approximation, it is not possible to obtain a metal-insulator transition when the band is half-filled (with one electron per site on average, i.e., $N_e = L$), unless some kind of magnetic order is imposed. As a consequence, these techniques turn the Mott insulator into a conventional band insulator, thus missing the essence of the Mott phenomenon, where a charge gap appears independently of spin order. Dynamical mean-field theory [8] offered an alternative route to this problem, giving a description of the Mott transition without the need for a symmetry breaking. However, this scheme fully neglects spatial correlations and becomes exact only in the limit of infinite dimensionality. Since charge fluctuations are very strong in low-dimensional systems and are expected to determine their physical properties, an alternative method, which allows one to take into account the role of charge fluctuations, would be very useful.

Here, we consider variational wave functions, which can be treated within Monte Carlo techniques, as a possible route to capture the low-energy properties of strongly-correlated systems. In particular, our approach is based on an approximate form for the ground-state wave function that contains the physically relevant terms for the correct description of the Mott insulating state, and, at the same time, is simple enough to allow a straightforward calculation of the physical quantities. In this way, we obtain a transparent and physically intuitive way to understand the correlation-induced localization of electrons.

In the limit of $U/t \to \infty$ (i.e., for Heisenberg or t-J models), the general form for correlated wave functions corresponds to *fully-projected* Slater determinants [9, 10], where the configurations having a finite number of double occupancies are completely removed. At half-filling, these wave functions are obviously insulating, since no charge fluctuations can occur. Within the Hubbard model, early calculations showed that the variational description of a Mott insulator is a non-trivial problem, whenever charge fluctuations are present. Indeed, the Gutzwiller on-site correlation factor [5], which is the natural extension of the full projector in the case of finite (on-site) interaction, gives an insulating state only in the limit of infinite repulsion (apart from infinite dimension), while for finite Coulomb interaction it always corresponds to a correlated metallic state. The reason for its failure has been widely discussed in the past, and an intuitive argument has been found in the lack of correlation among the charge carriers, which correspond to the empty sites (holons) and doubly occupied sites (doublons), created by charge fluctuations at finite interactions [11]. In fact, holons possess an effective positive charge, since one electron is missing, and doublons are negatively charged objects, having one more electron with respect to the average occupation number; if the system is perturbed with the insertion of an electric field, holons and doublons can move freely in opposite directions, thus leading to a metallic behavior. Variational attempts done by adding a short-range correlation term up to a distance ξ among holons and doublons, turned out to be likewise unsuccessful [11,12]. Naively, this happens because the configurations where holons and doublons are at distances larger than ξ are not subject to any correlation term, hence they can move freely on the lattice and conduct. Following this insight, it turns out that, in order to describe a correlated insulator without breaking any symmetry, it is necessary to correlate particles over all length scales. Let us consider a more general argument in view of the above considerations. For realistic Hamiltonians,

the dynamical properties of a system reflect the long-distance behavior of the static correlation functions of its ground state. Within a variational approach, this implies that a good *Ansatz* for an insulating state requires the correct description of the density-density correlations at large distances or, equivalently, the correct behavior of the charge structure factor at small momenta. For fermionic systems, the standard form for a correlated wave function is constituted by a correlation term acting on a Slater determinant, the latter corresponding to an uncorrelated metallic state. As a consequence, a variational wave function built with a short-range correlation factor cannot change the metallic character of the determinant, unless one fully suppresses charge fluctuations, since the large distance physics remains untouched.
The above arguments suggest that a long-range correlation factor is needed in order to correctly describe the insulating state. In particular, since we are interested in the density-density correlations, a natural choice of the correlation factor contains two-body terms, which corresponds to the so-called Jastrow factor [13]. It has been widely used in the context of liquid Helium on the continuum, where it gives the correct behavior of the density structure factor [14, 15]. Here, the analytic form of the Jastrow parameters can be deduced from weak-coupling calculations. Unfortunately, for the purpose of describing an insulating state, a proper analytic form of the Jastrow parameters cannot be obtained by weak-coupling techniques. The lack of a functional form for the Jastrow term, together with the large number of variational parameters required for a long-range correlation factor, represented the main obstacle to the use of this wave function in presence of strong correlation. Nowadays, this difficulty has been overcome with the help of advanced stochastic optimization techniques, which allow one to optimize many variational parameters independently, without assuming any functional form [16, 17].

2 The variational principle

In this section, we discuss the basic aspects of the variational principle, which represents one important pillar when searching for reliable approximations of strongly-correlated systems. Given any approximate state $|\Psi\rangle$ for the exact ground state $|\Upsilon_0\rangle$ of a given Hamiltonian, we can define its variational energy as

$$E = \frac{\langle\Psi|\mathcal{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle}\,. \tag{5}$$

Since any state in the Hilbert space can be expanded in terms of the eigenfunctions $|\Upsilon_i\rangle$ of the Hamiltonian (with energies E_i), the variational state can be written as

$$|\Psi\rangle = \sum_i a_i|\Upsilon_i\rangle, \tag{6}$$

with $a_i = \langle\Upsilon_i|\Psi\rangle$. The normalization condition reads as

$$\langle\Psi|\Psi\rangle = \sum_i |a_i|^2 = 1. \tag{7}$$

By using the expansion of Eq. (6), we easily obtain that

$$\epsilon \equiv E - E_0 = \sum_{i\neq 0} |a_i|^2 (E_i - E_0) \geq 0, \tag{8}$$

which implies that any trial state $|\Psi\rangle$ provides an upper bound of the exact energy, thus giving a controlled way to approximate the original problem. Then, all computational efforts are devoted to minimizing the variational energy E.

Let us now analyze in what sense an approximate wave function, with given "distance" in energy ϵ from the exact ground state, can be considered as a good approximation of the many-body ground state $|\Upsilon_0\rangle$. A crucial role is played by the gap to the first excited state, which is always finite in a system with N_e particles (apart from accidental degeneracies), namely $\Delta = E_1 - E_0 > 0$. From Eq. (8) and the fact that, for $i \neq 0$, $E_i - E_0 \geq \Delta$, it follows that

$$\epsilon \geq \Delta \sum_{i \neq 0} |a_i|^2 \, ; \tag{9}$$

then, by using the normalization condition (7), we obtain

$$\eta = 1 - |a_0|^2 \leq \frac{\epsilon}{\Delta} \, . \tag{10}$$

This relation tells us that, in order to have an accurate approximation of the exact ground state (i.e., $\eta \ll 1$), a sufficient condition is that the error ϵ in the variational energy has to be much smaller than the gap Δ to the first excited state.

The accuracy of generic correlation functions (i.e., expectation values of Hermitian operators, which do not commute with the Hamiltonian, over $|\Psi\rangle$) is usually worse than the one on the ground-state energy. In fact, let us consider a generic operator $\mathcal{O}$ and express the variational wave function as

$$|\Psi\rangle = a_0 |\Upsilon_0\rangle + \sqrt{\eta} |\Upsilon'\rangle, \tag{11}$$

where $|\Upsilon'\rangle$ is orthogonal to the ground state $|\Upsilon_0\rangle$. Then, the difference between the expectation value calculated with the variational state and the exact one is given by

$$\left| \langle \Psi | \mathcal{O} | \Psi \rangle - \langle \Upsilon_0 | \mathcal{O} | \Upsilon_0 \rangle \right| = \left| 2 a_0 \sqrt{\eta} \langle \Upsilon_0 | \mathcal{O} | \Upsilon' \rangle + \eta \langle \Upsilon' | \mathcal{O} | \Upsilon' \rangle - \eta \langle \Upsilon_0 | \mathcal{O} | \Upsilon_0 \rangle \right|, \tag{12}$$

where, for simplicity, we have assumed real wave functions. Then, whenever the variational state is close to the exact ground state, $\eta \ll \sqrt{\eta}$ and we can neglect all the terms that are proportional to η

$$\left| \langle \Psi | \mathcal{O} | \Psi \rangle - \langle \Upsilon_0 | \mathcal{O} | \Upsilon_0 \rangle \right| \approx 2 \sqrt{\eta} \left| \langle \Upsilon_0 | \mathcal{O} | \Upsilon' \rangle \right|, \tag{13}$$

which shows that the accuracy in correlation functions is more problematic than the one on the ground-state energy, with a term proportional to $\sqrt{\eta}$.

3 The variational Monte Carlo method

3.1 General principles

Let us start by describing the general framework in which variational Monte Carlo methods are defined. First of all, we fix a complete basis set $\{|x\rangle\}$ in the Hilbert space, in which (for simplicity) the states are taken to be orthogonal and normalized such that

$$\sum_x |x\rangle\langle x| = \mathbb{I}. \tag{14}$$

Then, any quantum state $|\Psi\rangle$ can be written as

$$|\Psi\rangle = \sum_x |x\rangle\langle x|\Psi\rangle = \sum_x \Psi(x)|x\rangle. \tag{15}$$

In turn, the expectation value of an operator $\mathcal{O}$ over a given variational wave function $|\Psi\rangle$ takes the following form

$$\langle \mathcal{O} \rangle = \frac{\langle \Psi|\mathcal{O}|\Psi\rangle}{\langle \Psi|\Psi\rangle} = \frac{\sum_x \langle \Psi|x\rangle\langle x|\mathcal{O}|\Psi\rangle}{\sum_x \langle \Psi|x\rangle\langle x|\Psi\rangle}\,. \tag{16}$$

The main problem in evaluating the expectation value is that the number of configurations in the sum is exponentially large with the number of particles. Therefore, for large systems, it is impossible to perform an exact enumeration of the configurations to compute $\langle \mathcal{O} \rangle$ exactly. Nevertheless, Eq. (16) can be recast into a form that can be easily treated by standard Monte Carlo methods. Indeed, we have that

$$\langle \mathcal{O} \rangle = \frac{\sum_x |\Psi(x)|^2 \mathcal{O}_L(x)}{\sum_x |\Psi(x)|^2}\,, \tag{17}$$

where we have defined the *local estimator* of the operator $\mathcal{O}$

$$\mathcal{O}_L(x) = \frac{\langle x|\mathcal{O}|\Psi\rangle}{\langle x|\Psi\rangle}\,. \tag{18}$$

The important point is that

$$\mathcal{P}(x) = \frac{|\Psi(x)|^2}{\sum_x |\Psi(x)|^2} \tag{19}$$

can be interpreted as a probability, since it is a non-negative quantity for all configurations $|x\rangle$ and is normalized, i.e., $\sum_x \mathcal{P}(x) = 1$. Therefore, the problem of computing a quantum average of the operator $\mathcal{O}$ can be rephrased into the calculation of the average of the random variable $\mathcal{O}_L(x)$ over the distribution probability $\mathcal{P}(x)$

$$\langle \mathcal{O} \rangle = \sum_x \mathcal{P}(x)\mathcal{O}_L(x). \tag{20}$$

In particular, if we consider the expectation value of the Hamiltonian, the local estimator corresponds to the so-called *local energy*, which is defined by

$$e_L(x) = \frac{\langle x|\mathcal{H}|\Psi\rangle}{\langle x|\Psi\rangle}\,. \tag{21}$$

An important feature of the variational Monte Carlo approach is the *zero-variance property*. Let us suppose that the variational state $|\Psi\rangle$ coincides with an exact eigenstate of $\mathcal{H}$ (not necessarily the ground state), namely $\mathcal{H}|\Psi\rangle = E|\Psi\rangle$. Then, it follows that the local energy $e_L(x)$ is constant

$$e_L(x) = \frac{\langle x|\mathcal{H}|\Psi\rangle}{\langle x|\Psi\rangle} = E\frac{\langle x|\Psi\rangle}{\langle x|\Psi\rangle} = E. \tag{22}$$

Therefore, the random variable $e_L(x)$ does not depend on $|x\rangle$, which immediately implies that its variance is zero, while its mean value E coincides with the exact eigenvalue (in other words, $e_L(x)$ is not a random variable). Clearly, this is an extreme case that is very rare for generic correlated models. However, in general, the variance of $e_L(x)$ will decrease its value whenever the variational state $|\Psi\rangle$ will approach an exact eigenstate. This fact is very important to reduce the statistical fluctuations and improve the numerical efficiency. The zero-variance property is a feature that exists only for quantum expectation values, while it is absent in classical calculations, where observables have thermal fluctuations.

3.2 Markov chains

Instead of an exact enumeration of all configurations $\{|x\rangle\}$ in the Hilbert space, the quantum average of the operator $\mathcal{O}$ is evaluated by sampling a set of configurations $\{|x_n\rangle\}$ that are distributed according to the probability $\mathcal{P}(x)$, such that

$$\langle \mathcal{O} \rangle \approx \frac{1}{N} \sum_{n=1}^{N} \mathcal{O}_L(x_n). \tag{23}$$

From now on, we denote configurations by using only x, dropping the ket notation. The idea to sample a generic probability distribution is to construct a non-deterministic, i.e., random, process for which a configuration x_n evolves as a function of a discrete iteration time n according to a stochastic dynamics. A particularly simple case is given by the so-called Markov chains, where the configuration at time $n+1$ just depends upon the one at time n

$$x_{n+1} = F(x_n, \xi_n), \tag{24}$$

where the function F is taken to be time independent. The stochastic nature of the dynamics (24) is due to the fact that F depends upon a random variable ξ_n that is distributed according to a probability density $\chi(\xi_n)$. Here, the main point is to define a suitable function F such that the configurations x_n will be distributed (for large enough time n) according to the probability that we want to sample. Notice that, although ξ_n and ξ_{n+1} are independent random variables, $x_n \equiv x$ and $x_{n+1} \equiv x'$ are not independent. The joint probability distribution of these variables can be decomposed into the product of the marginal and the conditional probability

$$\mathcal{P}_{\text{joint},n}(x', x) = \omega(x'|x)\, \mathcal{P}_n(x), \tag{25}$$

where the conditional probability is such that $\omega(x'|x) \geq 0$ for all x and x' and satisfies the following normalization

$$\sum_{x'} \omega(x'|x) = 1. \tag{26}$$

It represents the probability that, having the configuration x at the iteration n, x' appears at $n+1$; its actual form depends upon the function $F(x, \xi)$ and the probability distribution $\chi(\xi)$.

We are now in the position of deriving the so-called Master equation, associated to the Markov chain. Indeed, the marginal probability of the variable x' is given by

$$\mathcal{P}_{n+1}(x') = \sum_x \mathcal{P}_{\text{joint},n}(x', x), \tag{27}$$

so that, by using Eq. (25), we get

$$\mathcal{P}_{n+1}(x') = \sum_x \omega(x'|x)\, \mathcal{P}_n(x). \tag{28}$$

This equation allows us to calculate the evolution of the marginal probability $\mathcal{P}_n(x)$ as a function of n, since the conditional probability $\omega(x'|x)$ is determined by the stochastic dynamics in Eq. (24) and does not depend upon n.
The important question about the Markov process is to understand under which conditions the sequence of distributions $\mathcal{P}_n(x)$ converges to some limiting (i.e., equilibrium) distribution $\mathcal{P}(x)$ or not. In particular

1. Does a stationary distribution $\mathcal{P}(x)$ exist?

2. Is the convergence to $\mathcal{P}(x)$ guaranteed when starting from a given *arbitrary* $\mathcal{P}_0(x)$?

The first question requires that

$$\mathcal{P}(x') = \sum_x \omega(x'|x)\, \mathcal{P}(x). \tag{29}$$

In order to satisfy this condition, it is sufficient (but not necessary) to satisfy the so-called *detailed balance* condition

$$\omega(x'|x)\, \mathcal{P}(x) = \omega(x|x')\, \mathcal{P}(x'). \tag{30}$$

The second question requires the *ergodicity* condition, i.e., the possibility to reach any state x from any other one x' by performing a finite number of steps.

3.3 The Metropolis-Hastings algorithm

Finally, we present a practical way of constructing a conditional probability $\omega(x'|x)$ that satisfies the detailed balance condition (30), such that, for large values of n, the configurations x_n are distributed according to a given probability distribution $\mathcal{P}(x)$. Metropolis and collaborators [18] introduced a very simple scheme, which is also very general and can be applied to many different cases. Later, the so-called Metropolis algorithm has been extended to more general cases by Hastings [19] (very often, the name of "Metropolis-Hastings algorithm" is also used). As a first step, we split the transition probability $\omega(x'|x)$ into two pieces

$$\omega(x'|x) = T(x'|x)A(x'|x), \tag{31}$$

where $T(x'|x)$ defines a *trial probability* that proposes the new configuration x' from the present one x and $A(x'|x)$ is the *acceptance probability*. In the original work by Metropolis and coworkers, the trial probability has been taken symmetric, i.e., $T(x'|x) = T(x|x')$. However,

in the generalized version of the algorithm $T(x'|x)$ can be chosen with large freedom, as long as ergodicity is ensured. Then, in order to define a Markov process that satisfies the detailed balance condition, the proposed configuration x' is accepted with a probability

$$A(x'|x) = \min\left\{1, \frac{\mathcal{P}(x')T(x|x')}{\mathcal{P}(x)T(x'|x)}\right\}. \tag{32}$$

Without loss of generality, we can always choose $T(x|x) = 0$, namely we never propose to remain with the same configuration. Nevertheless, $\omega(x|x)$ can be finite, since the proposed move can be rejected. The actual value of $\omega(x|x)$ is fixed by the normalization condition $\sum_{x'} \omega(x'|x) = 1$.

The proof that detailed balance is satisfied by considering the acceptance probability of Eq. (32) is very simple. Indeed, let us consider the case in which x and $x' \neq x$ are such that

$$\frac{\mathcal{P}(x')T(x|x')}{\mathcal{P}(x)T(x'|x)} > 1, \tag{33}$$

in this case, we have that

$$A(x'|x) = 1, \tag{34}$$

$$A(x|x') = \frac{\mathcal{P}(x)T(x'|x)}{\mathcal{P}(x')T(x|x')}; \tag{35}$$

then, we can directly verify that the detailed balance is satisfied. A similar proof can be obtained in the opposite case where $\mathcal{P}(x')T(x|x')/[\mathcal{P}(x)T(x'|x)] < 1$.

4 Variational wave functions

4.1 The Hartree-Fock wave function

For fermionic models, the simplest example for a variational wave function is given by the Hartree-Fock approximation, where the many-body state is taken to be a product state of suitably optimized single-particle orbitals

$$|\Psi_{\mathrm{HF}}\rangle = \prod_{\alpha=1}^{N_e} \Phi_\alpha^\dagger |0\rangle; \tag{36}$$

here, $\Phi_\alpha^\dagger$ can be expressed in terms of the original fermionic operators as

$$\Phi_\alpha^\dagger = \sum_i W_{\uparrow,\alpha,i}^* c_{i,\uparrow}^\dagger + \sum_i W_{\downarrow,\alpha,i}^* c_{i,\downarrow}^\dagger\,, \tag{37}$$

where $\{W_{\sigma,\alpha,i}\}$ are complex coefficients that can be optimized to get the best variational state. The condition that orbitals are normalized and orthogonal to each other implies that

$$\sum_i \left(W_{\uparrow,\alpha,i} W_{\uparrow,\beta,i}^* + W_{\downarrow,\alpha,i} W_{\downarrow,\beta,i}^*\right) = \delta_{\alpha,\beta}\,. \tag{38}$$

The expectation value of any Hamiltonian can be easily evaluated analytically. On the lattice, it is relatively simple to obtain a solution for the Hartree-Fock equations by using iterative methods. However, while the Hartree-Fock approximation may give reasonable results in the weak-coupling regime, its accuracy becomes questionable for moderate and strong interactions. For example, a Mott insulator, with no symmetry breaking, cannot be described within this approximation; moreover, it is also not possible to stabilize superconducting phases in purely repulsive Hamiltonians. Therefore, a step forward is needed, in order to reach a better description of strongly-correlated systems.

4.2 The Gutzwiller wave function

The simplest example of a correlated state, which goes beyond the Hartree-Fock approximation, has been proposed by Gutzwiller to describe the effect of the Hubbard-U interaction in reducing the weight of configurations with doubly-occupied sites [5]. The Gutzwiller wave function is constructed by starting from the ground state $|\Phi_0\rangle$ of the $U = 0$ model and then applying an operator $\mathcal{P}_{\rm G}$ that suppresses the weight of configurations with doubly-occupied sites

$$|\Psi_{\rm G}\rangle = \mathcal{P}_{\rm G}|\Phi_0\rangle; \tag{39}$$

here, $\mathcal{P}_{\rm G}$ is the so-called Gutzwiller factor that depends upon a single variational parameter g (e.g., $g > 0$ for the repulsive Hubbard model)

$$\mathcal{P}_{\rm G} = \exp\left[-\frac{g}{2}\sum_i (n_i - n)^2\right], \tag{40}$$

where n is the average density.

The effect of the Gutzwiller factor becomes clear once the variational state is expanded in a basis set whose elements $\{|x\rangle\}$ represent configurations with particles sitting on the lattice sites. Indeed, since the Gutzwiller factor is diagonal in this basis (it contains the density operator on each site n_i), we have that

$$\langle x|\Psi_{\rm G}\rangle = \mathcal{P}_{\rm G}(x)\langle x|\Phi_0\rangle, \tag{41}$$

where $\mathcal{P}_{\rm G}(x) \leq 1$ is a number that depends on how many doubly-occupied sites are present in the configuration $|x\rangle$. Therefore, the amplitude of the non-interacting state $\langle x|\Phi_0\rangle$ is renormalized by $\mathcal{P}_{\rm G}(x)$.

For the Hubbard model with a generic hopping amplitude $t_{i,j}$, when the particle density is $n = 1$, a metal-insulator transition is expected at finite values of U/t. However, a simple argument suggests that the Gutzwiller wave function can describe such a transition only when the variational parameter g tends to infinity. Indeed, for $n = 1$, on average, there is one particle per site and density excitations are represented by doublons and holons. In the non-interacting state $|\Phi_0\rangle$, these objects are free to move and then responsible for the conductivity (a doublon is negatively charged with respect to the average background, while the holon is positively charged). The effect of the Gutzwiller factor is to penalize the formation of such objects; however, once

created, doublons and holons are no longer correlated, thus being free to move independently. Only when the energetic penalty is infinite, an insulator is obtained; here, all the density degrees of freedom are frozen and no transport is possible, implying an oversimplified description of a true insulator, where instead density fluctuations are always present. Extensive calculations have shown that g remains finite for all values of U/t and diverges only for $U/t = \infty$ [20,21].

4.3 The fully-projected Gutzwiller wave function

Here, we briefly discuss the limit of $g = \infty$, which corresponds to the *fully-projected* Gutzwiller state. For $n = 1$, the Gutzwiller factor becomes a projector in the sub-space with only singly-occupied sites

$$\mathcal{P}_\infty = \prod_i \left(n_{i,\uparrow} - n_{i,\downarrow}\right)^2 , \tag{42}$$

which implies that all configurations with empty or doubly occupied sites are annihilated. In fermionic models, there is still an exponentially large number of states with one electron per site, which differ by their spin configurations. Therefore, non-trivial spin fluctuations are still allowed. These kinds of fully-projected wave functions $|\Psi_\infty\rangle = \mathcal{P}_\infty|\Phi_0\rangle$ have been considered within the so-called resonating-valence bond (RVB) approach, which has been proposed by Anderson to capture the physics of exotic magnetic systems (spin liquids) [22] in frustrated Heisenberg models [23].

The case to include empty sites, which is relevant when $n < 1$, has been widely used to describe high-temperature superconductors in the t-J model [9,10]. In this case the Gutzwiller projection is written as

$$\mathcal{P}_\infty = \prod_i \left(1 - n_{i,\uparrow} n_{i,\downarrow}\right) , \tag{43}$$

which annihilates all configurations containing doubly-occupied sites, but leaving untouched configurations with only empty and singly-occupied sites.

4.4 The Jastrow wave function

As we have discussed above, the variational description of an insulator with density fluctuations is not captured by the simple Gutzwiller wave function (39) and requires a modification of the correlation term that is applied to the non-interacting wave function. A straightforward generalization of the Gutzwiller wave function is given by the inclusion of long-range terms in the correlator

$$|\Psi_{\mathrm{J}}\rangle = \mathcal{J}|\Phi_0\rangle, \tag{44}$$

where $\mathcal{J}$ is the Jastrow factor [13] that has been introduced in continuum models much before the Gutzwiller wave function. On the lattice, $\mathcal{J}$ takes a simple form

$$\mathcal{J} = \exp\left[-\frac{1}{2}\sum_{i,j} v_{i,j}(n_i - n)(n_j - n)\right] , \tag{45}$$

where $v_{i,j}$ is a pseudo-potential for density-density correlations in the variational state. For translationally invariant models, $v_{i,j}$ only depends upon the relative distance of the two sites i and j, i.e., $|\mathbf{R}_i - \mathbf{R}_j|$; moreover, the on-site term $v_{i,i}$ corresponds to the Gutzwiller parameter g. The Jastrow pseudo-potential can be either parametrized in some way, in order to reduce the number of variational parameters, or optimized for all possible distances, which are $O(L)$ in translationally invariant systems. The role of the long-range tail of the Jastrow factor is to create a bound state between holons and doublons, possibly impeding conduction, but still allowing local density fluctuations. Indeed, we have shown that such Jastrow terms may turn a non-interacting metallic state $|\Phi_0\rangle$ into an insulator [24, 25]. In particular, by denoting the Fourier transform of the (translationally invariant) pseudo-potential $v_{i,j}$ by v_q, the gapless (i.e., metallic) phase is described by having $v_q \approx 1/|\mathbf{q}|$ for $|\mathbf{q}| \to 0$, in any spatial dimension d; by contrast, a fully gapped (i.e., insulating) phase is obtained in one-dimension with $v_q \approx 1/|\mathbf{q}|^2$ for $|\mathbf{q}| \to 0$ [24]. This singular behavior of the pseudo-potential induces an exponential decay of the density-density correlations. In two and three spatial dimensions, a holon-doublon bound-state is generated by $v_q \approx \beta/|\mathbf{q}|^d$ for a sufficiently large value of β [25]. However, these behaviors of the pseudo-potential, which are obtained by an energy minimization, are not sufficient to have a fully gapped phase, since a residual power-law behavior in the density-density correlations is still present.

The Jastrow wave function of Eq. (44) has been introduced to study models on the continuum [13] and has been employed to perform the first quantum Monte Carlo calculation in a many-body system [14]. More precisely, a system of interacting bosons has been considered to model the ground-state properties of ^{4}He in three spatial dimensions with Lennard-Jones interactions. Here, in a first-quantization notation, the wave function with N_b bosons reads

$$\Psi_{\rm J}(\mathbf{r}_1, \ldots, \mathbf{r}_{N_b}) = \prod_{i<j} f(r_{i,j}) = \exp\left[-\sum_{i<j} u(r_{i,j})\right], \tag{46}$$

where $\{\mathbf{r}_i\}$ are the coordinates of the bosons and $f(r_{i,j}) = \exp[-u(r_{i,j})]$ is a function that depends upon the relative distance between two bosons $r_{i,j} = |\mathbf{r}_i - \mathbf{r}_j|$. Notice that the wave function of Eq. (46) is totally symmetric when exchanging two particles, thus having the correct symmetry for a bosonic wave function.

A suitable correlated wave function for N_e fermions can be obtained by applying the symmetric Jastrow factor $\prod_{i<j} f(r_{i,j})$ to a (anti-symmetric) Slater determinant, which, by using first-quantization notations, reads as

$$\Psi_{\rm HF}(\mathbf{r}_1, \ldots, \mathbf{r}_{N_e}) = \det\{\phi_\alpha(\mathbf{r}_j)\}, \tag{47}$$

where $\{\phi_\alpha(\mathbf{r}_j)\}$ is a set of one-particle orbitals. Then the Jastrow-Slater wave function is given by

$$\Psi_{\rm JS}(\mathbf{r}_1, \ldots, \mathbf{r}_{N_e}) = \prod_{i<j} f(r_{i,j}) \times \Psi_{\rm HF}(\mathbf{r}_1, \ldots, \mathbf{r}_{N_e}). \tag{48}$$

In total, this wave function is anti-symmetric when exchanging two particles and, therefore, has the correct symmetry for a fermionic state.

4.5 The backflow wave function

An alternative way to include some correlation inside the original variational state is to introduce a parametrization that allows the orbitals to depend upon the positions of the other particles, leading to the concept of *backflow* correlations. In quantum systems, a particle that moves is surrounded by a counter-flow generated by all the other particles; the existence of this flow pattern pushes away the particles, thus preventing a significant overlap among them. This idea has been originally introduced by Wigner [26] and then developed by Feynman [27, 28] in the context of excitations in ^{4}He and the effective mass of a ^{3}He impurity in liquid ^{4}He. In the fermionic case, the Slater determinant is not constructed with the actual positions of the electrons $(\mathbf{r}_1, \ldots, \mathbf{r}_{N_e})$, see Eq. (47), but with new "coordinates" given by

$$\mathbf{r}_i^b = \mathbf{r}_i + \sum_{j \neq i} \eta(|\mathbf{r}_i - \mathbf{r}_j|)(\mathbf{r}_j - \mathbf{r}_i), \tag{49}$$

where $\eta(|\mathbf{r}_i - \mathbf{r}_j|)$ is a suitable function that describes the effective displacement of the i-th particle due to the j-th one. The simplest wave function is built by taking plane-waves with positions given by $\{\mathbf{r}_i^b\}$. The effect of backflow correlations introduces many-body effects inside the Slater determinant, since, when the i-th electron is moved, all the new "coordinates" are modified, such that all particles respond to the movement of the single electron, adapting their positions accordingly.

Wave functions including both Jastrow factors and backflow correlations have been used to study Helium systems within the so-called hyper-netted chain approximation [29, 30]. They also have been used in Monte Carlo calculations to compute the properties of the homogeneous electron gas in two and three spatial dimensions [31, 32]. The advantage of the backflow wave function is that a single Slater determinant is used, allowing us to perform calculations with a large number of particles.

More recently, the same idea of modifying the single-electron orbitals to improve variational wave functions has been extended for lattice models [33, 34]. Here, the transformation (49) cannot be applied, since electrons live on the lattice sites. Nevertheless, we can imagine that the amplitudes of the Hartree-Fock orbitals (37) are changed according to the many-body configuration

$$W^b_{\sigma,i,\alpha} = \eta_0 W_{\sigma,i,\alpha} + \sum_{j \neq i} \eta_j \mathcal{O}_{i,j} W_{\sigma,j,\alpha}, \tag{50}$$

where $\{\eta_j\}$ is a set of variational parameters and $\mathcal{O}_{i,j}$ is a generic many-body operator that acts on the sites i and j. For example, within the repulsive Hubbard model, the formation of holon-doublon pairs is energetically expensive for large values of U/t; therefore, these objects tend to recombine into singly-occupied sites. In this case, we can consider a many-body operator $\mathcal{O}_{i,j} = D_i H_j$, where D_i (H_i) is the operator that gives 1 if the site i is doubly occupied (empty) and 0 otherwise. Then, the many-body state, which is constructed by taking the Slater determinant of these new "orbitals", will contain terms with single occupation, thus releasing the energy. More complicated expressions of the new "orbital" can be considered, as described in Refs. [33, 34].

5 Practical implementations

In the Monte Carlo evaluation of quantum averages, see Eq. (20), we must compute

- The ratio of probabilities with different configurations, which implies the ratio of overlaps between the given variational state and two configurations of the basis set

$$\frac{\mathcal{P}(x')}{\mathcal{P}(x)} = \left| \frac{\langle x'|\Psi\rangle}{\langle x|\Psi\rangle} \right|^2 , \tag{51}$$

  as required in the Metropolis algorithm.

- The local estimator $\mathcal{O}_L(x)$, which, in turn, implies ratios of overlaps and matrix elements of the observable between states of the basis set. For example, when considering the energy, we have

$$e_L(x) = \frac{\langle x|\mathcal{H}|\Psi\rangle}{\langle x|\Psi\rangle} = \sum_{x'} \langle x|\mathcal{H}|x'\rangle \frac{\langle x'|\Psi\rangle}{\langle x|\Psi\rangle} . \tag{52}$$

Naively, the computation of the local estimator looks a tremendously hard task, since it requires a summation over all the states of the many-body Hilbert space; however, thanks to the locality of the Hamiltonian (or, similarly, any other local operator or correlation function), only few terms actually contribute to the sum. Indeed, given the configuration $|x\rangle$, the matrix element $\langle x|\mathcal{H}|x'\rangle$ is non-zero only for $O(L)$ configurations $|x'\rangle$. As an example, let us consider the fermionic Hubbard model: by using the local basis, $|x\rangle$ is connected only to few other configurations that differ for the hopping of one electron from a given site to one of its neighbors; then, the maximum number of such processes is equal to the number of sites L times the number of bonds times 2 (due to the spin). Therefore, the computation of the local estimator only requires a small number of operations, usually proportional to the number of sites/particles.
Therefore, the building block of the variational Monte Carlo approach is the computation of $\langle x|\Psi\rangle$, which is the amplitude of the variational state over a generic element of the basis set. More precisely, along all the Markov process, only ratios of these overlaps must be computed. This calculation must be done for each configuration that is visited along the Markov process and, therefore, it must be done as fast as possible. This fact imposes some constraint on the form of the variational wave function. Usually, fermionic states require the calculation of determinants. Fortunately, there are fast (i.e., polynomial) algorithms to evaluate determinants, thus allowing us to consider these states as variational *Ansätze* for electron systems.

5.1 The Jastrow factor

Let us focus on the Jastrow factor of Eq. (45). For a translational invariant model, $v_{i,j}$ only depends upon the distance between i and j, thus the number of parameters can be reduced to $O(L)$. We would like to remark that, within the Monte Carlo approach, it is possible to treat exactly (but still having statistical errors) the limit of singular Jastrow factors with $v_{i,i} = \infty$, e.g., the Gutzwiller projector that eliminates all doubly-occupied sites (this case being relevant

for an infinite Hubbard-U interaction). Indeed, this case can be easily incorporated by building a Markov chain where only configurations $|x\rangle$ that satisfy this constraint are visited.
The advantage of considering the Jastrow-Slater wave function in the variational Monte Carlo technique is that the calculations can be extremely efficient and fast. Indeed, whenever the Jastrow factor is diagonal in the chosen basis, we have that

$$\langle x|\Psi_{\mathrm{J}}\rangle = \mathcal{J}(x)\langle x|\Phi_0\rangle, \tag{53}$$

where $\mathcal{J}(x)$ is the value of the Jastrow operator computed for the configuration $|x\rangle$, i.e., $\mathcal{J}|x\rangle = \mathcal{J}(x)|x\rangle$. Therefore, given the electronic configuration, $\mathcal{J}(x)$ is a number that can be evaluated in $O(L^2)$ operations, thus leading to the same complexity when computing the ratio $\mathcal{J}(x')/\mathcal{J}(x)$ that appears in the Metropolis ratio (51). However, whenever the two configurations differ only by few electron hoppings, it is possible to apply a fast computation of the ratio, which involves $O(L)$ operations.
Finally, we emphasize that, in order to have a polynomial algorithm, the Jastrow factor must only contain operators that are diagonal in the basis $|x\rangle$, otherwise $\mathcal{J}|x\rangle$ would generate an exponentially large number of states, ruling out any calculation on large systems.

5.2 Slater determinants

A generic Slater determinant can be obtained as the ground state of a suitable quadratic Hamiltonian $\mathcal{H}_0$. First of all, we contract the spin index σ and the lattice site i into a single index I running from 1 to $2L$

$$c_{i,\uparrow} \equiv d_i, \tag{54}$$
$$c_{i,\downarrow} \equiv d_{i+L}. \tag{55}$$

Then, we start from the simple case in which the non-interacting Hamiltonian is written as

$$\mathcal{H}_0 = \sum_{I,J} t_{I,J} d_I^\dagger d_J, \tag{56}$$

which contains hopping terms only, also including processes in which the spin along z is not conserved, i.e., the terms with $I \leq L$ and $J > L$ and vice-versa. In a compact form, the non-interacting Hamiltonian of Eq. (56) can be written as

$$\mathcal{H}_0 = \mathbf{d}^\dagger \mathbf{T} \mathbf{d} \ , \tag{57}$$

where

$$\mathbf{d}^\dagger = \begin{pmatrix} d_1^\dagger & \dots & d_{2L}^\dagger \end{pmatrix} \tag{58}$$

and

$$\mathbf{T} = \begin{pmatrix} t_{1,1} & \dots & t_{1,2L} \\ \vdots & \ddots & \vdots \\ t_{2L,1} & \dots & t_{2L,2L} \end{pmatrix}. \tag{59}$$

Since $\mathcal{H}_0$ commutes with the total number of electrons $N_e = \sum_I d_I^\dagger d_I$, the eigenstates are single-particle orbitals. In practice, the $2L \times 2L$ matrix $\mathbf{T}$ can be easily diagonalized by using standard libraries (e.g., LAPACK routines)

$$\mathcal{H}_0 = \mathbf{d}^\dagger \mathbf{U}\, \mathbf{U}^\dagger \mathbf{T} \mathbf{U}\, \mathbf{U}^\dagger \mathbf{d} = \mathbf{\Phi}^\dagger \mathbf{E} \mathbf{\Phi} = \sum_\alpha \varepsilon_\alpha \phi_\alpha^\dagger \phi_\alpha, \tag{60}$$

where $\mathbf{U}$ is a unitary matrix (that preserves anti-commutation relations of fermionic operators), $\mathbf{E} = \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_{2L})$ is the diagonal matrix containing the $2L$ eigenvalues ε_α of $\mathbf{T}$, and $\mathbf{\Phi}^\dagger = (\phi_1^\dagger, \ldots, \phi_{2L}^\dagger)$ is defined in terms of the eigenvectors of $\mathbf{T}$

$$\phi_\alpha^\dagger = \sum_I U_{I,\alpha} d_I^\dagger. \tag{61}$$

Now, the many-body state $|\Phi_0\rangle$ can be constructed by occupying the N_e lowest-energy orbitals

$$|\Phi_0\rangle = \prod_{\alpha=1}^{N_e} \phi_\alpha^\dagger |0\rangle = \left(\sum_I U_{I,1} d_I^\dagger \right) \ldots \left(\sum_I U_{I,N_e} d_I^\dagger \right) |0\rangle. \tag{62}$$

The generic configuration, with N_e electrons, which is visited along the Markov process, reads

$$|x\rangle = d_{R_1}^\dagger \ldots d_{R_{N_e}}^\dagger |0\rangle, \tag{63}$$

where $j = 1, \ldots, N_e$ includes both up and down spins and R_j assumes values from 1 to $2L$: the positions of spin-up electrons coincide with the site number, while the positions of spin-down electrons must be shifted by L. The overlap $\langle x|\Phi_0\rangle$ is then given by

$$\begin{aligned} \langle x|\Phi_0\rangle &= \langle 0| d_{R_{N_e}} \ldots d_{R_1} \left(\sum_I U_{I,1} d_I^\dagger \right) \ldots \left(\sum_I U_{I,N_e} d_I^\dagger \right) |0\rangle \\ &= \langle 0| d_{R_{N_e}} \ldots d_{R_1} \left[\sum_p (-1)^p \prod_{\alpha=1}^{N_e} U_{p\{R_j\},\alpha} \right] d_{R_1}^\dagger \ldots d_{R_{N_e}}^\dagger |0\rangle, \end{aligned} \tag{64}$$

where the sum inside the square bracket is over all the possible permutations of the $\{R_j\}$ in $|x\rangle$; the sign appears because of the anti-commutation relations of fermionic operators. Then, we get

$$\langle x|\Phi_0\rangle = \det\{U_{R_j,\alpha}\}. \tag{65}$$

When constructing the many-body state (62), we must pay attention to construct a unique many-body state, i.e., occupy the correct lowest-energy orbitals. When the highest-occupied orbital and the lowest-unoccupied one have different energies (*closed shell* configuration), the choice is unique. Instead, it can also happen that there is a degeneracy that does not allow a unique choice (*open shell* configuration). Whenever the non-interacting Hamiltonian $\mathcal{H}_0$ is diagonalized numerically, the eigenstates provided by standard libraries do not carry definite quantum numbers (like momentum), but are given by generic linear combinations of degenerate orbitals (which erratically depend on the numerical precision used for the computation). This is not a problem

whenever *all* the degenerate eigenstates are included in the many-body state (62), since taking any linear combination of columns in the matrix $\mathbf{U}$ will not change the value of the determinant in Eq. (65). By contrast, not including all degenerate eigenstates will cause a problem, since the determinant will depend on the particular combination of states that is considered. Therefore, whenever a numerical diagonalization is done, we must verify that a closed-shell configuration occurs. Otherwise, the problem of having a vanishing gap can be overcome by constructing suitable orbitals with definite quantum numbers, to obtain a reproducible simulation of the many-body state.

In practice, the diagonalization of the non-interacting Hamiltonian $\mathcal{H}_0$ must be performed at the beginning of the Monte Carlo calculation; then, we need to store the reduced part of the $\mathbf{U}$, obtained by keeping only the N_e columns that correspond to occupied orbitals. This is a $2L \times N_e$ matrix

$$\mathbf{U} = \begin{pmatrix} U_{1,1} & \dots & U_{1,N_e} \\ \vdots & \ddots & \vdots \\ U_{2L,1} & \dots & U_{2L,N_e} \end{pmatrix}. \tag{66}$$

Then, the overlap with a generic electronic configuration is given by the determinant of the matrix obtained by considering only the rows of (66) corresponding to the electron positions $\{R_j\}$, thus giving a $N_e \times N_e$ matrix.

Remarkably, the same kind of formalism can be used also if the non-interacting Hamiltonian contains an electron pairing that couples up and down spins. Indeed, let us consider a Bardeen-Cooper-Schrieffer (BCS) Hamiltonian described by

$$\mathcal{H}_0 = \sum_{i,j,\sigma} t_{i,j} c^\dagger_{i,\sigma} c_{j,\sigma} - \mu_0 \sum_{i,\sigma} c^\dagger_{i,\sigma} c_{i,\sigma} + \sum_{i,j} \Delta_{i,j} c^\dagger_{i,\uparrow} c^\dagger_{j,\downarrow} + \text{h.c.}, \tag{67}$$

where we have included a chemical potential μ_0, which fixes, on average, the number of electrons. In this case, the total number of particles is not conserved and the concept of single-particle orbitals is not defined. Indeed, the ground state of the BCS Hamiltonian is naturally written in terms of a pairing function. Nevertheless, we can perform a particle-hole transformation on the spin-down electrons

$$c_{i,\uparrow} \to f_{i,\uparrow} \equiv d_i \tag{68}$$

$$c_{i,\downarrow} \to f^\dagger_{i,\downarrow} \equiv d^\dagger_{i+L}. \tag{69}$$

Then, apart from constant terms, the transformed BCS Hamiltonian has the form of Eq. (57). Since, after the particle-hole transformation, the number of particles (but not the z component of the spin) is conserved, the eigenstates of the BCS Hamiltonian can be expressed into "orbitals", similarly to the ones of Eq. (61), but without having a definite spin component along z.

5.3 Fast computation of the determinants

Let us now show how to compute efficiently the ratio of determinants when the two configurations $|x\rangle$ and $|x'\rangle$ differ by one or few electron hoppings. According to Eq. (65) the overlap

$\langle x|\Phi_0\rangle$ is given by the determinant of $\tilde{U}_{j,\alpha} \equiv U_{R_j,\alpha}$

$$\tilde{\mathbf{U}} = \begin{pmatrix} U_{R_1,1} & \dots & U_{R_1,N_e} \\ \vdots & \ddots & \vdots \\ U_{R_l,1} & \dots & U_{R_l,N_e} \\ \vdots & \ddots & \vdots \\ U_{R_{N_e},1} & \dots & U_{R_{N_e},N_e} \end{pmatrix}, \tag{70}$$

which is obtained taking only the rows corresponding to the occupied sites of the matrix $\mathbf{U}$ of Eq. (66).
Let us start and consider the electronic configurations in which $|x'\rangle$ is obtained from $|x\rangle$ just by hopping the one electron from R_l to R'_l, i.e., $|x'\rangle = d^\dagger_{R'_l} d_{R_l}|x\rangle$. The new matrix $\tilde{\mathbf{U}}'$ will be equal to $\tilde{\mathbf{U}}$, except that the elements of the l-th row will be changed from $U_{R_l,\alpha}$ to $U_{R'_l,\alpha}$

$$\tilde{\mathbf{U}}' = \begin{pmatrix} U_{R_1,1} & \dots & U_{R_1,N_e} \\ \vdots & \ddots & \vdots \\ U_{R'_l,1} & \dots & U_{R'_l,N_e} \\ \vdots & \ddots & \vdots \\ U_{R_{N_e},1} & \dots & U_{R_{N_e},N_e} \end{pmatrix}. \tag{71}$$

Then, the ratio between two configurations that differ only by a single fermion hopping is

$$\frac{\langle x'|\Phi_0\rangle}{\langle x|\Phi_0\rangle} = \frac{\langle x|d^\dagger_{R_l} d_{R'_l}|\Phi_0\rangle}{\langle x|\Phi_0\rangle} = \frac{\det \tilde{\mathbf{U}}'}{\det \tilde{\mathbf{U}}}\,. \tag{72}$$

By denoting with K the new site of the l-th electron (i.e., $K \equiv R'_l$), the updated matrix elements are given by a compact form

$$\tilde{U}'_{j,\alpha} = \tilde{U}_{j,\alpha} + \delta_{j,l}(U_{K,\alpha} - \tilde{U}_{l,\alpha}) = \tilde{U}_{j,\alpha} + \delta_{j,l} v^{K,l}_\alpha, \tag{73}$$

where we have defined $v^{K,l}_\alpha \equiv U_{K,\alpha} - \tilde{U}_{l,\alpha}$; here the indices K and l are fixed, since they specify the site where the electron is hopping and the electron index, respectively. This equation can be rewritten in the following way

$$\tilde{U}'_{j,\alpha} = \sum_\beta \tilde{U}_{j,\beta} \left(\delta_{\beta,\alpha} + \tilde{U}^{-1}_{\beta,l} v^{K,l}_\alpha \right) = \sum_\beta \tilde{U}_{j,\beta} Q_{\beta,\alpha}\,, \tag{74}$$

where

$$Q_{\beta,\alpha} = \delta_{\beta,\alpha} + \tilde{U}^{-1}_{\beta,l} v^{K,l}_\alpha. \tag{75}$$

Therefore, $\tilde{\mathbf{U}}' = \tilde{\mathbf{U}}\mathbf{Q}$, which implies that the calculation of the ratio of the determinants of $\tilde{\mathbf{U}}'$ and $\tilde{\mathbf{U}}$ is equivalent to the calculation of the determinant of $\mathbf{Q}$

$$\frac{\det \tilde{\mathbf{U}}'}{\det \tilde{\mathbf{U}}} = \det \mathbf{Q}. \tag{76}$$

The great simplification comes from the fact that the determinant of $\mathbf{Q}$ can be easily computed. Indeed, $\mathbf{Q}$ has a particularly simple form that can be written as

$$Q_{\beta,\alpha} = \delta_{\beta,\alpha} + \mathcal{B}_\beta \mathcal{A}_\alpha, \tag{77}$$

where $\mathcal{B}_\beta = \tilde{U}^{-1}_{\beta,l}$ and $\mathcal{A}_\alpha = v^{K,l}_\alpha$. Although the matrix is not Hermitian, the eigenvalues of the matrix $\mathbf{Q}$ can be obtained from the secular equation

$$\sum_\alpha Q_{\beta,\alpha} v_\alpha = \lambda v_\beta; \tag{78}$$

by using the explicit form of Eq. (77), we obtain

$$v_\beta + \mathcal{B}_\beta \sum_\alpha \mathcal{A}_\alpha v_\alpha = \lambda v_\beta, \tag{79}$$

which implies that all vectors v_α that are orthogonal to $\mathcal{A}_\alpha$ are eigenvectors with eigenvalue $\lambda = 1$ (there are $N_e - 1$ of such vectors); in addition, $v_\alpha = \mathcal{B}_\alpha$ is also eigenvector with $\lambda = 1 + \sum_\alpha \mathcal{A}_\alpha \mathcal{B}_\alpha = \sum_\alpha U_{K,\alpha} \tilde{U}^{-1}_{\alpha,l}$. Therefore, we have that

$$\frac{\det \tilde{\mathbf{U}}'}{\det \tilde{\mathbf{U}}} = \det \mathbf{Q} = \sum_\alpha U_{K,\alpha} \tilde{U}^{-1}_{\alpha,l}. \tag{80}$$

Having stored (at the beginning of the simulation) the matrix $\tilde{\mathbf{U}}^{-1}$ for the configuration $|x\rangle$, this calculation requires $O(N_e)$ operations, instead of the $O(N_e^3)$ needed to evaluate a determinant. Then, once the new configuration $|x'\rangle$ is accepted along the Markov process, the matrix $\tilde{\mathbf{U}}^{-1}$ must be updated. This can be done in $O(N_e^2)$ operations. In fact, we have that $(\tilde{\mathbf{U}}')^{-1} = \mathbf{Q}^{-1}\tilde{\mathbf{U}}^{-1}$, the inverse of the matrix $\mathbf{Q}$ being given by (as easily verified)

$$Q^{-1}_{\alpha,\beta} = \delta_{\alpha,\beta} - \frac{1}{\det \mathbf{Q}} \mathcal{B}_\alpha \mathcal{A}_\beta. \tag{81}$$

Then, the updated matrix elements of $(\tilde{\mathbf{U}}')^{-1}$ are given by

$$\tilde{U}^{-1\prime}_{\alpha,j} = \tilde{U}^{-1}_{\alpha,j} - \frac{\tilde{U}^{-1}_{\alpha,l}}{\det \mathbf{Q}} \left(\sum_\beta U_{K,\beta} \tilde{U}^{-1}_{\beta,j} - \delta_{l,j} \right). \tag{82}$$

This is a closed equation for updating the matrix $\tilde{\mathbf{U}}^{-1}$.

We would like to emphasize that the previous results for the calculation of the ratio of determinants and the updating can be further simplified. Indeed, at the beginning of the calculation, we can compute and store a $2L \times N_e$ matrix $\mathbf{W}$, whose elements are given by

$$W_{I,j} = \sum_\alpha U_{I,\alpha} \tilde{U}^{-1}_{\alpha,j}; \tag{83}$$

then the ratio of determinants (80) costs $O(1)$ operations, since it consists in taking the element corresponding to the new site (row) and the electron performing the hopping process (column)

$$\frac{\det \tilde{\mathbf{U}}'}{\det \tilde{\mathbf{U}}} = W_{K,l}. \tag{84}$$

The evaluation of $\mathbf{W}$ requires the knowledge of the full matrix $\mathbf{U}$, which has been computed and stored once for all at the beginning of the simulation (it does not depend upon the electronic configuration), and $\tilde{\mathbf{U}}^{-1}$, which instead depends upon the configuration $|x\rangle$. Then, a simple updating scheme for $\mathbf{W}$ is possible. In fact, by multiplying both sides of Eq. (82) by $U_{I,\alpha}$ and summing over α, we obtain

$$W'_{I,j} = W_{I,j} - \frac{W_{I,l}}{W_{K,l}} \left(W_{K,j} - \delta_{l,j} \right), \tag{85}$$

where, we have used that $\det \mathbf{Q} = W_{K,l}$, according to Eq. (84) and the definition of Eq. (83). Since each matrix element must be updated with $O(1)$ operations, the total cost is $O(2LN_e)$.
Let us finish this part by generalizing the previous formalism to the case where more than one electron hop, i.e., $|x'\rangle = d^{\dagger}_{R'_{l_1}} d_{R_{l_1}} \dots d^{\dagger}_{R'_{l_m}} d_{R_{l_m}} |x\rangle$, thus leading to a modification of m rows of the $\tilde{\mathbf{U}}$ matrix; for example, this could be the case for pair-hopping or spin-flip processes. Then, Eq. (73) generalizes into

$$\tilde{U}'_{j,\alpha} = \tilde{U}_{j,\alpha} + \sum_{r=1}^{m} \delta_{j,l_r} (U_{K_r,\alpha} - \tilde{U}_{l_r,\alpha}) = \tilde{U}_{j,\alpha} + \sum_{r=1}^{m} \delta_{j,l_r} v_{\alpha}^{K_r,l_r}; \tag{86}$$

as before, the indices K_r and l_r (for $r = 1, \dots, m$) are fixed, because they specify the sites where the electrons are hopping and the electron indices, respectively. By performing the same algebra as before, we get

$$\tilde{U}'_{j,\alpha} = \sum_{\beta} \tilde{U}_{j,\beta} \left(\delta_{\beta,\alpha} + \sum_{r=1}^{m} \tilde{U}^{-1}_{\beta,l_r} v_{\alpha}^{K_r,l_r} \right) = \sum_{\beta} \tilde{U}_{j,\beta} Q_{\beta,\alpha}, \tag{87}$$

where now the matrix $\mathbf{Q}$ has the following form

$$Q_{\beta,\alpha} = \delta_{\beta,\alpha} + \sum_{r=1}^{m} \mathcal{B}^r_{\beta} \mathcal{A}^r_{\alpha}, \tag{88}$$

where $\mathcal{B}^r_{\beta} = \tilde{U}^{-1}_{\beta,l_r}$ and $\mathcal{A}^r_{\alpha} = v_{\alpha}^{K_r,l_r}$. As before, the determinant of $\mathbf{Q}$ can be easily computed by solving the corresponding eigenvalue problem

$$\sum_{\alpha} Q_{\beta,\alpha} v_{\alpha} = v_{\beta} + \sum_{r=1}^{m} \mathcal{B}^r_{\beta} \sum_{\alpha} \mathcal{A}^r_{\alpha} v_{\alpha} = \lambda v_{\beta}, \tag{89}$$

which implies that all vectors that are orthogonal to the subspace defined by the $\mathcal{A}^r_{\alpha}$'s are eigenvectors with $\lambda = 1$; moreover, $v_{\alpha} = \sum_{r=1}^{m} x_r \mathcal{B}^r_{\alpha}$ is an eigenvector provided that the coefficients x_r satisfy

$$\sum_{s=1}^{m} \left(\delta_{r,s} + \sum_{\alpha} \mathcal{A}^r_{\alpha} \mathcal{B}^s_{\alpha} \right) x_s = \lambda x_r. \tag{90}$$

Therefore, the m non-trivial eigenvalues of $\mathbf{Q}$ are those of the $m \times m$ matrix

$$C_{r,s} = \delta_{r,s} + \sum_{\alpha} \mathcal{A}^r_{\alpha} \mathcal{B}^s_{\alpha} = W_{K_r,l_s}. \tag{91}$$

The final expression of the ratio of the two determinants is given by

$$\frac{\det \tilde{\mathbf{U}}'}{\det \tilde{\mathbf{U}}} = \det(W_{K_r,l_s}). \tag{92}$$

Also in this case, once the move is accepted, we have to update the matrix $(\tilde{\mathbf{U}}')^{-1} = \mathbf{Q}^{-1}\tilde{\mathbf{U}}^{-1}$. As before, the inverse of the $\mathbf{Q}$ matrix can be obtained

$$Q^{-1}_{\alpha,\beta} = \delta_{\alpha,\beta} - \sum_{r,s=1}^{m} \mathcal{B}^r_\alpha C^{-1}_{r,s} \mathcal{A}^s_\beta \,. \tag{93}$$

Therefore, we get

$$\tilde{U}^{-1\prime}_{\alpha,j} = \tilde{U}^{-1}_{\alpha,j} - \sum_{r,s=1}^{m} \tilde{U}^{-1}_{\alpha,l_r} C^{-1}_{r,s} \left(\sum_\beta U_{K_s,\beta} \tilde{U}^{-1}_{\beta,j} - \delta_{l_s,j} \right). \tag{94}$$

Then, the updated $\mathbf{W}'$ is obtained by multiplying both sides of the previous equation by $U_{i,\alpha}$ and summing over α

$$W'_{I,j} = W_{I,j} + \sum_{r=1}^{m} W_{I,l_r} b^{(r)}_j, \tag{95}$$

where

$$b^{(r)}_j = -\sum_{s=1}^{m} C^{-1}_{r,s} \big(W_{K_s,j} - \delta_{l_s,j} \big). \tag{96}$$

5.4 Backflow correlations

Here, we would like to discuss how to implement the updating of the determinant part in the presence of backflow correlations that have been introduced on lattice problem. In practice, we consider a quadratic Hamiltonian to construct the non-interacting orbitals $\{U_{I,\alpha}\}$, see Eq. (61). In the simplest approach, in which the backflow correlations act on holon-doublon (nearest-neighbor) pairs [33], the overlap between the generic configuration $|x\rangle$ and the backflow wave function $|\Phi^b_0\rangle$ is constructed from the "correlated" orbital with backflow correction

$$U^b_{I,\alpha} = \eta_0 U_{I,\alpha} + \eta_1 \sum_{\langle j \rangle_i} D_i H_j U_{J,\alpha}, \tag{97}$$

where $I = i$ ($I = i + L$) for electrons with spin up (down), and equivalently for J and j; $\langle j \rangle_i$ indicates the sites j that are nearest neighbors of i; D_i (H_i) is the operator that gives 1 if the site i if doubly occupied (empty) and 0 otherwise; finally η_0 and η_1 are variational parameters. Then, the wave function for the given configuration $|x\rangle$ is obtained by taking the determinant of the matrix

$$\tilde{\mathbf{U}}^b = \begin{pmatrix} U^b_{R_1,1} & \dots & U^b_{R_1,N_e} \\ \vdots & \ddots & \vdots \\ U^b_{R_{N_e},1} & \dots & U^b_{R_{N_e},N_e} \end{pmatrix}, \tag{98}$$

in which the rows correspond to the sites occupied by the electrons.

In practice, we can consider more general forms for the "correlated" orbitals, by considering further terms [34], but still remaining with the spirit of considering a linear combination of non-interacting orbitals depending on the many-body configuration $|x\rangle$ (in the previous case, the linear combination is taken for configurations having holons and doublons at nearest-neighbor sites).

6 Optimization techniques

6.1 Calculation of derivatives

In this section, we consider wave functions that depend upon a set of p variational parameters, which are arranged into a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ and explicitly specified in the definition of the quantum state

$$\Psi_\alpha(x) = \langle x|\Psi_\alpha\rangle, \tag{99}$$

In the following, we discuss the basics ingredients that are necessary to compute the derivatives of the variational energy with respect to a given variational parameter α_k

$$f_k = -\frac{\partial E_\alpha}{\partial \alpha_k} = -\frac{\partial}{\partial \alpha_k}\frac{\langle\Psi_\alpha|\mathcal{H}|\Psi_\alpha\rangle}{\langle\Psi_\alpha|\Psi_\alpha\rangle}. \tag{100}$$

The dependence of E_α on the variational parameters is just a consequence of the fact that the wave function $|\Psi_\alpha\rangle$ depends upon $\boldsymbol{\alpha}$. Thus, in order to differentiate E_α, it is convenient to expand $|\Psi_\alpha\rangle$ for small changes $\alpha_k \to \alpha_k + \delta\alpha_k$. For a given configuration $|x\rangle$, where $\Psi_\alpha(x)$ is a complex number (in case of complex parameters, we can assume that all the α_k are real, once we consider their real and imaginary parts separately), we have that

$$\Psi_{\alpha+\delta\alpha_k}(x) = \Psi_\alpha(x) + \delta\alpha_k \frac{\partial\Psi_\alpha(x)}{\partial\alpha_k} + O(\delta\alpha_k^2), \tag{101}$$

where the notation $\Psi_{\alpha+\delta\alpha_k}(x)$ means that only the component α_k of the vector $\boldsymbol{\alpha}$ has been incremented by $\delta\alpha_k$. In the following, for simplicity, we assume that $\Psi_\alpha(x) \neq 0$ for all the configurations. For fermionic systems in the continuous space, the nodal region $\Psi_\alpha(x) = 0$ represents a negligible (i.e., with zero measure) integration domain. On the lattice, accidental configurations with $\Psi_\alpha(x) = 0$ can be removed by considering a tiny perturbation of the variational *Ansatz* (e.g., by adding a small noisy part) and considering the limit of vanishing perturbation. Then, Eq. (101) can be formally written in terms of a local operator $\mathcal{O}_k$, corresponding to the parameter α_k and defined by diagonal matrix elements $\mathcal{O}_k(x)$

$$\langle x|\mathcal{O}_k|x'\rangle = \delta_{x,x'}\mathcal{O}_k(x), \tag{102}$$

$$\mathcal{O}_k(x) = \frac{\partial \ln\Psi_\alpha(x)}{\partial\alpha_k} = \frac{1}{\Psi_\alpha(x)}\frac{\partial\Psi_\alpha(x)}{\partial\alpha_k}; \tag{103}$$

here, in principle, $\mathcal{O}_k(x)$ may depend upon the variational parameters $\boldsymbol{\alpha}$, however, to keep the notation simple, we prefer not to put the label α in the local operators. The important point is

that $\mathcal{O}_k(x)$ can be usually computed for the given *Ansatz* of the variational state. In this way, we can write a formal expansion of the many-body state as

$$|\Psi_{\alpha+\delta\alpha_k}\rangle = (1 + \delta\alpha_k \mathcal{O}_k)|\Psi_\alpha\rangle, \tag{104}$$

which can be readily verified by taking the overlap of both sides of the above equation with $|x\rangle$ and using Eqs. (102) and (103). Notice that the diagonal operator $\mathcal{O}_k$ is not necessarily Hermitian, as its diagonal elements are not necessarily real, for a generic complex case.
Let us now show how to obtain the explicit form of the energy derivative with respect to a given variational parameter. It is clear that the variational energy E_α (as well as any other correlation function) does not depend on the overall normalization (and global phase) of the wave function. In other words, by scaling the wave function by an arbitrary complex constant c, i.e., $|\Psi_\alpha\rangle \to c|\Psi_\alpha\rangle$, E_α remains unchanged. In order to exploit this property, it is better to consider explicitly normalized wave functions. First of all we define

$$|v_{0,\alpha}\rangle \equiv \frac{|\Psi_\alpha\rangle}{\|\Psi_\alpha\|}, \tag{105}$$

where $\|\Psi_\alpha\|$ indicates the norm of the state $|\Psi_\alpha\rangle$. Then, we define a set of states (one for each value of $k = 1, \ldots, p$)

$$|v_{k,\alpha}\rangle \equiv (\mathcal{O}_k - \overline{\mathcal{O}}_k)|v_{0,\alpha}\rangle, \tag{106}$$

where

$$\overline{\mathcal{O}}_k = \langle v_{0,\alpha}|\mathcal{O}_k|v_{0,\alpha}\rangle = \frac{\langle\Psi_\alpha|\mathcal{O}_k|\Psi_\alpha\rangle}{\langle\Psi_\alpha|\Psi_\alpha\rangle}. \tag{107}$$

The states $|v_{k,\alpha}\rangle$ are orthogonal to $|v_{0,\alpha}\rangle$, as easily verified when using Eq. (107); however, they are neither normalized nor orthogonal to each other, i.e., in general $\langle v_{k,\alpha}|v_{k',\alpha}\rangle \neq \delta_{k,k'}$ for $k, k' \neq 0$. Therefore, the set of states $|v_{0,\alpha}\rangle$ and $\{|v_{k,\alpha}\rangle\}$ defines a *semi-orthogonal* basis.
In order to compute the normalized wave function when the parameter α_k is changed, we first compute the norm of $|\Psi_{\alpha+\delta\alpha_k}\rangle$

$$\begin{aligned}\|\Psi_{\alpha+\delta\alpha_k}\|^2 &= \langle\Psi_\alpha|(1 + \delta\alpha_k\mathcal{O}_k)^*(1 + \delta\alpha_k\mathcal{O}_k)|\Psi_\alpha\rangle \\ &= \|\Psi_\alpha\|^2 \left[1 + 2\Re(\delta\alpha_k\overline{\mathcal{O}}_k) + O(\delta\alpha_k^2)\right].\end{aligned} \tag{108}$$

Then, we have that

$$\begin{aligned}|v_{0,\alpha+\delta\alpha_k}\rangle = \frac{|\Psi_{\alpha+\delta\alpha_k}\rangle}{\|\Psi_{\alpha+\delta\alpha_k}\|} &= |v_{0,\alpha}\rangle + \left[\delta\alpha_k\mathcal{O}_k - \Re(\delta\alpha_k\overline{\mathcal{O}}_k)\right]|v_{0,\alpha}\rangle + O(\delta\alpha_k^2) \\ &= [1 + i\Im(\delta\alpha_k\overline{\mathcal{O}}_k)]|v_{0,\alpha}\rangle + \delta\alpha_k|v_{k,\alpha}\rangle + O(\delta\alpha_k^2),\end{aligned} \tag{109}$$

which can be finally recast as

$$|v_{0,\alpha+\delta\alpha_k}\rangle = \exp(i\delta\phi)\left[|v_{0,\alpha}\rangle + \delta\alpha_k|v_{k,\alpha}\rangle\right] + O(\delta\alpha_k^2), \tag{110}$$

where $\delta\phi = \Im(\delta\alpha_k\overline{\mathcal{O}}_k)$.

By using the above expression, it is immediate to work out the derivative of the variational energy E_α with respect to a given variational parameter α_k

$$\frac{\partial E_\alpha}{\partial \alpha_k} = \lim_{\delta\alpha_k \to 0} \frac{\langle v_{0,\alpha+\delta\alpha_k}|\mathcal{H}|v_{0,lpha+\delta\alpha_k}\rangle - \langle v_{0,\alpha}|\mathcal{H}|v_{0,\alpha}\rangle}{\delta\alpha_k} = \langle v_{k,\alpha}|\mathcal{H}|v_{0,\alpha}\rangle + \langle v_{0,\alpha}|\mathcal{H}|v_{k,\alpha}\rangle$$
$$= 2\Re\left[\frac{\langle \Psi_\alpha|\mathcal{H}(\mathcal{O}_k - \overline{O}_k)|\Psi_\alpha\rangle}{\langle\Psi_\alpha|\Psi_\alpha\rangle}\right]. \tag{111}$$

Notice also that, as expected, the phase factor $\delta\phi$ does not enter in the above expression.
In order to evaluate the force f_k by a standard Monte Carlo sampling, we introduce a completeness relation to have

$$f_k = -2\Re\left[\frac{\sum_x \langle\Psi_\alpha|\mathcal{H}|x\rangle\langle x|(\mathcal{O}_k - \overline{\mathcal{O}}_k)|\Psi_\alpha\rangle}{\sum_x \langle\Psi_\alpha|x\rangle\langle x|\Psi_\alpha\rangle}\right] =$$
$$= -2\Re\left[\sum_x \frac{e_L^*(x)(\mathcal{O}_k(x) - \overline{\mathcal{O}}_k)|\Psi_\alpha(x)|^2}{\sum_x |\Psi_\alpha(x)|^2}\right], \tag{112}$$

where $e_L^*(x)$ is the complex conjugate of the local energy (here, we omitted the index α, in harmony with the notation adopted for the local operators $\mathcal{O}_k$), as it is generally a complex-valued function. Then, f_k can be evaluated by considering

$$f_k \approx -2\Re\left[\frac{1}{N}\sum_{i=1}^N e_L^*(x_i)\big(\mathcal{O}_k(x_i) - \overline{\mathcal{O}}_k\big)\right], \tag{113}$$

$$\overline{\mathcal{O}}_k \approx \frac{1}{N}\sum_{i=1}^N \mathcal{O}_k(x_i). \tag{114}$$

6.2 The stochastic reconfiguration

The knowledge of energy derivatives of Eq. (100) allows us to employ the steepest-descent method [35] to change the variational parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$, even when p is very large

$$\alpha_k' = \alpha_k + \delta\alpha_k, \tag{115}$$

$$\delta\alpha_k = \Delta f_k, \tag{116}$$

where Δ is an arbitrary (small) constant. In principle, its value can be optimized to reach the lowest possible energy at each iteration; however, in most applications, it is common practice to keep Δ constant along the minimization procedure. Then, the variational parameters are iteratively improved along a Markov chain procedure. In absence of noise, the steepest-descent method always converges to a minimum, where the Euler conditions $f_k = 0$ are satisfied. Indeed, let us suppose that $f_k \neq 0$, then the energy for $\boldsymbol{\alpha}'$ is given by a Taylor expansion to linear order in Δ

$$E_{\alpha'} = E_\alpha + \sum_k \frac{\partial E_\alpha}{\partial \alpha_k}\delta\alpha_k + O(\Delta^2) = E_\alpha - \Delta\sum_k f_k^2 + O(\Delta^2), \tag{117}$$

where we used that $\partial E_\alpha/\partial\alpha_k = -f_k$ and $\delta\alpha_k = \Delta f_k$. Therefore, for small Δ, when the linear truncation is accurate enough in the Taylor expansion, we obtain that

$$\Delta E \equiv E_{\alpha'} - E_\alpha = -\Delta \sum_k f_k^2 \leq 0; \tag{118}$$

here, the equality sign holds only when $f_k = 0$. Thus, the method converges to a minimum for a large number of iterations just because the energy monotonically decreases with the number of iterations. Within the steepest-descent method only the first derivative of the energy is computed and it is certain that a small change of the parameters $\boldsymbol{\delta\alpha} = (\delta\alpha_1, \ldots, \delta\alpha_p)$ parallel to the force $\mathbf{f} = (f_1, \ldots, f_p)$ will decrease the energy; the only issue concerns the size of Δ, which must be taken sufficiently small to make the quadratic term in Eq. (117) negligible.

Although the steepest-descent approach eventually converges to a local minimum, its effectiveness may suffer from cases where the dependence on the variational parameters is highly non-linear (e.g., in the Jastrow factors). In this case, a small change of a given variational parameter can produce very different wave functions and physical quantities, whereas another parameter may weakly affect the wave function. In order to overcome these difficulties, it is important to introduce an appropriate metric δs^2 that is used to estimate the "proximity" of two normalized (complex) wave functions $|v_{0,\alpha}\rangle$ and $|v_{0,\alpha+\delta\alpha}\rangle$

$$\delta s^2 = \min_{\delta\theta} \| \exp(-i\delta\theta) v_{0,\alpha+\delta\alpha} - v_{0,\alpha} \|^2. \tag{119}$$

Here, the minimization on the phase factor $\delta\theta$ is necessary because we do not want to distinguish between two wave functions that differ only by an overall phase factor, as they produce the same correlation functions. In other words, we want to define a distance δs^2 that vanishes when we have physically equivalent wave functions. Then, we replace in Eq. (119) the expression for $|v_{0,\alpha+\delta\alpha}\rangle$ that is obtained by generalizing Eq. (110) to the case where several parameters are changed

$$|v_{0,\alpha+\delta\alpha}\rangle = \exp(i\delta\phi) \left[|v_{0,\alpha}\rangle + \sum_k \delta\alpha_k |v_{k,\alpha}\rangle \right] + O(|\delta\alpha|^2), \tag{120}$$

where $\delta\phi = \sum_k \Im(\delta\alpha_k \overline{\mathcal{O}}_k)$. Now, the minimization over $\delta\theta$ gives $\delta\theta = \delta\phi$, thus leading to

$$\delta s^2 = \sum_{k,k'} \langle v_{k,\alpha} | v_{k',\alpha} \rangle \delta\alpha_k \delta\alpha_{k'} + O(|\delta\alpha|^2). \tag{121}$$

Since all increments $\delta\alpha_k$ are assumed real (as discussed previously, here we assume that all parameters are real), we can symmetrize the previous expression with respect to the indices k and k' and neglect the terms that are $O(|\delta\alpha|^2)$, obtaining

$$\delta s^2 = \frac{1}{2} \sum_{k,k'} \left(\langle v_{k,\alpha} | v_{k',\alpha} \rangle + \langle v_{k',\alpha} | v_{k,\alpha} \rangle \right) \delta\alpha_k \delta\alpha_{k'}. \tag{122}$$

In this way, we can finally identify a matrix $\mathbf{S}$ that fully determines the metric in the space of normalized wave functions

$$S_{k,k'} = \Re \left(\langle v_{k,\alpha} | v_{k',\alpha} \rangle \right), \tag{123}$$

which implies that the distance between two wave functions reads

$$\delta s^2 = \sum_{k,k'} S_{k,k'}\, \delta\alpha_k \delta\alpha_{k'}. \tag{124}$$

At this point, it is natural to improve the steepest-descent method by using the metric given by $\mathbf{S}$. The minimization of $\Delta E + \mu \delta s^2$ with the metric δs^2 given in Eq. (124) improves the convergence to the minimum of the variational energy with respect to the simple steepest-descend approach, as non-equivalent parameters can be appropriately changed with a different scale. This approach is called *stochastic reconfiguration* [16]. The minimization of $\Delta E + \mu \delta s^2$ gives

$$\sum_{k'} S_{k,k'}\, \delta\alpha_{k'} = \frac{f_k}{2\mu}, \tag{125}$$

which is a set of linear equations for the unknown vector $\boldsymbol{\delta\alpha}$. After having solved this linear system, we can update the variational parameters until convergence is reached; as in the steepest-descent method, we can set $\Delta = 1/(2\mu)$ small enough, which may be kept fixed during the optimization. We would like to stress the fact that, since the matrix $\mathbf{S}$ is strictly positive definite, the energy is monotonically decreasing along the optimization as

$$\Delta E = -\Delta \sum_{k,k'} S^{-1}_{k,k'} f_k f_{k'} < 0. \tag{126}$$

Within a Monte Carlo procedure, the matrix $\mathbf{S}$ is evaluated by a finite sampling of N configurations $\{x_i\}$ as

$$S_{k,k'} \approx \Re \left[\frac{1}{N} \sum_{i=1}^{N} (\mathcal{O}_k(x_i) - \overline{\mathcal{O}}_k)(\mathcal{O}_{k'}(x_i) - \overline{\mathcal{O}}_{k'}) \right]; \tag{127}$$

note that the forces are computed in a similar way, see Eq. (113).

7 Selected results

Here, we would like to show a few selected results that have been obtained by using Jastrow-Slater wave functions, also including backflow correlations. First of all, we would like to emphasize the importance of the long-range tail in the Jastrow factor, in order to correctly reproduce the strong-coupling regime of the Hubbard model. In Fig. 1, we report the variational energies for the one-dimensional Hubbard model at half filling for two different clusters with $L = 18$ and 82 sites. The Slater determinant is obtained by filling the lowest-energy states of a the quadratic Hamiltonian (56) with only nearest-neighbor hopping (dubbed as "Fermi sea"). The comparison is between the cases with (on-site) Gutzwiller and (long-range) Jastrow factors. It is well known that, the fully-projected Gutzwiller state gives very accurate energies for the Heisenberg model [36]. Instead, considering the Hubbard model with finite repulsion U/t, it turns out that, in the strong-coupling limit, the Gutzwiller state gives a rather poor variational description, missing completely the super-exchange energy generated by the virtual-hopping processes. This happens because, by increasing U/t, the Gutzwiller parameter g increases, and

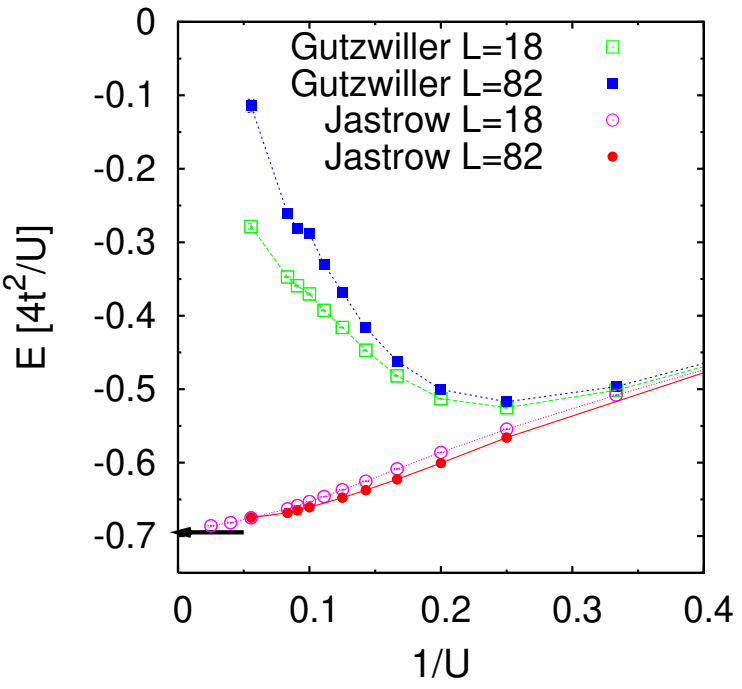

Fig. 1: *Energy per site (in units of $4t^2/U$) for the one-dimensional Hubbard model at half filling. The results for wave functions in which the Gutzwiller and Jastrow factors are applied to the Fermi sea are reported for two cluster sizes ($L = 18$ and 82). The arrow indicates the energy per site for the fully-projected Fermi sea in the Heisenberg model.*

the hopping processes, which create double occupancies, become less probable, with a consequent kinetic-energy loss. By contrast, the long-range Jastrow factor enables us to connect the fully-projected insulator valid in the strong-coupling limit to an insulating state at finite U/t, as demonstrated by the fact that the variational energy of $|\Psi_{\rm J}\rangle$ approaches the one calculated with the fully-projected Gutzwiller wave function.

In two dimensions (or for frustrated one-dimensional cases, with next-nearest-neighbor hopping t'), the situation is more delicate: the Jastrow factor is no longer sufficient to correctly reproduce the ground-state properties in the strongly-correlated regime and, in many cases, backflow correlations are necessary to reach high accuracies. For example, we consider the Hubbard model with $t' = 0$ and $t'/t = 0.7$, for two different variational wave functions. The first is constructed by applying a Jastrow factor on top of a BCS state, which is suitable to describe non-magnetic (i.e., spin-liquid) states. The second is obtained from a quadratic Hamiltonian that explicitly includes antiferromagnetic order, to describe a magnetically ordered phase

$$\mathcal{H}_{\rm AF} = -t \sum_{\langle i,j\rangle,\sigma} c^{\dagger}_{i,\sigma} c_{j,\sigma} + h.c. + \Delta_{\rm AF} \sum_{j} \left(e^{i\,\mathbf{Q}\cdot\mathbf{R}_j} c^{\dagger}_{j,\uparrow} c_{j,\downarrow} + e^{-i\,\mathbf{Q}\cdot\mathbf{R}_j} c^{\dagger}_{j,\downarrow} c_{j,\uparrow} \right), \tag{128}$$

where $\langle \dots \rangle$ indicates neighboring sites and $\mathbf{Q} = (\pi, \pi)$ is the pitch vector for the Néel order. In order to have the correct spin-spin correlations at large distance, we can apply a long-range spin Jastrow factor

$$\mathcal{J}_s = \exp\left[-\frac{1}{2} \sum_{i,j} u_{i,j} S^z_i S^z_j \right], \tag{129}$$

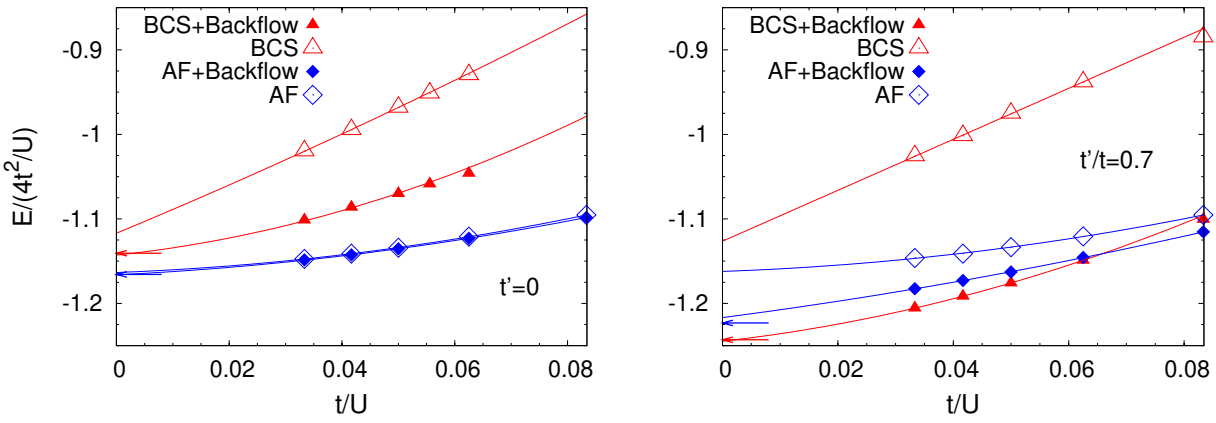

Fig. 2: *Energies per site (in units of $J = 4t^2/U$) for the two-dimensional Hubbard model at half filling, for both the unfrustrated ($t' = 0$) and frustrated ($t'/t = 0.7$) case. The cases with and without backflow correlations are reported (for the BCS state). The results for the wave function with antiferromagnetic order and no BCS pairing are also shown. Arrows indicate the energies per site for the corresponding fully-projected states in the Heisenberg model.*

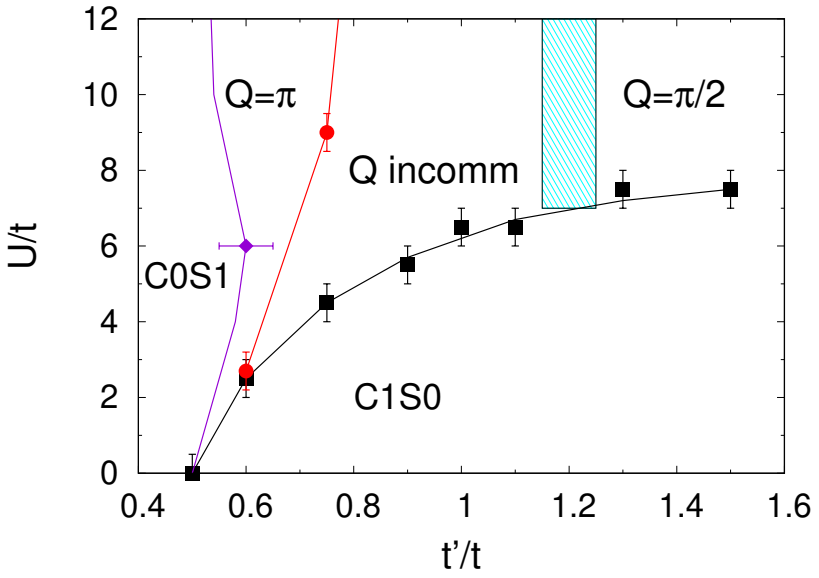

Fig. 3: *Phase diagram of the t-t' Hubbard model at half-filling with the metallic phase with gapped spin excitations (C1S0) and the insulating phase with gapless spin excitations (C0S1). The insulating phase with gapped spin excitations (C0S0) has regions with commensurate ($Q{=}\pi$ and $Q = \pi/2$) and incommensurate (Q incomm) spin-spin correlations.*

where $S^z_i{=}1/2(n_{i,\uparrow}{-}n_{i,\downarrow})$ and $u_{i,j}$ is a pseudo-potential that can be optimized for each independent distance. In analogy with the density Jastrow factor of Eq. (45), it governs spin fluctuations orthogonal to the magnetic field Δ_{AF} [37]. It is important to stress that the uncorrelated state, obtained from Eq. (128), can easily satisfy the single-occupancy constraint by taking $\Delta_{AF} \to \infty$. In this limit, it also contains the virtual hopping processes, which are generated by the kinetic term, implying that it is possible to reproduce super-exchange processes.

The results for the energies of these two variational states are reported in Fig. 2. In general, the results for the Hubbard model (at finite interaction U/t) are not smoothly connected to the ones

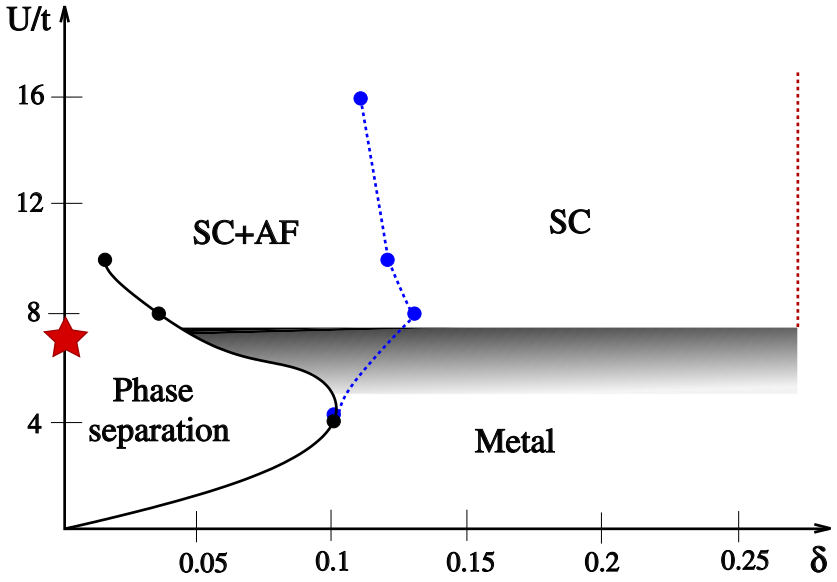

Fig. 4: *Schematic phase diagram for the two-dimensional Hubbard model by varying U/t and the filling factor $\delta = 1 - n$ (where n is the density of electrons). The red star labels the location of the hidden Mott transition U_{Mott}/t at half filling. The black line with black dots denotes the boundary of the phase-separation region, that shrinks for $U/t \gtrsim U_{Mott}/t$. The dashed blue line with blue dots marks the disappearance of $\Delta_{\rm AF}$ in the variational state. The dashed red line indicates the boundary of the region where sizable pairing correlations are detected. Finally, in the shaded gray region finite-size effects are strong and precise results cannot be obtained in the thermodynamic limit.*

obtained with the fully-projected states and the Heisenberg model, except for the unfrustrated case with $t' = 0$ and the antiferromagnetic wave function. However, thanks to backflow correlations, it is possible to obtain a correct extrapolation to the infinite-U limit. The importance of backflow correlations is extremely important in the frustrated case, where they are essential also for improving the accuracy of the antiferromagnetic wave function.

Therefore, by using Jastrow-Slater states, which include backflow correlations, it is possible to get quite accurate results in a variety of models. As an example, we report the phase diagram of the t-t' Hubbard model in one dimension at half filling, see Fig. 3 [38]. There are three main phases: one is stabilized for small values of the frustrating ratio t'/t, which is gapless in the spin sector and gapped in the charge one (denoted by C0S1); another, for small values of U/t and sufficiently large t'/t, which is gapless in the charge sector and gapped in the spin one (C1S0); and the last one, for large U/t and t'/t, which is fully gapped and dimerized (C0S0).

As a final example, we show the phase diagram that is obtained for the two-dimensional Hubbard model, as a function of U/t and the filling factor $\delta = 1 - n$, see Fig. 4 [39]. Here, the antiferromagnetic phase, which is stable at half filling (i.e., $\delta = 0$), gives rise to phase separation at small values of the interaction strength. Remarkably, these results suggest that a reminiscence of the Mott transition, hidden by the antiferromagnetic phase at half-filling, emerges after a careful analysis of the BCS pairing. This hidden Mott transition is intimately related with the change from Slater to Mott antiferromagnetism, the former being related to a Fermi surface in-

stability towards antiferromagnetic order, while the latter being connected to a super-exchange mechanism. For Coulomb interactions that are smaller than this "critical" value, the system is unstable towards phase separation and there is no strong evidence that superconductivity may emerge; by contrast, for larger values of U/t, hole doping drives the Mott antiferromagnet into a homogeneous superconducting phase, with the condensation energy gain shifting from potential to kinetic by increasing U/t.

8 Conclusions

Variational wave functions, such as Jastrow-Slater states, represents a very powerful and useful tool to investigate strongly-correlated systems on the lattice. In the recent past, there has been increasing evidence that it is possible to construct many-body states that may compete with other numerical methods, such as density-matrix renormalization group or its recent developments based upon tensor networks [40]. Future developments, in which neural networks may be used to generalize the Jastrow factor and better describe correlation effects [41], may be beneficial to solve the many-body problem.

References

[1] N. Mott, Proc. Phys. Soc. (London) **62**, 416 (1949)

[2] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)

[3] P. Lee, N. Nagaosa, and X.G. Wen, Rev. Mod. Phys. **78**, 17 (2006)

[4] J. Hubbard, Proc. Royal Soc. of London **276**, 238 (1963)

[5] M. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963)

[6] J. Kanamori, Prog. Theor. Phys. **30**, 275 (1963)

[7] H. Bethe, Z. Phys. **71**, 205 (1931)

[8] A. Georges, G. Kotliar, W. Krauth, and M. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)

[9] P. Anderson, Science **235**, 1196 (1987)

[10] F. Zhang and T. Rice, Phys. Rev. B **37**, 3759 (1988)

[11] T. Kaplan, P. Horsch, and P. Fulde, Phys. Rev. Lett. **49**, 889 (1982)

[12] H. Yokoyama and H. Shiba, J. Phys. Soc. Jpn. **59**, 3669 (1990)

[13] R. Jastrow, Phys. Rev. **98**, 1479 (1955)

[14] W. McMillan, Phys. Rev. **138**, A442 (1965)

[15] L. Reatto and G. Chester, Phys. Rev. **155**, 88 (1967)

[16] S. Sorella, Phys. Rev. B **64**, 024512 (2001)

[17] S. Sorella, Phys. Rev. B **71**, 241103 (2005)

[18] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1957)

[19] W. Hastings, Biometrika **57**, 97 (1970)

[20] H. Yokoyama and H. Shiba, J. Phys. Soc. Jpn. **56**, 1490 (1987)

[21] H. Yokoyama and H. Shiba, J. Phys. Soc. Jpn. **56**, 3582 (1987)

[22] P. Anderson, Mater. Res. Bull. **8**, 153 (1973)

[23] L. Capriotti, F. Becca, A. Parola, and S. Sorella, Phys. Rev. Lett. **87**, 097201 (2001)

[24] M. Capello, F. Becca, M. Fabrizio, S. Sorella, and E. Tosatti, Phys. Rev. Lett. **94**, 026406 (2005)

[25] M. Capello, F. Becca, S. Yunoki, and S. Sorella, Phys. Rev. B **73**, 245116 (2006)

[26] E. Wigner and F. Seitz, Phys. Rev. **46**, 509 (1934)

[27] R. Feynman, Phys. Rev. **94**, 262 (1954)

[28] R. Feynman and M. Cohen, Phys. Rev. **102**, 1189 (1956)

[29] V. Pandharipande and N. Itoh, Phys. Rev. A **8**, 2564 (1973)

[30] K. Schmidt and V. Pandharipande, Phys. Rev. B **19**, 2504 (1979)

[31] Y. Kwon, D. Ceperley, and R. Martin, Phys. Rev. B **48**, 12037 (1993)

[32] Y. Kwon, D. Ceperley, and R. Martin, Phys. Rev. B **58**, 6800 (1998)

[33] L. Tocchio, F. Becca, A. Parola, and S. Sorella, Phys. Rev. B **78**, 041101 (2008)

[34] L. Tocchio, F. Becca, and C. Gros, Phys. Rev. B **83**, 195138 (2011)

[35] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery:
Numerical recipes 3rd edition: the art of scientific computing
(Cambridge University Press, 2007)

[36] C. Gros, Ann. Phys. **189**, 53 (1989)

[37] F. Becca, M. Capone, and S. Sorella, Phys. Rev. B **62**, 12700 (2000)

[38] L. Tocchio, F. Becca, and C. Gros, Phys. Rev. B **81**, 205109 (2010)

[39] L. Tocchio, F. Becca, and S. Sorella, Phys. Rev. B **94**, 195126 (2016)

[40] J. LeBlanc, A. Antipov, F. Becca, I. Bulik, G.L. Chan, C.M. Chung, Y. Deng, M. Ferrero, T. Henderson, C. Jiménez-Hoyos, E. Kozik, X.W. Liu, A. Millis, N. Prokof'ev, M. Qin, G.E. Scuseria, H. Shi, B. Svistunov, L. Tocchio, I. Tupitsyn, S. White, S. Zhang, B.X. Zheng, Z. Zhu, and E. Gull, Phys. Rev. X **5**, 041041 (2015)

[41] G. Carleo and M. Troyer, Science **355**, 602 (2017)