Generalized Lanczos algorithm for variational quantum Monte Carlo

Sandro Sorella
Istituto Nazionale per la Fisica della Materia and International School for Advanced Studies, Via Beirut 4, 34013 Trieste, Italy
(Received 12 September 2000; published 19 June 2001)

We show that the standard Lanczos algorithm can be efficiently implemented statistically and self-consistently improved, using the stochastic reconfiguration method, which has been recently introduced to stabilize the Monte Carlo sign problem instability. With this scheme a few Lanczos steps over a given variational wave function are possible even for large size as a particular case of a more general and more accurate technique that allows to obtain lower variational energies. This method has been tested extensively for a strongly correlated model like the *t-J* model. With the standard Lanczos technique it is possible to compute any kind of correlation functions, with no particular computational effort. By using the fact that the variance $\langle H^2\rangle-\langle H\rangle^2$ is zero for an exact eigenstate, we show that the approach to the exact solution with few Lanczos iterations is indeed possible even for ~ 100 electrons for reasonably good initial wave functions. The variational stochastic reconfiguration technique presented here allows in general a many-parameter energy optimization of any computable many-body wave function, including for instance generic long-range Jastrow factors and arbitrary site-dependent orbital determinants. This scheme improves further the accuracy of the calculation, especially for long-distance correlation functions.

DOI: 10.1103/PhysRevB.64.024512 PACS number(s): 02.70.Rr, 75.10.Jm, 75.40.Mg

I. INTRODUCTION

The study of strongly correlated systems is becoming a subject of increasing interest due to the realistic possibility that in many physical materials such as high-T_c superconductors, a strong correlation between electrons may lead to an unexpected physical behavior that cannot be explained within conventional schemes, such as for instance mean field or Fermi liquid theories.

One of the most important models, which is still the subject of intense numerical studies, due to its possible relevance for high-T_c superconductivity, is the so-called *t-J* model:[1,2]

$$\hat{H}=J\sum_{\langle i,j\rangle}\left(\hat{\mathbf{S}}_i\cdot\hat{\mathbf{S}}_j-\frac{1}{4}\hat{n}_i\hat{n}_j\right)-t\sum_{\langle i,j\rangle,\sigma}\tilde{c}^{\dagger}_{i,\sigma}\tilde{c}_{j,\sigma}, \qquad (1)$$

where $\tilde{c}^{\dagger}_{i,\sigma}=\hat{c}^{\dagger}_{i,\sigma}(1-\hat{n}_{i,\bar{\sigma}})$, $\hat{n}_i=\Sigma_\sigma\hat{n}_{i,\sigma}$ is the electron density on site i, $\hat{\mathbf{S}}_i=\Sigma_{\sigma,\sigma'}\tilde{c}^{\dagger}_{i,\sigma}\tau_{\sigma,\sigma'}\tilde{c}_{i,\sigma'}$ is the spin operator, and $\tau_{\sigma,\sigma'}$ are Pauli matrices. After many years of intense numerical and theoretical efforts there is no general consensus on the properties of this simple Hamiltonian and of the related Hubbard model. In this paper the *t-J* model is studied by means of a recent numerical technique,[3] apparently very promising, to improve systematically the accuracy of a starting variational wave function, even when the Monte Carlo simulation is severely affected by the so-called ''sign problem'' instability. We will show in this paper that in a particular case this technique allows us to apply very efficiently the well-known Lanczos technique, by means of quantum Monte Carlo simulations. This allows us to obtain accurate variational wave functions even on rather large sizes, as first pointed out by Heeb and Rice.[4]

The Green function Monte Carlo with stochastic reconfiguration (GFMCSR) was introduced recently[3] for—so-called—projection techniques, aiming to determine—as accurately as possible—the ground-state (g.s.) wave function ψ_0 of a given Hamiltonian H.

In the statistical approach the electron g.s. wave function is sampled over a set of configurations $\{|x\rangle\}$, denoting spins and electron positions, and belonging to a complete basis with large or even infinite dimension. The ground-state component ψ_0 of a given trial state ψ_G—henceforth assumed to be nonzero on each configuration x—is filtered out by applying to it some projection operator G^n for a large number n of power iterations. The matrix G may be given by $G=e^{-H\Delta t}$ for short imaginary time Δt, as is the case for the conventional diffusion Monte Carlo simulation for continuous models, or the one that will be considered in the following sections,

$$G=(\Lambda I-H)^{k_p}, \qquad (2)$$

for lattice models, where a suitably large constant shift Λ allows convergence to the ground state ($\Lambda=0$ is used for the *t-J* model). The integer k_p, determining the number of powers of the Hamiltonian in the Green function, may be in principle larger than 1,[5] but in the following we avoid this complication and we consider k_p fixed at its minimum value: $k_p=1$.

In order to perform stable simulations with a large signal-to-noise ratio even for large n the many-body propagation $\psi_n\rightarrow G^n|\psi_G\rangle$ needs to be stabilized at each iteration, using an approximation that can be efficiently implemented statistically by means of the stochastic reconfiguration (SR) scheme.[5]

The essential step in the SR is to replace the many-body state $\psi_{n+1}(x)=G\psi_n$, obtained by applying to ψ_n the exact Green function G, with the approximate state $\psi'_{n+1}(x)$, determined by the following conditions. A given set of $p+1$ operators $\{O^k\}$, not restricted to be Hermitian operator, satisfies the equalities

$$\langle \psi_G | O^k | \psi'_{n+1} \rangle = \langle \psi_G | O^k | \psi_{n+1} \rangle \quad \text{for} \quad k=0,\ldots,p, \quad (3)$$

with O^0 for $k=0$ being the identity here for simpler and more compact notation, $\psi'_{n+1}=r_x\psi^f(x)$, ψ^f being a reference state known exactly (e.g., the standard variational approach) or statistically (e.g., the fixed node approximate state) on each configuration x, and

$$r_x=\sum_{k=0}^{p} \alpha_k O_x^k, \quad (4)$$

where

$$O_x^k=\frac{\langle \psi_G | O^k | x \rangle}{\langle \psi_G | x \rangle}, \quad (5)$$

and the constants α_k are determined by the conditions (3):

$$\alpha_i=\sum_k s_{i,k}^{-1} \langle \psi_G | O^k | \psi_{n+1} \rangle, \quad (6)$$

where the covariance matrix is given by

$$s_{i,k}=\sum_x \psi_G(x)\psi^f(x)O_x^i O_x^k. \quad (7)$$

At equilibrium, for large n ψ'_n converges to a well-defined many-body state ψ_{SR}, up to a trivial normalization constant, which maybe also infinite but irrelevant for physical expectation values. In this limit therefore the conditions (3) determine the strong constraints on the many-body state ψ_{SR}, namely,

$$\frac{\langle \psi_G | O^k | \psi_{SR} \rangle}{\langle \psi_G | \psi_{SR} \rangle}=\frac{\langle \psi_G | O^k G | \psi_{SR} \rangle}{\langle \psi_G | G | \psi_{SR} \rangle} \quad \text{for} \quad k=0,\ldots,p. \quad (8)$$

These constraints are exactly fulfilled by an exact eigenstate of the Green function G, so that they represent physically relevant conditions restricting the possible values of the correlations functions for a reasonable approximation ψ_{SR} of the ground state of H (an exact eigenstate of G). As will be shown later on in Sec. V these constraints (8) coincide in some special cases with the Euler equations of minimum energy, but this is not generally true. The SR conditions represent therefore more restrictive criteria to judge the quality of a given approximate state. They are directly related to correlation functions and not necessarily to the energy alone. This property is particularly important, since for the same strongly correlated Hamiltonian it is often possible to find several variational states, with very accurate energy expectation values, but with completely different correlation functions.[6] Thus the SR conditions (8), which are clearly not limited to a statistical approach, represent a useful tool to alleviate the above very important difficulty of approximate variational techniques.

Going back to the more formal derivation, the SR can be considered a projection P_{SR} of the exactly propagated wave function $\psi_{n+1}=G\psi_n$ onto a subspace spanned by the states $|\psi_f^k\rangle$, defined by $\langle x|\psi_f^k\rangle=O_x^k\psi^f(x)$, for $k=0,\ldots,p$. Notice also that *only* when $\psi^f=\psi_G$, $\psi_f^k=(O^k)^\dagger|\psi_G\rangle$; namely, ψ_f^k is obtained by applying the operator $(O^k)^\dagger$ to the state ψ_G.

Solving the linear system determined by the SR conditions (3) we obtain a closed expression for the linear operator P_{SR}, which explicitly depends on the state ψ_G and the reference state ψ^f:

$$P_{SR}=\sum_{i,j} s_{i,j}^{-1} |\psi_f^i\rangle\langle\psi_G|O^j. \quad (9)$$

The operator P_{SR} is not a true projection operator. Though it satisfies the simple requirement (i) $P_{SR}^2=P_{SR}$, (ii) $P_{SR}^\dagger=P_{SR}$ is not generally satisfied. A way to implement statistically this ''pseudoprojection'' operator in the large number of walker limit was discussed in Ref. 5.

After each reconfiguration the state ψ_n is replaced by $\psi'_n=P_{SR}\psi_n$, but also the reference state ψ^f is changed. In fact in the original formulation[3,5] also the reference state explicitly depends on n and is updated after each SR:

$$\psi_n^f(x) \to \mathrm{sgn}[\psi_G(x)]|\psi'_n(x)|. \quad (10)$$

The reason for the choice (10) is to optimize the reference state and obtain more accurate results, as explained in the forthcoming paragraphs.

With the restriction of a stable simulation the signs of the reference wave function have to be fixed; otherwise the average sign will drop exponentially to zero in the simulation, Therefore, in Eq. (10), ψ^f has been restricted to have the same signs (or the same nodes) of the guiding wave function ψ_G. On the other hand, the amplitudes of the reference wave function ψ^f can be considerably improved from our best variational guess ψ_G, since during the exact dynamic the state ψ_n gets closer to the ground-state wave function. The optimal choice for the amplitudes has naturally led to the definition (10).

This technique, with the choice (10), has been shown to be remarkably accurate for frustrated spin or boson systems, allowing one in many test cases to obtain essentially exact results within statistical errors.[7,8] However, for fermion problems the situation is much different. Though this technique allows a significant improvement in the energy and correlation function calculations,[9,10] the bias remains still sizable and difficult to eliminate completely by increasing the number of correlation functions used in the SR technique. Due to the antisymmetry of the fermion many-body wave function, it appears that the nodes in this case play a much more important role.

It is instructive to consider the case of continuous models. In this case, for fermion systems, nodes have to appear in the many-body wave function just by antisymmetry considerations. On the other hand, symmetry alone does not restrict the nodal surface [the locus where the wave function $\psi_0(x)$ vanishes], implying that correlation effects can significantly change the nodal surface. In this case it may be useless or irrelevant to improve the amplitudes without changing the nodes in the reference wave function (10).

For fermion systems, due to the above difficulty in determining a nodal surface which is weakly modified by correlation effects, it appears at the moment difficult to avoid a sizable ''nodal'' error in energies and correlation functions.

Therefore in this case it is extremely important that any approximation is at least controlled by the variational principle. In this way an approach different from the one proposed before[3] can be used. The first step to obtain a rigorous variational method is to consider the reference wave function ψ^f fixed to the variational or the fixed node state.[11,12]

II. FIXED REFERENCE DYNAMIC

In the GFMC propagation a large number M of walkers is used; the jth walker is characterized by two weights w_j^f, w_j, acting on a configuration state x_j, e.g., a state with definite electron positions and spins. The reference weights w_j^f sample statistically the reference state $\psi_n^f(x)$ whereas the weights w_j refers to the state $\psi_n(x)$ propagated by the exact Green function $\psi_n = G\psi_{n-1}$. More precisely, taking into account the importance sampling transformation[5]

$$\langle\langle w_j^f \delta_{x,x_j}\rangle\rangle = \psi_n^f(x)\psi_G(x), \tag{11}$$

$$\langle\langle w_j \delta_{x,x_j}\rangle\rangle = \psi_n(x)\psi_G(x), \tag{12}$$

where the brackets $\langle\langle\cdots\rangle\rangle$ indicate both the statistical average and the one over the number of walkers, at a given Markov iteration n. The two wave functions ψ_n^f and ψ_n are propagated using the statistical approach. For the first state a reference Green function $G_{x',x}^f$ with all positive matrix elements is used, whereas the latter one is propagated by means of the exact Green function (2) which is related to the reference one by a simple relation $G_{x',x} = s_{x',x}G_{x',x}^f$ with finite and known matrix elements $s_{x',x}$:

$$\psi_{n+1}(x')\psi_G(x') = \sum_x G_{x',x}\psi_n(x)\psi_G(x), \tag{13}$$

$$\psi_{n+1}^f(x')\psi_G(x') = \sum_x G_{x',x}^f\psi_n^f(x)\psi_G(x). \tag{14}$$

The reference Green function $G_{x',x}^f = p_{x',x}b_x$ is written in terms of a stochastic matrix $p_{x',x}$ ($p_{x',x} \geqslant 0$ and $\Sigma_{x'}p_{x',x} = 1$) times an x dependent normalization factor $b_x = \Sigma_{x'}G_{x',x}^f$, so that a statistical implementation of the iteration (13) is possible. Namely, given the configuration x_j of the jth walker a new configuration x_j' is selected statistically with probability p_{x_j',x_j} (notice that $\Sigma_{x'}p_{x_j',x_j} = 1$ by definition):

$$x_j' = x' \text{ with probability } p_{x',x_j}. \tag{15}$$

Then the reference weight

$$w_j^f \to b_{x_j}w_j^f \tag{16}$$

is scaled by the normalization factor b_{x_j}, whereas the weight related to the exact Green function G,

$$w_j \to s_{x_j',x_j}b_{x_j}w_j, \tag{17}$$

is further multiplied by the s_{x_j',x_j} matrix element.

According to the above Markov iteration, defined by Eqs. (15)–(17), all the walkers propagate independently of each other. The reference weights w^f remain positive, whereas w_j, related to the exact propagation, accumulate many sign changes, leading, for large n, to an exponential increase of the signal-to-noise ratio. This is in fact determined by the corresponding exponential drop of the average walker sign $\langle s\rangle = \Sigma_j w_j/\Sigma|w_j|$ (the infamous ''sign problem''). The stochastic reconfiguration allows one to alleviate this problem, by implementing statistically the operator P_{SR} described in the previous section. In practice the weights w_j are replaced by approximate weights p_j, but with a much larger average sign $\langle s\rangle$. The weights p_j sample statistically $\psi_n'(x)\psi_G(x)$ and are determined by solving the corresponding linear system (3). This is done at a given Markov iteration by averaging Eq. (3) *only* over the walker samples. This means that there exists a bias due to the statistical uncertainty of the quantum averages in Eq. (3). This bias, however, can be controlled efficiently[5] because for a large number of walkers it vanishes as $1/M$. In this limit r_x in Eq. (4) depends only on the configuration x as the statistical fluctuations of the constants α_k can be neglected for $M \gg p$.

Another method to control much more efficiently the statistical fluctuations of the constants $\{\alpha_k\}$, without using a too large number of walkers, is to perform the SR scheme only after applying, at each iteration n, a large number L_b of statistical steps defined in Eqs. (15)–(17) In this way the statistical averages in Eq. (3) have only small fluctuations $\propto 1/\sqrt{L_b}$, which can be neglected in the limit of large bin length L_b, even when a small number of walkers is used. In each bin we are considering only the iteration $n \to n+1$ of the power method when the parameters α_k of the wave function $P_{SR}\psi_n$ are known and computed in the previous iteration. Thus $\{w_j^f\}$ evolve statistically according to G^f [Eqs. (15) and (16)], and new configurations $\{x_j\}$ are generated for the statistical sampling of $\psi_G(x)$ $\psi^f(x) = \langle\langle w_j^f \delta_{x,x_j}\rangle\rangle$, whereas, inside the bin, the weights $\{w_j\}$ are always reset to $w_j = r_{x_j}w_j^f$—i.e., with the same configurations and a simple scaling of the weights it is possible to sample the propagated state $P_{SR}\psi_n$ in all the L_b statistical steps. Then at each step inside the bin the exact Green function G is applied statistically [Eq. (17)] in order to sample $\psi_{n+1} = GP_{SR}\psi_n$, which is required to calculate the right-hand side (RHS) of the SR conditions (3). At the end of the bin new α_k can be computed by solving the linear system (3). It is understood that even inside the bin a branching scheme at fixed number of walkers[13] and with a fixed number of correcting factors can be used to improve importance sampling. Whenever the length of the bin is so large that the statistical fluctuations can be neglected, the algorithm is deterministic inside the statistical uncertainties and becomes much more efficient compared to the original scheme,[3,5] where the SR conditions were applied at each step ($L_b = 1$). We have found instead that the most efficient scheme is to change these constants α_k only when the SR conditions (3) are not satisfied within statistical errors (e.g., they are off by more than three error bars), by increasing systematically the bin length at each iteration. This scheme allows one also to eliminate the bias

due to the finite number of walkers M, which was rather sizable in the original formulation.[5]

After the SR, say, at the iteration n_0, in order to continue the power method iteration for $n>n_0$, the new approximate state $\psi'_{n_0}=P_{SR}\psi_{n_0}$ replaces ψ_{n_0} but the reference state ψ^f is still arbitrary. Instead of changing the reference weights with the choice[3,5] $w_j^f=|p_j|$ [implying Eq. (10) in the large number of walker limits] it is instead possible to remain with the same reference state $\psi^f_{n_0}$, without changing it in the statistical sense. This is obtained by the following simple scheme. After the SR new configurations $x'_i=x_{j(i)}$ are selected among the old ones x_j according to the probability $\Pi_j=|p_j|/\Sigma_k|p_k|$. This scheme naturally defines the random index table $j(i)$,[13,5] used to improve importance sampling—as in the branching scheme for the standard diffusion Monte Carlo simulation[14]—and allows one to continue the simulation more efficiently with equally weighted walkers $|w'_i|\simeq$const. In fact in order to sample statistically the states ψ'_n and ψ^f with corresponding new weights w'_i and $w_i^{f\prime}$,

$$\psi'_n(x)\psi_G(x)=\langle\langle p_i\delta_{x,x_i}\rangle\rangle=\langle\langle w'_i\,\delta_{x,x'_i}\rangle\rangle,$$

$$\psi^f_n(x)\psi_G(x)=\langle\langle w_i^f\delta_{x,x_i}\rangle\rangle=\langle\langle w_i^{f\prime}\,\delta_{x,x'_i}\rangle\rangle, \tag{18}$$

it is enough to use the so-called reweighting method, which makes the above equations *exact* in the statistical sense:

$$w'_i=\bar{w}\,\mathrm{sgn}\;p_{j(i)}, \tag{19}$$

$$w_i^{f\prime}=\bar{w}w^f_{j(i)}/|p_{j(i)}|=\frac{|w'_i|}{|r_{x'_i}|}, \tag{20}$$

where $\bar{w}=(1/M)\Sigma_j|p_j|$ is a constant common to all the walker weights and $\mathrm{sgn}\,a=\pm1$ represents the sign of the number a. The correcting factor $\bar{w}$ is taken into account only for a finite number L of past iterations, starting, e.g., from $n-L+1$; otherwise, the weights of the walkers may increase or decrease exponentially, leading to a divergent variance. This introduces a systematic bias, which vanishes, however, exponentially in L and decreases as $1/M$.[5] In practice for a large number of walkers it is enough to consider only few (or even none) ''correcting'' factors in the statistical averages, as common practice in Green function Monte Carlo simulations.[14] From the reweighting method it is also clear that the choice of the probability function Π_j is not restricted to be proportional to p_j. In particular we have found it more convenient to use the weights corresponding to $G\psi_n$ determined after applying Eq. (17), so that the choice $\Pi_j=|w_j|/\Sigma|w_k|$ further improves the importance sampling, with a minor change in Eqs. (19) and (20):

$$w'_i=\bar{w}\frac{p_{j(i)}}{|w_{j(i)}|},$$

$$w_i^{f\prime}=\frac{|w'_i|}{|r_{x'_i}|},$$

with $\bar{w}=\Sigma|w_j|/M$.

Using the previous scheme the reference state equilibrates necessarily to the largest right eigenstate of the reference matrix G^f. At equilibrium the state ψ_{SR} has therefore a very well-compact and clear definition. It represents the maximum right eigenstate of the matrix

$$P_{SR}(\Lambda I-H)P_{SR}, \tag{21}$$

with given ψ^f in the definition of P_{SR} Eq. (9).

The method is therefore rigorously variational provided the pseudoprojector P_{SR} is a true projector operator, namely, for $\psi^f=\psi_G$ when $P_{SR}^\dagger=P_{SR}$, as implied by Eq. (9). In this case in fact by standard linear algebra, the maximum right eigenstate of the operator (21) is the best variational state of H belonging to the subspace defined by the projector P_{SR} (see Appendix A).

A. Variational energy when $P_{SR}=P_{SR}^\dagger$

In the original formulation of the SR the reference Green function was defined with a slight generalization of the fixed node Green function.[3] The Green function proposed by Hellberg and Manousakis[15] is instead more appropriate in this context and much more convenient from the practical point of view of reducing statistical fluctuations:

$$G^f_{x',x}=\frac{1}{z_{x'}}|\Lambda\,\delta_{x',x}-\psi_G(x')H_{x',x}/\psi_G(x)|. \tag{22}$$

where $z_x=\Sigma_{x'}|\Lambda\,\delta_{x',x}-\psi_G(x')H_{x',x}/\psi_G(x)|$, $|a|$ meaning the absolute value of the number a. It is simple to show that, by applying the power method with the above Green function, convergence is reached when the maximum right eigenvector $\psi_G(x)^2$ is filtered out:

$$\sum_x G^f_{x',x}\psi_G(x)^2=\psi_G(x')^2. \tag{23}$$

Thus the above Green function can be used to generate statistically configurations [see Eqs. (15) and (16)] distributed according to $\psi_G(x)^2$ with a stochastic matrix $p_{x',x}=G^f_{x',x}z_{x'}/z_x$.

The advantage of using the reference Green function (22) is evident when we consider its *very simple* relation with the exact Green function, namely,

$$s_{x',x}=G_{x',x}/G^f_{x',x}=\pm z_{x'}, \tag{24}$$

where the sign $\pm$ is given by the sign of the Green function matrix element $\Lambda\,\delta_{x',x}-\psi_G(x')H_{x',x}/\psi_G(x)$, and depends of course on x and x'. Using the fixed reference algorithm in Eq. (20) $\psi^f=\psi_G$, and the operator P_{SR} represents in this special case a true projector one $P_{SR}^\dagger=P_{SR}$. Thus in this case the method is rigorously variational as pointed out in the last part of the previous section.

We notice also an important property of this method. If in the SR conditions (3) only operators defined by powers of the Hamiltonian $O^k=H^k$ are used, the projector P_{SR} acts on the same Krilov basis (spanned by $H^k|\psi_G\rangle$, $k=0,\ldots,p$) of

the well-known Lanczos algorithm. Thus $(P_{SR}GP_{SR})^n|\psi_G\rangle$ filters out the lowest-energy variational state in this Krilov basis, i.e., by definition the state obtained by applying p Lanczos iterations to ψ_G. We recover in particular a known property of the Lanczos algorithm, valid also for the SR method: the method is exact if p equals the dimension of the Hilbert space.

Due to the equivalence of the Lanczos algorithm with the SR technique, it is clear why, with the latter technique, it is possible to obtain a rather good approximate ground state with p reasonably small.[4,7–10] In fact the convergence of the Lanczos algorithm is at least exponential in p.[16]

We have therefore derived that the Lanczos algorithm can be implemented statistically using the SR method. This allows us to perform easily two Lanczos iterations on a given variational wave function for fairly large size systems. Furthermore, the SR method allows us to put several correlation functions in the Eq. (3). Since the method is strictly variational, the variational energy has necessarily to decrease by increasing the mentioned number of correlation functions.

B. Improving the variational energy

Following Ref. 17, by applying a finite number of exact Green function iterations to the wave function ψ_{SR}, the corresponding quantum average

$$E_{SR}^k=\frac{\langle\psi_{SR}|G^kHG^k|\psi_{SR}\rangle}{\langle\psi_{SR}|G^{2k}|\psi_{SR}\rangle} \tag{25}$$

remains obviously variational for any k.

Taking into account $2k$ statistical factors $s_{x',x}$, the above quantum averages can be statistically evaluated with the same Markov chain for which E_{SR} $(k=0)$ is computed.

The sign problem can be controlled for not too large k and systematically improved variational energies can be obtained compared to the $k=0$ result. However, experience has shown that it is very difficult to have significant improvement over the $k=0$ result for a large system size.

III. VARIATIONAL ENERGY WHEN $P_{SR}\neq P_{SR}^\dagger$

We have seen that the method is rigorously variational once the reference weights are changed according to Eq. (20). However, as we have explained in the Introduction, a better choice is to continue after the SR with $w^{f'}=|w'|$, the rationale for this choice being that the wave function $\psi'=P_{SR}\psi$ has much better amplitudes than the variational wave function ψ_G. This allows us to improve self-consistently the reference wave function in order to be as close as possible to the true ground state. In this way, however, $\psi^f\neq\psi_G$ and the method is no longer variational in the sense that the SR state defined by the right eigenstate with maximum modulus $|\Lambda-E_{SR}|$ eigenvalue of the matrix,

$$P_{SR}(\Lambda I-H)P_{SR}|\psi_{SR}\rangle=(\Lambda-E_{SR})|\psi_{SR}\rangle, \tag{26}$$

is no longer the lowest-energy one in the basis defined by $|\psi_f^k\rangle$, $k=0,1,\ldots,p$, and E_{SR} does not necessarily bound the ground-state energy.

A compromise between the two methods is to introduce a parameter r that interpolates smoothly the two limits. This parameter enters in the reweighting relation (20) in a very simple manner:

$$w_i^{f'}==\frac{|w_i'|}{|r_{x_i'}|^{1-r}}. \tag{27}$$

Notice that, when we release the fixed reference dynamic for $r\neq0$ and when for large n the constants $\{\alpha_k\}$ converge to nonzero values, the reference wave function ψ^f is not given by averaging the configurations with the weights w^f since the relation $\psi_{SR}=P_{SR}\psi_n(x)=r_x\psi^f(x)$ is implied by the definition of the projector P_{SR} in Eq. (9). In fact, by using Eq. (27) and that $\langle\langle w_i'\,\delta_{x,x_i}\rangle\rangle=\psi_{SR}(x)\psi_G(x)=r_x\psi^f(x)\psi_G(x)$, it is simple to derive that

$$\begin{aligned}\langle\langle w_i^{f'}\,\delta_{x,x_i}\rangle\rangle&=|r_x|^{r-1}\operatorname{sgn}r_x\langle\langle w_i'\,\delta_{x,x_i}\rangle\rangle\\&=|r_x|^{r-1}\operatorname{sgn}r_x\psi_{SR}(x)\,\psi_G(x)\\&=|r_x|^r\psi^f(x)\,\psi_G(x).\end{aligned} \tag{28}$$

For $r=0$ the method is rigorously variational, in the sense that the parameters α_k appearing in the definition of r_x Eq. (4), are the ones that minimize the energy expectation value, once the SR conditions (3) are satisfied for a large number n of iterations. It is reasonable to expect that such a property remains valid even for $r\ll1$ when $P_{SR}^\dagger\simeq P_{SR}$. We have empirically verified that for small r, in particular for the one that minimizes the energy expectation value, this property is indeed verified. On the other hand, for $r\neq0$ the expectation value of the energy, which can be computed by using the forward walking technique described in Appendix B, remains obviously a rigorous upper bound of the energy.

For $r\neq0$ we assume that the infinite number of walkers or large bin L_b is taken so that the parameters α_k can be considered constants for large n. In this case, if we take into account the reweighting (27), the reference wave function ψ^f is obtained as the right eigenvector $\psi_R(x)=\langle\langle w_i^{f'}\,\delta_{x,x_i}\rangle\rangle$ with maximum eigenvalue of the renormalized reference Green function $\overline{G^f}$:

$$\overline{G_{x',x}^f}=|r_{x'}|^rG_{x',x}^f, \tag{29}$$

namely, as $\psi_f(x)=|r_x|^{-r}\psi_R(x)/\psi_G(x)$

Then the SR state $\psi_{SR}(x)=r_x\psi^f(x)$ is simply given in terms of this right eigenvector:

$$\psi_{SR}(x)=R_x\psi_R(x), \tag{30}$$

with

$$R_x=|r_x|^{1-r}/\psi_G(x)\ \operatorname{sgn}r_x. \tag{31}$$

Thus even when $r\neq0$ the SR state can be uniquely determined. It is also clear that, since r_x is not necessarily positive or negative, the nodes can be changed and improved with respect to the nodes of the initial guess ψ_G, both for $r=0$ with the standard Lanczos algorithm and for $r>0$. If the

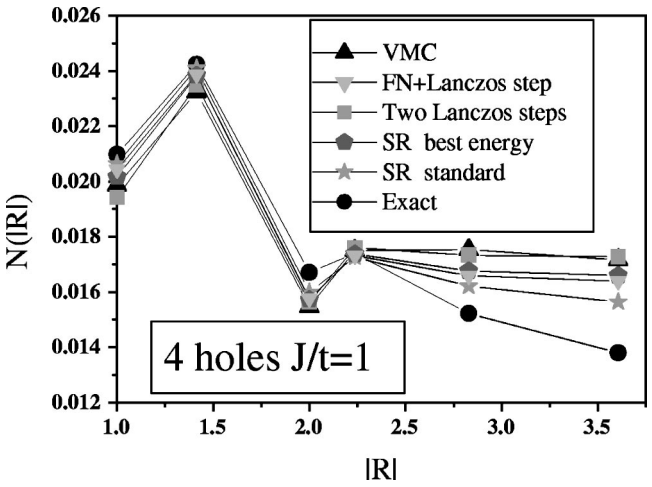

FIG. 1. Hole-hole correlations for the 26-site *t-J* cluster for various methods. ''SR standard'' indicates the original GFMCSR implementation ($r=1$ here) (Refs. 3 and 5), whereas ''SR best energy'' indicates the optimal variational SR wave function obtained with $r=0.25$. The ''VMC'' is the lowest-energy Jastrow-Slater variational singlet wave function as discussed in Sec. V, FN +Lanczos step the fixed node over the one Lanczos step wave function.

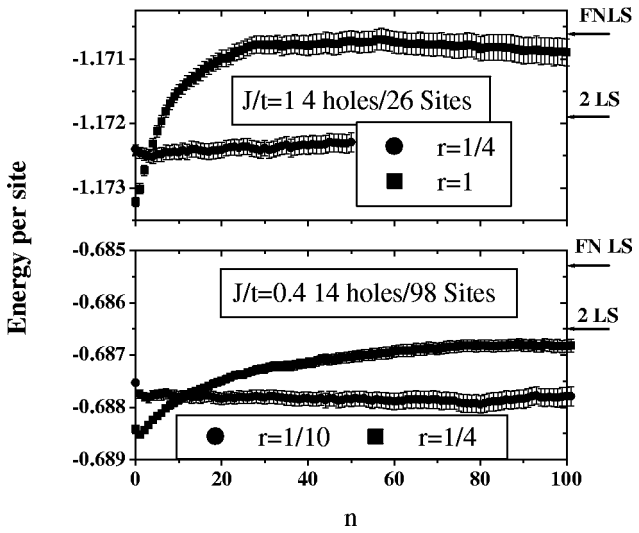

FIG. 2. Energy per site as a function of the forward iteration n as described in Appendix B for the 26-sites–4-hole case and the 98-site–14-hole case. The value of the variational parameter r is also indicated.

nodes of ψ_G are exact and the Hamiltonian is not frustrated, it is also possible to show that $r=1$ and $\Lambda\to\infty$ provide the exact result, with r_x positive definite.

Is it possible to compute correlation functions over ψ_{SR}?

As is shown in Appendix B the answer is yes, and not only for correlation functions diagonal in the basis x, as in the standard forward walking technique,[13] but also for all the ones with off-diagonal elements contained in the nonzero Hamiltonian matrix elements. In particular it is possible to compute

$$\langle \psi_{SR}|H|\psi_{SR}\rangle \geqslant E_0,$$

i.e., the expectation value over the state defined by the SR conditions. This estimate is obviously variational and can be further improved by applying a finite number of exact Green function powers to the right and to the left of the Hamiltonian, as in the power Lanczos algorithm,[17] with the difference that in this case Eq. (25) has to be evaluated with the ''forward walking technique,'' as described in Appendix B. In Fig. 2 we plot the evolution of the expectation value of the energy over the state ψ_{SR} as a function of the number of iterations, n, required by the forward walking to filter out from ψ_G its component over ψ_{SR}, leading to a true variational energy estimate. We see that for $r=1$, within the original SR technique,[3,5] the energy expectation value can be much higher than the corresponding ''mixed average'' estimate $(n=0)$.

This behavior can be understood in the following way: for $r=1$ and for $\Lambda\to\infty$ there is no way to improve the sign of the wave function over ψ_G because $r_x\to 1$ [G and G^f tend to the identity up to a constant and so the correction r_x to $\psi^f \simeq \psi$ for $r=1$ has to become unity up to $O(1/\Lambda)$], whereas for $r=0$ the Lanczos algorithm, which is Λ independent, certainly modifies and improves the nodes. For fermion systems therefore it appears important to work with small r because the nodes of the wave function play a particular important role in determining a good variational energy.

A different behavior is seen for correlation functions. For large $J/t=1$ when a four-particle bound state is likely to be formed, our Jastrow-Slater wave function is not appropriate enough, and the large distance behavior is not exactly reproduced (see Fig. 1) even when we apply to it a couple of Lanczos iterations, which, remarkably, provide a very accurate variational energy (see Table I).

That the qualitative behavior of this correlation function is different from the variational starting wave function can be understood only when the algorithm with $r>0$ or the fixed node (FN) (which is worse in energy) are used. Especially successful is the original technique $r=1$, which improves by a factor of 2 the important long-range behavior of $N(|R|)$, clearly displaying the features of a genuine bound state, by a decaying probability to have holes at larger and larger distances. For correlation functions diagonal in the chosen basis the nodes do not play any role and $r=1$ or the FN itself is likely to provide much better correlation functions. However, from the previous argument about the impossibility to correct the nodes for $r=1$ and $G,G^f\sim I$ we expect indeed that for large sizes the Green function G tends to the identity, either because $\Lambda\propto L$, as required by the power method to converge, or because the gap to the first excited state decreases and a power iteration $\psi_{n+1}=G\psi_n$ is less and less effective for changing the wave function. Thus r has to scale to zero for large sizes if we do not want to spoil too much the variational expectation value of the energy (see Fig. 2). It is remarkable that the gain in variational energy is larger and larger when the size is increased. Thus the $r>0$ technique seems to overcome, at least partially, a serious limitation of the Lanczos algorithm, namely, that in the thermodynamic limit the energy per site cannot be improved by a technique which is not size consistent. The gain in energy with $r>0$

TABLE I. Energy per site in the *t-J* model for various variational methods. VMC is the variational method, obtained by adjusting in the Jastrow and determinant part of the wave function all possible parameters being compatible with the symmetries of the finite lattice (see Sec. V). VMC+LS is obtained by applying to this wave function a Lanczos step, VMC+FN is the lattice fixed node approach (Refs. 12 and 11), VMC+2 LS indicates the two-Lanczos-step variational wave function, VMC+LS+FN is the fixed node over the VMC+LS wave function, and the ''Best *r*'' is $\langle \psi_{SR}|H|\psi_{SR}\rangle$, computed by ''forward walking'' as described in Appendix B, by optimizing the parameter *r*. The exact energy values for the largest size were estimated by the variance extrapolation. On the 98 sites, the FNLS computation takes 10 h CPU time on a Pentium-II 400 MHz, whereas the VMC+2LS wave function takes about 40 hours with $M=500$ walkers for a statistical accuracy of $10^{-4}t$ on the energy per site, the ''best SR'' another factor of 8 more due to the forward walking. The computation of diagonal correlation functions instead takes a similar amount of time for all methods; thus it is safer to compute them with the best variational method. Error bars are indicated in brackets.

N	L	J/t	VMC	VMC+LS	VMC+FN	VMC+2 LS	VMC+LS +FN	Best SR	Best r	Exact
22	26	0.3	-0.6138(1)	-0.6332(1)	-0.6277(1)	-0.6381(1)	-0.6371(1)	-0.6387(1)	0.375	-0.64262
22	26	0.5	-0.7647(1)	-0.7812(1)	-0.7759(1)	-0.7852(1)	-0.7841(1)	-0.7855(1)	0.25	-0.78812
22	26	1.0	-1.1476(1)	-1.1672(1)	-1.1608(1)	-1.1719(1)	-1.1706(1)	-1.1724(1)	0.25	-1.17493
30	32	0.3	-0.4543(1)	-0.4628(1)	-0.4611(1)	-0.46522(3)	-0.46524(3)	-0.4661(1)	0.375	-0.470175
84	98	0.4	-0.6653(1)	-0.6807(1)	-0.6777(1)	-0.6865(1)	-0.68530(5)	-0.6879(2)	0.1	-0.692(1)
50	98	0.4	-0.9656(1)	-0.9832(1)	-0.98225(5)	-0.9886(1)	-0.98781(6)	-0.9901(2)	0.1	-0.9920(5)

can instead be size consistent, since, as shown in Appendix B, $r>0$ corresponds to modifying the reference Green function $G^f \to \overline{G}^f$, similarly to what the fixed node algorithm—which is size consistent for size consistent ψ_G—does.

IV. NUMERICAL IMPLEMENTATION

In order to put efficiently a finite number p of Hamiltonian powers in the SR scheme it is by far more convenient to use an importance sampling strategy (see, e.g., Appendix C), by using information of the previous $p-1$ Lanczos iterations. A guiding wave function corresponding to the previous $(p-1)$ approximation, $\psi'_G(x)=\langle x|\Sigma_{k=0}^{p-1}\alpha_k H^k|\psi_G\rangle = r_{p-1}(x)\psi_G(x)$, can be used.[18] With this new guiding function the powers of the Hamiltonian can be put in the SR conditions by computing the corresponding mixed average estimators:

$$O_x^k=\frac{\langle \psi_G|H^k|x\rangle}{\langle \psi'_G|x\rangle} \tag{32}$$

for $k=0,\ldots,p$.

The advantage of using the SR scheme is clear even when we restrict this method only to evaluation of the first few Lanczos iterations. In order to perform p Lanczos iterations, it is enough to compute only p Hamiltonian powers on a given configuration x. In the conventional variational method it is always necessary to compute the expectation value of the Hamiltonian amounting to $2p+1$ powers of the Hamiltonian, leading to a much more demanding numerical effort. It is also important to emphasize that within this technique the parameters α_k defining the SR state are given at the end of the SR simulation $n\to\infty$. Once the $\{\alpha_k\}$ are determined it is then convenient to compute correlation functions with fixed constants $\{\alpha_k\}$ by performing statistical averages over a large bin L_b without applying the SR conditions (3) as discussed in Sec. II. This method significantly improves the statistical fluctuations of the quantum averages over the variational state $P_{SR}|\psi_G\rangle$.

V. ENERGY OPTIMIZATION FOR A MANY-PARAMETER VARIATIONAL WAVE FUNCTION

The most important advantage of the SR technique is that many variational parameters can be handled at little expense of solving a linear system (6) of corresponding size. The few-Lanczos-step technique, as we have discussed before, is determined only by few coefficients due to the difficulty of the large size to compute many powers of the Hamiltonian on a given configuration x, which is required for evaluation of the corresponding mixed estimator $O_x^k = \langle \psi_G|H^k|x\rangle/\langle \psi_G|x\rangle$.

Clearly the method discussed in Sec. II is not limited only to the Hamiltonian momenta correlation functions, but remains variational for arbitrary ones. In particular many kinds of correlation functions can be thought to be a renormalization of the guiding wave function ψ_G, allowing a powerful multiparameter energy optimization similar to Ref. 19, where the variance was instead minimized. In the present section we assume for simplicity that *all* the correlation functions are used for the purpose of optimizing the variational wave function ψ_G, and we restrict our very general analysis to variational wave functions of the Jastrow-Slater form for a strongly correlated system such as the *t-J* model defined in Eq. (1).

A. Variational wave function

The Jastrow-Slater variational wave function can be generally written as

$$|\psi_G\rangle=\hat{J}|S\rangle, \tag{33}$$

where $|S\rangle$ is a determinant wave function that can be obtained as an exact ground state of a generic one-body Hamiltonian of the bilinear form[20] $H_{1\text{-}body}\simeq c^\dagger c\cdots+c^\dagger c^\dagger\cdots+\text{H.c.}$ whereas the Jastrow factor

$$\hat{J}=\exp\left(\sum_{i,j} v(i,j)n_i n_j\right)\exp\left(\sum_{i,j} v^z(i,j)\sigma_i^z\sigma_j^z\right) \tag{34}$$

introduces arbitrary Gaussian correlations between local charges $n_i=\Sigma_\sigma c^\dagger_{i,\sigma}.c_{i,\sigma}$ and local spins along the z axis: $\sigma^z_i=c^\dagger_{i,\uparrow}c_{i,\uparrow}-c^\dagger_{i,\downarrow}c_{i,\downarrow}$. Both operators are defined on each configuration $\{x\}$ of the chosen Hilbert space. The reason for the exponential form in the Jastrow wave function comes from size-consistency considerations, implying that for macroscopic and disconnected regions of spaces A and B the wave function factorizes, $\psi_{A,B}=\psi_A\otimes\psi_B$, this factorization being fulfilled by the exponential form.

Considering the t-J model the restriction of no doubly occupied sites can be thought of an infinitely negative $v(i,i)=-\infty$ charge correlation, whereas the restriction at a fixed number N of particles can be thought of as another singular Jastrow term $\exp[-\infty(N_{tot}-N)^2]$ where $N_{tot}=\Sigma_i n_i$ is the total charge operator. The latter two singular terms in the Jastrow factor are simply a restriction on the Hilbert space that can be very easily implemented without numerical instabilities.

We now consider how the symmetries of the finite-size t-J model drastically restrict the still too large number of variational parameters in the Jastrow-Slater wave function. Translation invariance implies that the Jastrow potential depends only on the vector difference $v(i,j)=v(\vec{R}_i-\vec{R}_j)$, the function v being invariant (s-wave) for all rotation and reflection symmetries of $\vec{R}_i-\vec{R}_j$. Moreover $v^z=0$ for a singlet wave function.

The singlet and translation symmetries imply also strong restrictions on the one-body Hamiltonian defining the Slater determinant. This Hamiltonian can be generally written

$$\hat{H}_S=\hat{H}_0+(\hat{\Delta}^\dagger+\hat{\Delta}), \tag{35}$$

$$\hat{\Delta}^\dagger=\sum_{\langle R,\vec{\tau}\rangle}\Delta(\vec{\tau})(\tilde{c}^\dagger_{R,\uparrow}\tilde{c}^\dagger_{R+\vec{\tau},\downarrow}+\tilde{c}^\dagger_{R+\vec{\tau},\uparrow}\tilde{c}^\dagger_{R,\downarrow}), \tag{36}$$

where $\hat{H}_0=\Sigma_{k,\sigma}\epsilon_k\tilde{c}^\dagger_{k,\sigma}\tilde{c}_{k,\sigma}$ is the free-electron tight-binding nearest-neighbor Hamiltonian, $\epsilon_k=-2t(\cos k_x+\cos k_y)-\mu$, μ is the free-electron chemical potential, and $\hat{\Delta}^\dagger$ creates all possible pairs at the various distances $|\tau|$ with definite rotation-reflection symmetry [e.g., $d_{x^2-y^2}$ implies $\Delta(1,0)=-\Delta(0,1)$].

For a generic Jastrow-Slater singlet state, satisfying all symmetries of the t-J model, a quite large number of variational parameters are therefore available corresponding to $v(\vec{\tau})$ and $D(\vec{\tau})$ for all distances $|\vec{\tau}|$. Not all these parameters are independent; namely, the substitution $v(|\tau|)\to v(\vec{\tau})+\text{const}$ does not change the wave function up to a constant, so that $v(\vec{\tau})$ can be assumed to be zero at the maximum distance. An analogous dependence exists between the various parameters $\Delta(\vec{\tau})$, since after projecting it at a fixed number of particles, the ground state of the Slater determinant Hamiltonian (35) can be written

$$|S\rangle=\left(\sum_{R,\tau}f(\vec{\tau})(\tilde{c}^\dagger_{R,\uparrow}\tilde{c}^\dagger_{R+\vec{\tau},\downarrow}+\tilde{c}^\dagger_{R+\vec{\tau},\uparrow}\tilde{c}^\dagger_{R,\downarrow})\right)^{N/2}|0\rangle, \tag{37}$$

where $f(\tau)$ is the pairing wave function simply related to $\Delta(\tau)$ in Fourier transform:

$$f(k)=\frac{\Delta(k)}{\epsilon_k+\sqrt{\Delta_k^2+\epsilon_k^2}},$$

where ϵ_k in the above expression is not limited to nearest-neighbor hopping.

Thus, if we scale $f(\tau)\to\text{const}\times f(\tau)$, the many-body wave function (37) remains unchanged, implying that the number of independent parameters is equal to $N_{shell}-1$, where N_{shell} is the number of independent shells at $|\tau|>0$ consistent with the rotation-reflection symmetry of f. Notice that, once in the determinant part of the wave function $N_{shell}-1$ variational parameters are independently varied, it is useless to consider other terms, such as, for instance, the chemical potential μ or next-nearest-neighbor hopping in H_0: they always provide a suitable renormalization of f that can be sampled by the first $N_{shell}-1$ parameters. Moreover, by performing a particle-hole transformation in Eq. (35) on the spin down $\tilde{c}^\dagger_{i,\downarrow}\to(-1)^i\tilde{c}_{i,\downarrow}$, the ground state of the Hamiltonian (35) is just a Slater determinant with $N=L$ particles.[21] This is the reason why this variational wave function represents a generic Jastrow-Slater state, a standard variational wave function used in QMC. Using the particle-hole transformation, it is also possible to control exactly the spurious finite system divergences related to the nodes of the d-wave order parameter.

B. Stochastic minimization

Among the correlation functions important to define the variational wave function two classes are important for guiding functions of the Jastrow-Slater form (33).

(i) The first class of operators renormalizes the Slater determinant and has been identified by Filippi and Fahy.[22] Here O^k are defined by means of the one-body operators $O^k_{1\text{-}body}$ by the following relation:

$$O^k_x=\frac{\langle x|O^k|\psi_G\rangle}{\langle x|\psi_G\rangle}=\frac{\langle x|O^k_{1\text{-}body}|S\rangle}{\langle x|S\rangle}. \tag{38}$$

Thus for $|\alpha_k|$ small, $(1+\Sigma_k\alpha_kO^k)|\psi_G\rangle\simeq\hat{J}\exp(\Sigma_k\alpha_kO^k_{1\text{-}body})|S\rangle$, which remains a Jastrow-Slater wave function of the same form $J|S'\rangle$ with $|S'\rangle=\exp(\Sigma_k\alpha_kO^k_{1\text{-}body})|S\rangle$. Since one-body operators are bilinear (e.g., $c^\dagger c$) in fermion second-quantization operators, S' remains a Slater determinant.[20] In the Jastrow-Slater case for the t-J model considered here the one-body operators read

$$O^k_{1\text{-}body}=\sum_{R,\tau\in shell\#k}S(\tau)(\tilde{c}^\dagger_{R,\uparrow}\tilde{c}^\dagger_{R+\tau,\downarrow}+\tilde{c}^\dagger_{R+\tau,\uparrow}\tilde{c}^\dagger_{R,\downarrow}), \tag{39}$$

where the sign $S(\tau)=\pm1$ is determined by symmetry. Also the bar kinetic energy H_0 is considered in this approach. According to the previous discussion the chemical potential μ is fixed to the free-electron one in H_0.

(ii) The second class of correlation functions is composed of the ones that appear in the Jastrow factor. They are the diagonal operators O^k—density-density $\Sigma_R n_R n_{R+\tau}$ or spin-spin $\Sigma_R\sigma^z_R\sigma^z_{R+\tau}$—in the chosen basis x of configurations

with fixed spins and electron positions. Again for small α they can be considered a renormalization of the Jastrow factor: $J \to J \exp(\Sigma_k \alpha_k O^k)$.

The multiparameter minimization method can be summarized as follows.

(i) After each reconfiguration the factor $r_x = \Sigma_k \alpha_k O_x^k$ is computed with given α_k, whose statistical fluctuations can be arbitrarily reduced by increasing the number of walkers or the bin length L_b as described in Sec. II. In this case of nonlinear optimization the bin technique is particularly important because it allows one to avoid, for large enough bin length L_b, unphysical fluctuations of the guiding wave function.

After an exact Green function step a wave function better than ψ_G is obtained and is parametrized by the coefficient α_k contained in the factor r_x: In fact,

$$P_{SR}\psi_{n+1} = P_{SR} G P_{SR} \psi_n = r_x \psi_G(x)$$

has to be by definition a variational state better than $P_{SR}\psi_n$,[23] which in turn is better than ψ_G, since for instance we can assume $\psi_{n=0} = \psi_G$.

(ii) It is therefore convenient to change at each iteration n the guiding wave function

$$|\psi_G'\rangle \to \exp\left(\sum_{k=1}^{p} \bar{\alpha}_k O^k \right) |\psi_G\rangle. \tag{40}$$

In the above equation we have introduced the new scaled coefficients $\bar{\alpha}_k = \alpha_k / C$ simply related to the ones α_k defined by the SR conditions (3). This is obtained by recasting r_x in a form that is more suitable for exponentiation:

$$r_x = C\left[1 + \sum_{k=1}^{p} \bar{\alpha}_k (O_x^k - \bar{O}^k) \right], \tag{41}$$

where

$$\bar{O}^k = \frac{\sum_x \psi_G(x)^2 O_x^k}{\sum_x \psi_G(x)^2} = \langle\langle w_j^f O_x^k \delta x, x_j \rangle\rangle$$

and $C = 1 + \Sigma_{k=1}^{p} \alpha_k \bar{O}^k$. The above exponentiation is justified, provided $\psi_G'/||\psi_G'|| \simeq \psi_G/||\psi_G||$. This is certainly fulfilled at equilibrium when for large n $\psi_{n+1}' = P_{SR}\psi_{n+1} \propto P_{SR}\psi_n \propto \psi_G^*$, ψ_G^* being the Jastrow-Slater wave function with the lowest-energy expectation value. In fact at equilibrium the SR conditions turn exactly to the Euler equations of minimum energy for ψ_G^*:

$$\frac{\langle \psi_G^* | O^k | \psi_G^* \rangle}{\langle \psi_G^* | \psi_G^* \rangle} = \frac{\langle \psi_G^* | O^k H | \psi_G^* \rangle}{\langle \psi_G^* | H | \psi_G^* \rangle}, \tag{42}$$

as implied by Eq. (3) for $\psi_{n+1}' = \psi_G^*$ and $\psi_{n+1} = G\psi_n' \propto G\psi_G^*$, and taking also into account that here for simplicity O^0 is the identity.

This implies that ψ_G^* is a lowest-energy wave function of the Jastrow-Slater form. There may be many local minima when Euler's equations are identically satisfied. In this case the SR represents a very useful tool for global minimization. In fact the bin length L_b or the number of walkers M represents at each iteration n an effective inverse temperature that can be increased gradually following the well-known ''simulated annealing'' statistical algorithm.[24] As is shown in Fig. 3 we apply this technique with a very short bin length using a full translation-invariant singlet wave function and using x,y reflection symmetries without rotation symmetry. This amounts to 24 independent parameters for a 26-site cluster ($N_{shell} = 13$). In the plot we show the evolution of the short-range BCS parameters $\Delta(1,0), \Delta(0,1)$ when at the beginning $\psi_{n=0} = \psi_G$ was set with no Jastrow term $v = 0$ and the s-wave symmetric determinant defined by $\Delta(1,0) = \Delta(0,1) = 0.1t$, $\Delta_\tau = 0$ for $|\tau| > 1$. The s-wave symmetric solution is locally stable, but for short enough bin there is a finite tunneling probability to cross the barrier and stabilize the much lower-energy solution with d-wave symmetry.

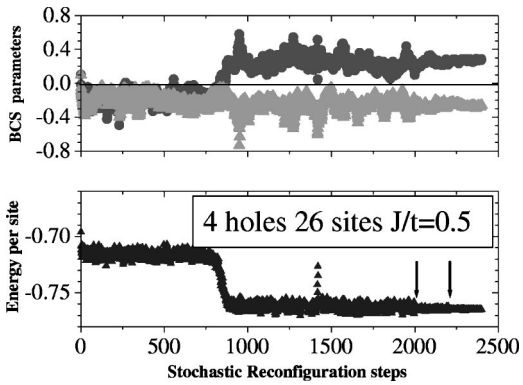

FIG. 3. Energy per site and evolution of the pairing amplitudes for four holes in the t-J model at $J/t = 0.5$ as a function of the stochastic power method iteration n. Here 200 walkers were used. Upper panel: triangles and circles denote nearest-neighbor $\Delta(\tau)$ amplitudes in Eq. (35) along the x and y directions, respectively. Lower panel: the arrows indicate the power iterations when the Monte Carlo bin length L_b has increased from 10 to 100 (left arrow) and from 100 to 500 (right arrow) Monte Carlo steps.

(iii) In order to continue with the new guiding wave function (40) without another long equilibration, walker weights w_j and w_j^f in Eqs. (11) and (12) can be reweighted as follows:

$$w_j \to w_j [\psi_G'(x_j)/\psi_G(x_j)],$$

$$w_j^f \to w_j^f [\psi_G'(x_j)/\psi_G(x_j)]^2 \tag{43}$$

in order that the new weights acting on the same configurations x represent statistically Eq. (11) with $\psi^f = \psi_G = \psi_G'$ and Eq. (12) with the new guiding wave function ψ_G'.

(iv) For large n the one-body operators corresponding to the Slater determinant may become linearly dependent because D may approach an eigenstate of a one-body Hamiltonian $\Sigma h_k O_{1\text{-}body}^k$, $\Sigma h_k O_{1\text{-}body}^k |S\rangle = E|S\rangle$, with suitable constants h_k and E. Thus the covariance matrix $s_{k,k'}$ quickly becomes singular, leading to a numerical instability which is difficult to control statistically. A stable method to overcome

this difficulty was found by Filippi and Fahy,[22] essentially by taking out one operator (the one-body kinetic energy) from the ones used in the linear system (3). Here we have found it more convenient to solve the linear system (3),

$$\sum s_{k,k'}\alpha_{k'}=\langle\psi_G|O^kG|\psi_n\rangle=f_k,$$

by diagonalizing first the symmetric matrix

$$s_{k,k'}=\sum_j U_{k,j}\Lambda_j U_{k',j} \tag{44}$$

and taking out the lowest eigenvalue $\Lambda_0\geqslant 0$ component of the positive-definite matrix s. This is equivalent to selecting in the SR another set of $p-1$ operators which are no longer singularly dependent. This is a perfectly legal operation for $\Lambda_0\sim 0$, since as shown in Ref. 5, for the singular operator $O^*=\Sigma_k U_{k,0}O^k$, such that $O^*|\psi_G\rangle=0$, the SR condition (3) is identically satisfied. Thus the resulting linear system is no longer affected by the above numerical instability:

$$\Lambda_j\alpha'_j=\sum_k U_{k,j}f_k \tag{45}$$

with $\alpha'_0=0$ and

$$\alpha_k=\sum_{j>0} U_{k',j}\alpha'_j. \tag{46}$$

Finally we obtain a much more stable determination of α_k, which does not affect the result at equilibrium, where $\Lambda_0\rightarrow 0$ and the correct Euler equations are satisfied. With this scheme also optimization of the Jastrow parameters together with the Slater determinant ones is possible without too much effort.

We have found that the generic situation for Jastrow-Slater wave functions is that the optimal determinant is actually the ground state of a one-body Hamiltonian $H_{1\text{-}body}=\Sigma_k h_k O^k_{1\text{-}body}$, a particular linear combination of the chosen one-body operators used in the SR conditions. This is in agreement with the Filippi-Fahy ansatz.[22] Occasionally, however, the optimal determinant turns out to be the excited state of a one-body Hamiltonian.

In Figs. 4 and 5 we show the full Jastrow-Slater optimization for the t-J model in the largest-size cases where the exact solution is known: 4 holes in 26 sites[25] and 2 holes in 32 sites,[26] respectively. We display the hole-hole correlation functions $N(R)=\langle(1-n_0)(1-n_R)\rangle$ on the variational wave function with and without the Jastrow factor. We see that the improvement towards an exact solution is crucially dependent on the Jastrow density-density factor especially at long-range distance. This behavior seem to be analogous to the one-dimensional (1D) case where long-range Jastrow factors are enough to determine the anomalous long-range behavior of correlation functions in one-dimensional Luttinger liquids.[27,28] The remarkable accuracy of the Jastrow-Slater wave functions is clearly limited (see, e.g., Fig. 2) to the region $J/t\lesssim 0.5$ where pairs with d-wave symmetry repel each other and do not form many-particle bound states as is the case for large J/t when phase separation occurs.

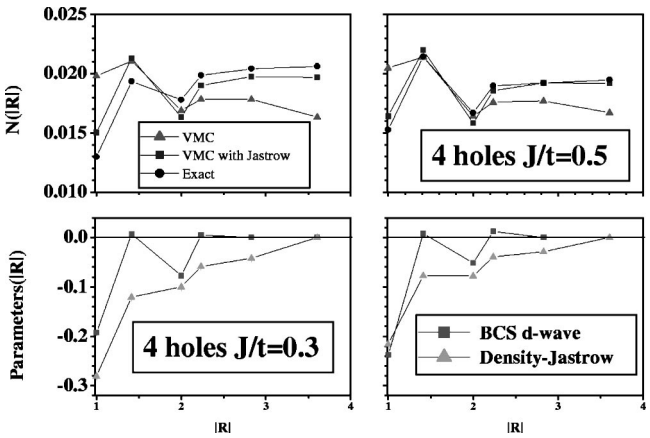

FIG. 4. Hole-hole correlations (upper panels) and Jastrow-Slater parameters (lower panels) for the optimal variational singlet wave functions with pairing wave function f with d-wave symmetry in a 26-site cluster t-J model.

For the 32-site cluster the spin-translation symmetry has to be explicitly broken in the variational wave function by introducing a staggered magnetization $H_0\rightarrow H_0-\Delta_{AF}\Sigma_R(-1)^R\sigma^z_R$ along the z axis. The best wave function compatible with reflection and rotation and translation with spin interchange ($\uparrow\leftrightarrow\downarrow$) can be conveniently parametrized by Δ_{AF} and a next-neighbor hopping t' in the Slater determinant[29] and also a spin Jastrow factor is allowed within this class of wave functions. Even in this case, as shown in the corresponding Fig. 5, the hole-hole correlations are almost exactly reproduced by the strong attractive Jastrow term at long distance. This means that in this small doping regime it is important to have a broken symmetry ground state, which suppress the d-wave BCS order parameter.

The next step is to perform few Lanczos steps over these variational wave functions which have been shown to be very accurate but not an exact representation of the ground-

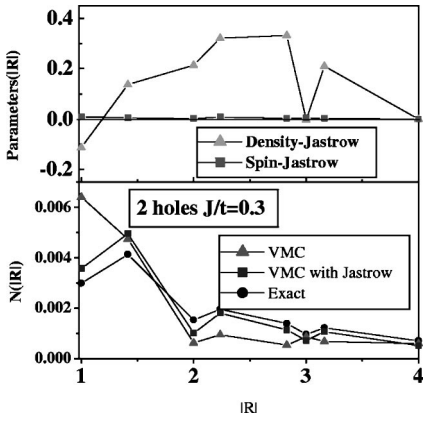

FIG. 5. Same as in Fig. 4 for the 32-site cluster, with a broken symmetry Slater determinant wave function with $\Delta_{AF}=0.2t$, $\Delta_{BCS}=0.1t$, and $t'/t=-0.15$.

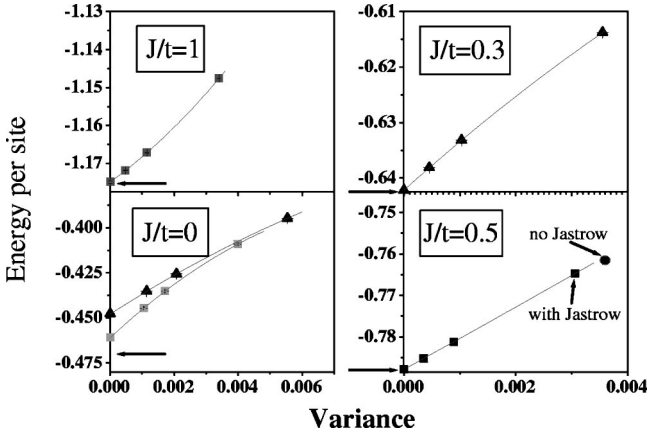

FIG. 6. Energy per site as a function of the number of Lanczos steps $p=0$ (higher energy variance), $p=1$ (medium-energy variance), and $p=2$ (lowest energy with nonzero variance) starting from the optimal Jastrow-Slater wave function for 4 holes in the 26-site clusters and several J/t. Only for $J/t=0$ is the broken symmetry solution (with the spin Jastrow factor $v^z \neq 0$) better than the best singlet wave function (triangles). The arrows indicate the exact energies, whereas the zero-variance energies are the extrapolated results with a quadratic fit (solid lines). For $J/t=0.5$ we show the corresponding variational $p=0$ energy and variance for the wave function without a Jastrow term.

state many-body wave function. In the singlet case with no broken symmetry the energy as a function of the variance of the energy per site,

$$\frac{\langle \psi_G | (H/L)^2 | \psi_G \rangle}{\langle \psi_G | \psi_G \rangle} - \left(\frac{\langle \psi_G | H/L | \psi_G \rangle}{\langle \psi_G | \psi_G \rangle} \right)^2, \tag{47}$$

is indeed smoothly related to the exact ground-state energy. The reason is that for a good variational state the variance approaches zero as the energy becomes exact, and this property can be used to estimate the exact energy by a simple linear extrapolation energy versus variance for several variational wave functions, as pointed out by several authors before.[30–32] The combination of the Lanczos approach and the variance extrapolation is particularly effective since the Lanczos technique converges remarkably fast for a good initial wave function, so that in the variance extrapolation a systematic and reliable extrapolation can be easily obtained.[33]

Whenever on a finite system a broken symmetry variational solution has much lower energy than the fully symmetric one, the Lanczos method is less effective. This is shown, for instance, for the 26 sites at $J/t=0$, Fig. 6, or in the 32-site–2-hole case, Fig. 7. In the latter case we show also the energy as a function of the variance for the fully symmetric solution. We see in this case that the approach to the exact solution for a singlet wave function is rather difficult but indeed possible.

On a finite system there is always a small energy gain to recover a state with definite spin. It is very difficult to obtain this residual energy with few Lanczos step iterations, since in order to average over the various directions of the order parameter many Hamiltonian power iterations are required.

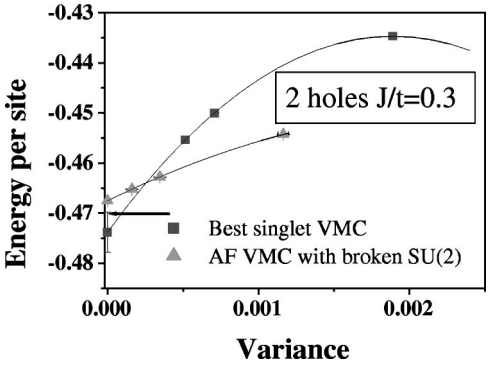

FIG. 7. Energy per site as a function of the number of Lanczos steps $p=0$ (higher energy with nonzero variance), $p=1$ (medium energy with nonzero variance), and $p=2$ (lowest energy with nonzero variance) starting from the optimal Jastrow-Slater wave function for a 32-site t-J square lattice. The arrows indicate the exact energies, whereas the zero-variance energies are the extrapolated results with a quadratic fit (solid lines). The broken symmetry solution is described in Fig. 5 whereas the best singlet wave function is obtained by optimizing only the density-density Jastrow and the d-wave parameters as in Eq. (35).

However, this residual energy should vanish in the thermodynamic limit if the symmetry is indeed spontaneously broken. Therefore we conclude that the small discrepancy between exact results and extrapolated ones in Figs. 6 and 7 is irrelevant within the above assumption, which is confirmed by simulations on a much larger size, showing antiferromagnetic long-range order at small doping[13] and ferromagnetism for the $J=0$ case.[34] In any case the variance extrapolation with few Lanczos steps provides always a much more accurate estimate of the exact ground-state energy compared to the lowest variational estimate.

VI. RESULTS AND DISCUSSION

The variational approach is certainly limited and ''biased'' by the ''human'' choice of the variational wave function, ''believed'' to be the correct one for the physical problem considered. In this paper we have described a variational approach that improves systematically any given variational wave function with a couple (and in principle more) Lanczos steps (LS's), with reasonable computational effort. This approach is certainly limited, especially for large sizes, when few Lanczos iterations cannot remove the possible large bias of the initial variational guess. However, for 2D fermionic systems on a lattice, in the strong correlation regime, i.e., close to a Mott insulator state, it is very difficult to improve the best variational wave function obtained with the Lanczos scheme (see Table I). This result is particularly meaningful if we consider that in principle the FN technique is size consistent (lowers the energy per site of the variational guess even in the thermodynamic limit) and the LS's technique with a fixed number of iterations p is not. Thus, close to a Mott insulator, it is very important to change the nodes of the wave function—which the Lanczos technique allows—rather than improving only the amplitudes—as in the FN technique. It is worth mentioning, however, that the FN energy reported

in Table I is only an upper bound of the expectation value of the Hamiltonian on the FN state. We do not know exactly how much this will lower the energy in favor of the FN technique, but we expect that this should be only a minor correction, especially for large sizes.

As shown in Table I the variational energy can be further improved by using the $r>0$ technique, which, within the formalism of the present paper, can be thought as a self-consistent improvement of the amplitudes and the nodes of the few-Lanczos-step variational wave function. For $r=1$ the original SR scheme is recovered, which as seen in (Fig. 2), may not be the optimal choice from the variational point of view.[3,5]. As far as the variational energy is concerned the self-consistent approach ($r>0$) is not very effective (the improvement is between 20% and 30% on small size systems) for small system sizes but appears to be more and more important as the size is increased. Also correlation functions may be qualitatively improved (Fig. 1) by the self-consistent approach, especially when a many-particle bound state appears (''stripes'' or phase separation in this model) which is not contained at the variational level.

An important advantage of the standard variational approach ($r=0$) is that the error in the ground state energy and correlation functions can be estimated using that the variance of the energy per site (47) in an exact calculation should be zero. The variance can be estimated systematically with high statistical accuracy for the first p Lanczos steps acting on the initial variational wave function ψ_G. We show that the approach to the exact result may be smooth, even for large system size and number of electrons, even when ψ_G is not particularly close to the exact result. Obvious exceptions exist and are shown here in Fig. 2 for correlation functions, whereas the energy seems always better behaved (see Fig. 6).

We have tested this simple scheme in the 2D-Heisenberg model where an exact solution of energies and correlation functions is easily available by using standard techniques. The 2D-Heisenberg model has off-diagonal long-range order in the ground state, the order parameter

$$m^2=\frac{1}{L^2}\sum_{R,R'}\vec{S}_R\cdot\vec{S}_{R'}(-1)^{R-R'}$$

being finite in the thermodynamic limit $L\to\infty$. We start with the variational wave function in Eq. (35) obtained by projecting out the doubly occupied states to a wave function with d-wave nearest-neighbor BCS pairing correlations,[35] but without any explicit antiferromagnetic order parameter.

This wave function represents an accurate wave function for quantum antiferromagnets as far as the energy is concerned, but certainly has not the right behavior at large distances and may be considered a resonating valence band (RVB) disordered variational wave function.[1] After applying only two Lanczos iterations the 18-site size is almost exactly reproduced by this simple wave function, showing that at short distance the quantum antiferromagnetic wave function is almost indistinguishable from a RVB one.[6] As we increase the size, the variational energy calculation (see Fig. 8) clearly loses accuracy, since the gap to the first excited state scales to zero and the Lanczos algorithm becomes correspondingly less effective. This loss of accuracy is, however, not dramatic (asymptotically it scales as $1/\sqrt{L}$), as shown by the similar quantitative agreement with the exact result obtained with both the 50- and 98-site lattices (Fig. 8).

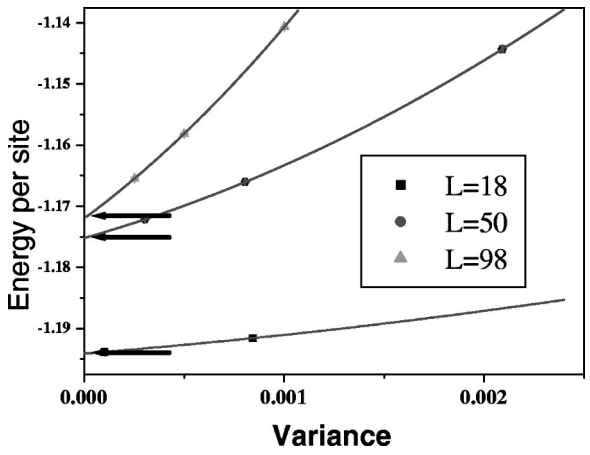

FIG. 8. Energy per site of the finite-size Heisenberg model. Comparison of exact results (indicated by arrows) and the approximate $p=0,1,2$ Lanczos step iterations over the projected d-wave wave function. Solid lines are a quadratic fit of the data.

Also remarkable is the correct trend of long-range correlation functions as we increase the number of Lanczos iterations. As shown in the inset of Fig. 9 the linear extrapolation with variance, which is valid for correlation functions averaged over the system volume (see Appendix D), is capable of detecting almost the right long-range magnetic order in the Heisenberg model. This is remarkable if we consider that the starting wave function ψ_G is disordered, as also shown in the same inset. We remark also that in this particular case the original SR algorithm[3,5] for $r=1$ and $\Lambda\to\infty$ is *exact* since the nodes of the variational wave function are the correct ones.

As anticipated the estimate of the variational error, by using the fact that the variance has to approach zero in an exact calculation, is really effective in this case and shows

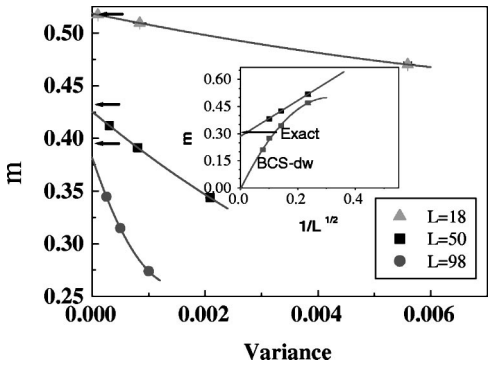

FIG. 9. Order parameter $m=\sqrt{S(\pi,\pi)/L}$ in the finite-size Heisenberg model [$S(\pi,\pi)$ being the spin isotropic antiferromagnetic structure factor]. Comparison of exact results (indicated by arrows) and the approximate $p=0,1,2$ Lanczos step iterations over the projected d-wave wave function. Solid lines are a quadratic fit of the data. Inset: finite-size scaling with the variational (BCS d-wave) wave function and with the variance extrapolated one.

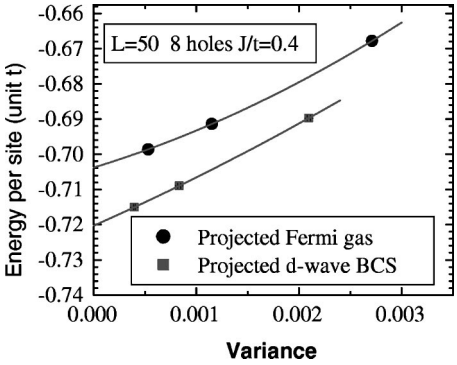

FIG. 10. Energy per site vs variance for two different initial wave functions (highest-energy dots) after applying one (medium-energy points) and two (lowest-energies points) Lanczos steps. Lines are a quadratic fit to the data.

that, even for a large size, it is possible to reach very good quantitative estimates of the energies and correlation functions (see Ref. 33), even when the starting wave function is not particularly close to the exact one. In the Lanczos algorithm the variance becomes zero only when the lowest-energy state nonorthogonal to the initial wave function is reached. However, we expect that the variance may reach very small values close to a ''quasieigenstate,'' implying the failure of the variance variational error estimate. In fact, in this case the energy optimization Lanczos technique is trapped in a local minimum, with no possibility of tunneling to the true global minimum energy (see Fig. 10).

Clearly this is a well-known problem in numerical optimization and the only possibility, when the exact solution is not known, is to check the various candidates for the ground state and determine the one with the lowest energy. From another point of view, this property of the Lanczos algorithm may be useful to estimate physically observable quantities, such as the condensation energy for a metal to become a superconductor, which is just the macroscopic difference in energy between two thermodynamically stable states. From Fig. 10 an estimate of this condensation energy is $0.02t \sim 100$ K, in reasonable agreement with experiments on high-T_c cuprates,[36] suggesting that the main features of *d*-wave superconductivity can be understood with this simple model.

At the moment, the approach we have presented here has to be limited to very few Lanczos iterations for large sizes with the given computer resources. Nevertheless, it is certainly systematic and unbiased as far as the approach to the ground state is concerned: the corrections to the initial guiding function depend only on the Hamiltonian H and no other biased approximation. Compared to the standard FN technique, it allows a systematic improvement of the starting variational wave function ψ_G by correcting not only its amplitudes but also the nodes.

The extension of this technique to continuous models is straightforward. For the reference dynamic [given by G^f in Eq. (22)] one can use the Langevin dynamic, as is done in Ref. 37, so that it is possible to determine the lowest-energy state obtained by applying to ψ_G few Lanczos steps with ($r>0$) or without self-consistency ($r=0$) or by using a many-parameter variational wave function, with a nonlinear optimization as described in Sec. V.

ACKNOWLEDGMENTS

I am deeply indebted to V. Leuveen who convinced me in CECAM-Toulose about the importance of having a rigorous variational method for the ground state of strongly correlated systems. I am also particularly grateful to D. Poilblanc and P. W. Leung who kindly sent me unpublished data on the 26- and 32-site *t-J* model, respectively. Special thanks go to F. Becca, L. Capriotti, M. Capone, A. Parola, and G. Santoro for many useful comments about this paper. This work was partially supported by INFM and MURST (COFIN99).

APPENDIX A: PROPERTIES OF THE OPERATOR $P_{SR}(\Lambda I-H)P_{SR}$

In this appendix we focus on some properties of the matrix $P_{SR}(\Lambda I-H)P_{SR}$, defined in Eq. (21).

(i) The maximum eigenvalue $\Lambda-E_{SR}$ of the Hermitian matrix $P_{SR}(\Lambda I-H)P_{SR}$ is certainly smaller than the corresponding one $\Lambda-E_0$ of the exact Green function $\Lambda I-H$. In fact $\psi_{SR}=P_{SR}\psi_{SR}$ is the normalized eigenstate of $P_{SR}(\Lambda I-H)P_{SR}$ with eigenvalue E_{SR}; then $P_{SR}^{\dagger}=P_{SR}$ implies that $E_{SR}=\langle\psi_{SR}|P_{SR}HP_{SR}|\psi_{SR}\rangle=\langle\psi_{SR}|H|\psi_{SR}\rangle\geqslant E_0$, as ψ_{SR} can be considered a variational state of the exact Hamiltonian H with energy E_{SR}.

(ii) Since ψ_{SR} is obtained by applying $[P_{SR}(\Lambda I-H)P_{SR}]^n$ to a given trial wave function, for $n\to\infty$ such a propagated wave function will converge therefore to the lowest-energy state in the subspace projected by P_{SR}. This implies clearly that E_{SR} is lower than or at most equal to the variational energy on the reference wave function, $\langle\psi_G|H|\psi_G\rangle$, simply because ψ_G belongs to this subspace.

(iii) Since ψ_G belongs to the subspace projected by P_{SR}, $\langle\psi_G|HP_{SR}=\langle\psi_G|P_{SR}H$. Therefore the mixed average estimate, statistically much more convenient,

$$E_{MA}=\frac{\langle\psi_G|H|\psi_{SR}\rangle}{\langle\psi_G|\psi_{SR}\rangle}=\frac{\langle\psi_G|P_{SR}HP_{SR}|\psi_{SR}\rangle}{\langle\psi_G|P_{SR}|\psi_{SR}\rangle}=E_{SR},$$

coincides with the variational bound E_{SR} of the ground-state energy as ψ_{SR} is an exact eigenstate of $P_{SR}HP_{SR}$ with eigenvalue E_{SR}.

APPENDIX B: FORWARD WALKING

In order to compute correlation functions over ψ_{SR} it is necessary to use a slight generalization of the forward walking technique, generalized to a nonsymmetric matrix such as (29). Moreover, since in the meaningful SR limit of a large number of walkers or bin length $L_b\to\infty$ the parameters α_k can be assumed constants in r_x, Eq. (4), it is much more convenient to implement the forward walking technique without allowing any fluctuations of the random variables α_k. This can be done easily by first evaluating the expectation value ratios $\bar{\alpha}_k=\langle\alpha_k\rangle/\langle\alpha_0\rangle$, $k=1,\ldots,p$, with the standard SR algorithm, i.e., allowing the α_k fluctuations for each

Markov iteration n. The second step is to perform a different simulation, usually much more efficient as far as the error bars are concerned, with r_x determined by nonrandom constants $\bar{\alpha}_k$:

$$r_x = 1 + \sum_{k=1}^{p} \bar{\alpha}_k O_x^k. \tag{B1}$$

If the $\bar{\alpha}_k$ are determined accurately for M, L_b large, the SR conditions (3) will be automatically verified within error bars. The statistical or systematic error related to the determination of the constants $\bar{\alpha}_k$ is also not much important. In fact even assuming that with the first simulation the constants $\bar{\alpha}_k$ are determined with a non-negligible statistical error and an unavoidable systematic bias due to the finite number of walkers, the method that we will describe in the following will provide also in such a case a variational estimate of the energy with the chosen constants $\bar{\alpha}_k$. The analogy of this method with the Lanczos method is evident also in this case. Even in the latter technique, a first run is usually implemented to determine the coefficients of the ground state $\bar{\alpha}_k$ in the Krilov basis spanned by the initial wave function ψ_G and the powers of the Hamiltonian applied to it, $\psi_0 = \psi_G + \Sigma_{k=1,p} \alpha_k H^k |\psi_G\rangle$ (with some more technical ingredient to work with an orthonormal basis). Then correlation functions over ψ_0 are computed by recovering the ground-state wave function in this basis using the determined coefficients $\bar{\alpha}_k$.

Let us now focus on the implementation of the forward walking technique within the SR scheme at fixed constants $\bar{\alpha}_k$. Since $\bar{G}^f$ in Eq. (29) is not symmetric, its left eigenvector $\langle \psi_L | \propto \lim_{n\to\infty} \langle \psi_G | (\bar{G}^f)^n$ does not necessarily coincide with the corresponding right eigenvector ψ_R. Fortunately the matrix $\bar{G}^f$ can be easily written in terms of a symmetric matrix G^0:

$$\bar{G}^f_{x',x} = \alpha_{x'} G^0_{x',x} / \alpha_x, \tag{B2}$$

with

$$\alpha_x = \frac{|\psi_G(x)||r_x|^{r/2}}{\sqrt{z_x}}, \tag{B3}$$

$$G^0_{x',x} = \frac{|r_x|^{r/2} |r_{x'}|^{r/2}}{\sqrt{z_x z_{x'}}} |\Lambda \delta_{x',x} - H_{x',x}|. \tag{B4}$$

Therefore the right and left eigenvectors of $\bar{G}^f$ are easily written in terms of the maximum eigenstate ϕ_0 of the symmetric matrix G_0, namely, $\psi_R(x) = \alpha_x \phi_0(x)$ and $\psi_L(x) = \phi_0(x)/\alpha(x)$. Then using the definition of the SR state, Eq. (30), it follows that also the left eigenvector of $\bar{G}^f$ can be written in terms of ψ_{SR}:

$$\psi_{SR}(x) = L_x \psi_L(x), \tag{B5}$$

with

$$L_x = \psi_G(x) r_x / z_x \tag{B6}$$

and

$$R_x = |r_x|^{1-r} / \psi_G(x) \operatorname{sgn} r_x. \tag{B7}$$

After applying several times the Green function $\bar{G}^f$ the walkers w^f, x determine the state $\psi_R(x)$. Then it is possible to evaluate expectation values of any operator O with given matrix elements $O_{x',x}$ by applying the following relation, which corresponds to propagating n times forward $\psi_R(x)$:

$$\frac{\langle \psi_{SR}|O|\psi_{SR}\rangle}{\langle \psi_{SR}|I|\psi_{SR}\rangle} = \lim_{n\to\infty} \frac{\sum_{x'',x',x} (\bar{G}^f)^n_{x'',x'} \bar{O}_{x',x} \psi_R(x)}{\sum_{x'',x',x} (\bar{G}^f)^n_{x'',x'} \bar{I}_{x',x} \psi_R(x)}, \tag{B8}$$

where the matrix elements of O and the identity I are replaced by the ones of the left-right transformed matrices $\bar{O}$ and $\bar{I}$, respectively. The explicit matrix elements of $\bar{O}$ and $\bar{I}$ in the RHS of the above equation are given by

$$\bar{O}_{x',x} = L_{x'} O_{x',x} R_x, \tag{B9}$$

$$\bar{I}_{x',x} = L_x R_x \delta_{x',x}. \tag{B10}$$

This means that in the standard forward walking technique,[13] instead of using the importance-sampled matrix elements obtained with $L_x = \psi_G(x) = 1/R_x$ in Eq. (B9), the slightly more involved ones (B9) and (B10) have to be considered. In fact by simple substitutions of these matrix elements into Eq. (B8), using also that $\Sigma_{x'',x',x} (\bar{G}^f)^n_{x'',x'} \propto \psi_L(x') = \psi_{SR}(x')/L_{x'}$, Eq. (B5), and that $\psi_R(x) = \psi_{SR}(x)/R_x$, Eq. (30), Eq. (B8) is easily verified. The statistical algorithm used to evaluate the ratio in Eq.(B8) is very similar to the standard ''forward walking'' technique[13] for diagonal operators. The few differences are the following.

(i) Also the denominator in Eq. (B8) has to be ''forward'' propagated for n iterations, since in this case the diagonal elements of $\bar{I}$ are not trivially one (since $L_x \neq R_x^{-1}$). The error bars have to be then calculated taking into account that the numerator and the denominator are very much correlated.

(ii) Off-diagonal operators can be computed without performing another simulation, provided the matrix elements of the operator O are contained in the nonvanishing ones of the Green function G (or some power of G if the operator is evaluated statistically). In particular the expectation value of the Hamiltonian and the even more accurate ones (25) can be computed altogether with a single Markov chain.

(iii) Similarly the accuracy of diagonal and off-diagonal operators can be further improved by computing

$$\frac{\langle \psi_{SR}|G^k O G^k|\psi_{SR}\rangle}{\langle \psi_{SR}|G^{2k}|\psi_{SR}\rangle}.$$

In fact an important advantage of the SR technique is that the reference Green function $G^f_{x',x}$ is nonzero for all nonzero

elements of the exact Green function G (whereas in the FN technique the matrix elements with negative sign are suppressed). Thus exact sampling of the Green function G can be done with the standard reweighting method, requiring only the finite multiplicative factors $s_{x',x}=G_{x',x}/G^f_{x',x}$, calculated for each iteration n and each walker of the Markov chain. The same technique can be obviously generalized when the reference Green function G^f is the fixed node one—slightly generalized to have the possibility to cross the nodes[5]—simply replacing G^f and $z_x=1$ in the above expressions. However, the statistical accuracy for the determination of the constants $\{\alpha_k\}$ is very bad with the FN reference G^f, about an order of magnitude less efficient than the one in Eq. (22), without a significant improvement in variational energies. The reason for such bad behavior [or the successful one for Eq. (22)] is not clear at present.

APPENDIX C: EFFICIENT CALCULATION OF THE SINGLE-LANCZOS-STEP WAVE FUNCTION

In this appendix we describe an efficient way to find the optimal LS wave function $|\psi_\alpha\rangle=(1+\alpha H)|\psi\rangle$, starting from a chosen variational guess $|\psi\rangle$, i.e., to calculate the value of α for which the energy

$$E(\alpha)=\frac{\langle\psi_G|(1+\alpha H)H(1+\alpha H)|\psi_G\rangle}{\langle\psi_G|(1+\alpha H)^2|\psi_G\rangle} \tag{C1}$$

has a minimum. A standard method is to calculate statistically the various powers of the Hamiltonian

$$h_n=\frac{\langle\psi_G|H^n|\psi_G\rangle}{\langle\psi_G|\psi_G\rangle}, \tag{C2}$$

using configurations x generated by the Metropolis algorithm according to the weight $\psi_G(x)^2$. This method is, however, inefficient since much better importance sampling is obtained when configurations are instead generated according to the optimal Lanczos wave function $\psi_{\alpha^*}(x)=[1+\alpha^* e_{\psi_G}(x)]\psi_G(x)$, where $e_\psi(x)=\langle\psi|H|x\rangle/\langle\psi|x\rangle$ (Ref. 18) is the local energy corresponding to a generic guiding wave function ψ, and α^* minimizes the above expectation value (C1) for $\alpha=\alpha^*$. This wave function ψ_{α^*} may be much better leading to much lower variances especially for the higher momenta h_2 and h_3.

In this appendix we describe an efficient way to find the optimal LS wave function $|\psi_{\alpha^*}\rangle$, starting from a chosen variational guess $|\psi_\alpha\rangle$ with energy

$$E(\alpha)=\frac{h_1+2\alpha h_2+\alpha^2 h_3}{1+2\alpha h_1+\alpha^2 h_2}, \tag{C3}$$

which easily written in terms of the energy momenta h_n.

In order to minimize Eq. (C3), given an arbitrary value of α, it is convenient first to compute the energy expectation value h_1 with the standard statistical method and then, in place of the remaining Hamiltonian higher momenta h_2 and h_3, generate statistical configurations according to $\psi_\alpha(x)^2$ and compute

$$E(\alpha)=\frac{\langle\psi_\alpha|H|\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle},$$

$$\chi=\frac{\langle\psi_\alpha|(1+\alpha H)^{-1}|\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle}.$$

Here $E(\alpha)$ is obtained by averaging over the chosen configurations the local energy corresponding to ψ_α, namely, $\langle e_{\psi_\alpha}\rangle$, whereas χ is obtained by averaging over the same configurations $\langle[1+\alpha e_{\psi_G}(x)]^{-1}\rangle$. Given χ it is straightforward to compute

$$h_2=[(\chi^{-1}-2)(1+\alpha h_1)+1]/\alpha^2,$$

and therefore given h_1 and h_2, the value of $E(\alpha)$ implicitly defines the highest momentum $h_3=[E(\alpha)(1+2\alpha h_1+\alpha^2 h_2)-h_1-2\alpha h_2]/\alpha^2$. Notice that the most difficult energy momentum h_3 is given by sampling an energy expectation value, which is by far statistically more accurate compared to the direct determination of h_3.

It is then possible to minimize analytically $E(\alpha)$, yielding

$$\alpha^*=\frac{-(h_3-h_1h_2)\pm\sqrt{(h_3-h_1h_2)^2-4(h_2-h_1^2)(h_1h_3-h_2^2)}}{2(h_1h_3-h_2^2)}, \tag{C4}$$

where the above sign $\pm$ is such as to minimize $E(\alpha^*)$.

The analytic minimization of $E(\alpha)$, Eq. (C3), given the values of χ, h_1, and $E(\alpha)$ itself, provides the exact value of α^* in Eq. (C4) within the statistical uncertainties. They become smaller and smaller whenever $\alpha\sim\alpha^*$. Typically two or at most three attempts are enough to reach an accurate determination of α^* when the condition

$$\chi=\frac{1}{1+\alpha^* E(\alpha^*)} \tag{C5}$$

is exactly fulfilled. This condition is true in general only for eigenstates of the Hamiltonian, but remains valid for the single-Lanczos-step wave function.

APPENDIX D: VARIANCE ESTIMATE OF THE ERROR IN "BULK" CORRELATION FUNCTIONS

In this appendix we estimate the error in correlation functions assuming that the ground state $|\psi_0\rangle$ is approximated by the wave function $|\psi_p\rangle$ distant ϵ_p from $|\psi_0\rangle$. Namely, with no loss of generality we write

$$|\psi_p\rangle=|\psi_0\rangle+\epsilon_p|\psi'\rangle, \tag{D1}$$

with $\langle\psi_0|\psi_0\rangle=\langle\psi'|\psi'\rangle=1$, ψ' representing a normalized wave function orthogonal to the exact one $\langle\psi_0|\psi'\rangle=0$. We restrict our analysis to thermodynamically averaged correlation functions O, the ones which can be written as a bulk average of local operators O_R: $O=(1/L)\Sigma_R O_R$. This class of operators includes for instance, the average kinetic or potential energy or the spin-spin correlation function at a given

distance τ, $O_R = S_R \cdot S_{R+\tau}$. If we use periodic boundary conditions, the expectation value of O_R on a state with a given momentum does not even depend on R and the bulk average does not represent an approximation:

$$\langle \psi_0 | O_R | \psi_0 \rangle = \langle \psi_0 | O | \psi_0 \rangle = C. \tag{D2}$$

We show here that the expectation value of bulk-averaged operators O on the approximate state ψ_p satisfies the following relation:

$$\frac{\langle \psi_p | O | \psi_p \rangle}{\langle \psi_p | \psi_p \rangle} = C + O(\epsilon_p^2, \epsilon_p / \sqrt{L}), \tag{D3}$$

thus implying that for large enough size the expectation value (D3) approaches the exact correlation function C linearly with variance. This allows one to obtain a good accuracy with a good variational calculation, which is not easy to obtain if a term $\sim \epsilon_p$ dominates.

The validity of the above statement is very simple to show under very general grounds. In fact by definition

$$\langle \psi_p | O | \psi_p \rangle = C + 2\epsilon_p \langle \psi' | O | \psi_0 \rangle + \epsilon_p^2 \langle \psi' | O | \psi' \rangle. \tag{D4}$$

The most important term proportional to ϵ in the above equation can be easily bounded by use of the Schwartz inequality

$$|\langle \psi' | O | \psi_0 \rangle|^2 = |\langle \psi' | O - C | \psi_0 \rangle|^2 \leqslant \langle \psi_0 | (O-C)^2 | \psi_0 \rangle. \tag{D5}$$

The final term in the latter inequality can be estimated under the general assumption that the correlation functions $C(\tau) = \langle (O_R - C)(O_{R+\tau} - C) \rangle$ decay sufficiently fast with distance $|\tau|$, as a consequence of the cluster property:

$$\langle \psi_0 | (O-C)^2 | \psi_0 \rangle = \frac{1}{L} \sum_{\tau} C(\tau).$$

This concludes the proof of the statement of this appendix, provided $\Sigma_\tau C(\tau)$ is finite for $L \to \infty$.

[1] G. Baskaran, Z. Zou, and P. W. Anderson, Solid State Commun. **63**, 973 (1987).

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

[3] S. Sorella, Phys. Rev. Lett. **80**, 4558 (1998).

[4] E. S. Heeb and T. M. Rice, Europhys. Lett. **27**, 673 (1994).

[5] S. Sorella and L. Capriotti, Phys. Rev. B **61**, 2599 (2000).

[6] S. Liang, B. Doucot, and P. W. Anderson, Phys. Rev. Lett. **61**, 365 (1988).

[7] L. Capriotti, A. E. Trumper, and S. Sorella, Phys. Rev. Lett. **82**, 3899 (1999).

[8] L. Capriotti and S. Sorella, Phys. Rev. Lett. **84**, 3173 (2000).

[9] M. Calandra, F. Becca, and S. Sorella, Phys. Rev. Lett. **81**, 5185 (1998).

[10] M. Calandra and S. Sorella, Phys. Rev. B **61**, 11 894 (2000).

[11] D. F. B. ten Haaf, H. J. M. van Bemmel, J. M. J. van Leeuwen, W. van Saarloos, and D. M. Ceperley, Phys. Rev. B **51**, 13 039 (1995).

[12] H. J. M. van Bemmel, W. van Saarlos, J. M. J. van Leeuwen, and G. An, Phys. Rev. Lett. **72**, 2442 (1994).

[13] M. Calandra and S. Sorella, Phys. Rev. B **57**, 11 446 (1998).

[14] See e.g., N. Trivedi and D. M. Ceperley, Phys. Rev. B **41**, 4552 (1990).

[15] C. S. Hellberg and E. Manousakis, Phys. Rev. B **61**, 11 787 (2000).

[16] See, e.g., J. K. Cullum and R. A. Willoughby, *Lanczos Algorithm for Large Symmetric Eigenvalue Computations* (Birkhauser, 1985).

[17] C. T. Shih, Y. C. Chen, and T. K. Lee, Phys. Rev. B **57**, 627 (1998).

[18] If there exist configurations x such that $\psi_G(x) = 0$ and $\psi'_G(x) \neq 0$, it is computationally convenient to regularize the given guiding function by adding to it a random noise which may be arbitrarily small within computer accuracy, thus preventing the above pathology with probability 1. By reducing this noise to zero the ''clean'' limit can be easily achieved without too much effort or complicated changes in the algorithm (Ref. 15).

[19] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).

[20] In this paper we adopt a more general definition for a Slater determinant, as any state which is an eigenstate of a one-body Hamiltonian, not necessarily commuting with the particle number.

[21] H. Yokoyama and H. Shiba, J. Phys. Soc. Jpn. **57**, 2482 (1988).

[22] C. Filippi and S. Fahy, J. Chem. Phys. **112**, 3523 (2000).

[23] This is strictly valid provided G is a positive-definite operator, which generally holds for large enough Λ. In fact, by the Schwartz inequality, it is simple to show that the expectation value of the energy on $P_{SR}\psi_{n+1}$ is lower than or at most equal to the one over $P_{SR}\psi_n$.

[24] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).

[25] D. Poilblanc, Phys. Rev. B **52**, 9201 (1995).

[26] A. L. Chernyshev, P. W. Leung, and R. J. Gooding, Phys. Rev. B **58**, 13 594 (1998).

[27] C. S. Hellberg and E. J. Mele, Phys. Rev. Lett. **67**, 2080 (1991); R. Valenti and C. Gros, *ibid.* **68**, 2402 (1992).

[28] F. Franjic and S. Sorella, Mod. Phys. Lett. B **10**, 873 (1996).

[29] A. Himeda and M. Ogata, preprint, Phys. Rev. Lett. **85**, 4345 (2000).

[30] See e.g., C.J. Umrigar, K. G. Wilson, and J. M. Wilkins, in *Computer Simulations Studies in Condensed Matter Systems: Recent Developments*, edited by D. P. Lanadau, K. K. Mon, and H. B. Schuttler (Springer, Berlin, 1988).

[31] C. Gros, Phys. Rev. B **42**, 6835(R) (1990).

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

[33] F. Becca, L. Capriotti, and S. Sorella, cond-mat/006353 (unpublished).

[34] F. Becca and S. Sorella, Phys. Rev. Lett. **86**, 3396 (2001).

[35] C. Gros, Phys. Rev. B **38**, 931 (1988).

[36] See e.g., M. R. Norman, M. Randeria, B. Janko, and J. C. Campuzano, Phys. Rev. B **61**, 14 742 (2000) and references therein.

[37] S. Baroni and S. Moroni, Phys. Rev. Lett. **82**, 4745 (1999).