Adaptive optimization of wave functions for fermion lattice models

Matteo Beccaria and Antonio Moro
Dipartimento di Fisica dell'Università di Lecce, I-73100, Italy
and Istituto Nazionale di Fisica Nucleare - INFN, Sezione di Lecce, Italy
(Received 23 May 2001; published 5 September 2001)

We present a simulation algorithm for Hamiltonian fermion lattice models. A guiding trial wave function is adaptively optimized during Monte Carlo evolution. We apply the method to the two dimensional Gross-Neveu model and analyze systematic errors in the study of ground state properties. We show that accurate measurements can be achieved by a proper extrapolation in the algorithm-free parameters.

DOI: 10.1103/PhysRevD.64.077502 PACS number(s): 05.50.+q, 11.10.Ef, 11.10.Kk, 71.10.Fd

Lattice field theory is a constructive framework where nonperturbative properties of quantum models can be addressed both analytically and by numerical techniques. The main standing theoretical viewpoints are the traditional Lagrangian approach [1] and the Hamiltonian formulation [2]. In the study of fermionic models, Lagrangian simulations suffer the drawback of requiring Grassmann variables that are difficult to handle numerically and must be integrated out explicitly leading to large nonlocal determinants. Instead, in the Hamiltonian approach, the treatment of Fermi anticommuting operators is straightforward. In particular, this holds in one spatial dimension where notoriously difficult *sign problems* [3] are tame.

Another important reason to resort to Hamiltonian methods is that they rely on powerful well founded many-body techniques [4]. In particular, a direct analysis of the ground state structure is often feasible through a guiding *trial wave function* [5]. This is an approximation to the exact ground state that can provide deep physical insights into the model under consideration [6]. Also, it plays a central role in the simulation algorithms, and the quality of the results depends critically on its accuracy [7]. Usually, it contains a set of free parameters that deserves optimization by rather expensive variational calculations [8].

Here we present a Monte Carlo (MC) algorithm that includes an automatic optimization of the trial wave function by means of a nonlinear feedback between state sampling and guiding. The MC core is based on a general stochastic representation of matrix evolution problems [9], and was discussed in the specific case of the Hubbard model [10]. The adaptive optimization strategy has been already applied to diffusion MC studies of purely bosonic models with continuous state space [5].

In this Brief Report, we focus on fermionic models and present an algorithm suitable for the study of Hamiltonians acting on a finite-dimensional fully discrete state space. In fact, for a local fermion model discretized on a finite lattice, the Hamiltonian is a large sparse matrix $H=\{H_{ss'}\}_{s,s'\in S}$, with S denoting the discrete state space. The ground state can be obtained by acting on a given initial state with the evolution semigroup $\Omega=\{e^{-tH}\}_{t\geqslant 0}$ in the $t\to\infty$ limit. For simplicity, we assume a nondegenerate ground state; in the general case, Ω projects onto the lowest eigenspace.

To build a MC algorithm, we need a probabilistic representation of Ω. For each pair $s,s'\in S$ such that $s\neq s'$ and $H_{s's}\neq 0$ we define $\Gamma_{s's}=-H_{s's}$. We assume that all $\Gamma_{s's}>0$ (no sign problem), and build an S-valued Markov stochastic process s_t by identifying $\Gamma_{s's}$ as the rate for the transition $s\to s'$. Hence the average occupation $P_s(t)=\mathbf{E}(\delta)_{s,s_t}$, with $\mathbf{E}(\cdot)$ denoting the average with respect to s_t, obeys the master equation $\dot{P}_s(\beta)=\Sigma_{s'\neq s}(\Gamma_{ss'}P_{s'}-\Gamma_{s's}P_s)$.

Related to s_t, we also define the real valued stochastic process $W_t=\exp(-\int_0^t\omega_{s_t}dt)$, with $\omega_s=\Sigma_{s'\in S}H_{s's}$. It can be shown that the weighted expectation value $\psi_s(t)=\mathbf{E}(\delta_{s,s_t}W_t)$ reconstructs Ω, $(d/dt)\,\psi_s(t)=-\Sigma_{s'\in S}H_{ss'}\psi_{s'}(t)$, with $\psi_s(0)=\mathrm{Prob}(s_0=s)$. Matrix elements of Ω can be identified with certain expectation values. In particular, the ground state energy E_0 can be obtained by

$$E_0=\lim_{t\to+\infty}[\mathbf{E}(\omega_{s_t}W_t)/\mathbf{E}(W_t)], \tag{1}$$

that gives E_0 as the asymptotic average of ω_s over realizations of s_t with weight W_t, called *walkers* in the following. The actual construction of the process is straightforward. A realization of s_t is a piecewise constant map $\mathbf{R}\to S$ with isolated jumps at times $t=t_0,t_1,\ldots$, with $t_0<t_1<t_2<\cdots$. An algorithm to compute the triples $\{t_n,s_{t_n},W_{t_n}\}$ is the following.

(1) We simply denote $s_{t_n}\equiv s$ and define the set T_s of target states connected to s: $T_s=\{s',\Gamma_{s's}>0\}$. We also define the total width $\Gamma_s=\Sigma_{s'\in T_s}\Gamma_{s's}$. (2) Extract $\tau\geqslant 0$ with probability density $p_s(\tau)=\Gamma_s e^{-\Gamma_s\tau}$. In other words, $\tau=-(1/\Gamma_s)\log\xi$, with ξ uniformly distributed in $[0,1]$. (3) Extract a new state $s'\in T_s$ with probability $p_{s'}=\Gamma_{s's}/\Gamma_s$. (4) Define $t_{n+1}=t_n+\tau$, $s_{t_{n+1}}=s'$ and $W_{t_{n+1}}=W_{t_n}e^{-\omega_s\tau}$.

The above algorithm is the explicit zero imaginary time limit of power algorithms [14].

For a better performance, it is useful to introduce a trial state $|\Phi(\alpha)\rangle$ depending on some parameters α. The original Hamiltonian H is replaced by the isospectral $H_{ss'}(\alpha)=\Phi_s(\alpha)H_{ss'}\Phi_{s'}^{-1}(\alpha)$ with $\Phi_s(\alpha)=\langle s|\Phi(\alpha)\rangle$. The algorithm is unchanged [the hermiticity of $H(\alpha)$ has not been assumed], but everything, in particular ω_s, becomes α dependent. In the ideal case when $|\Phi(\alpha)\rangle$ is an exact ground state, then $\omega_s\equiv E_0$, and the ground state energy is estimated by Eq. (1) with zero fluctuations.

As is well known, a naive implementation of Eq. (1) fails because the variance of the right hand side diverges as $t\to+\infty$. A possible way out is stochastic reconfiguration (SR)

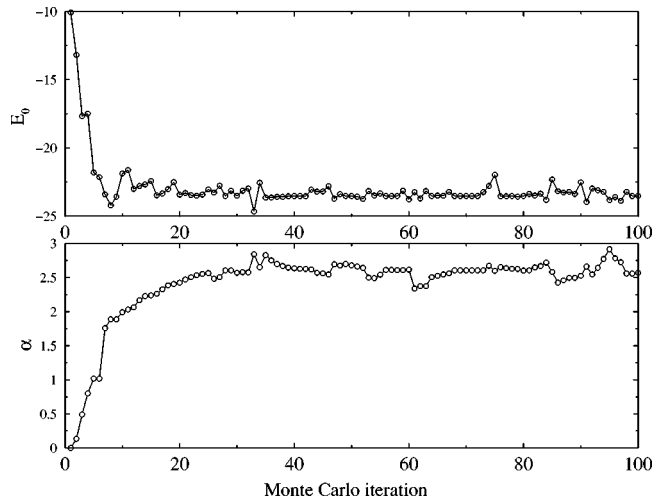

FIG. 1. $L=10$, $N_f=2$, $g=3.0$, $K=10$, and $\beta=0.5$. MC evolution of the ground state energy estimate and of the α parameter.

[11–14]. An ensemble with a large fixed number K of walkers is introduced, and a branching procedure deletes walkers with low weight and makes copies of the ones with larger weight. In the end, we take the numerical limit $K\to\infty$. If β is the time between two SR's, then we denote the estimate of the ground state energy by $\hat{E}_0(\beta,K,\alpha)$, where we do not write the dependence on physical parameters (lattice size, couplings). Usually, the dependence on α is quite strong and requires optimization to make $|\Phi(\alpha)\rangle$ the closest possible to the exact ground state.

As we remarked, a possible way to optimize α is to minimize the fluctuations of $\omega_{s_t}(\alpha)$ [15]. To this aim, following the general ideas of [16], we promote α to a sequence $\{\alpha_n\}$ and, after each SR, we compute the variance of $\omega(\alpha)$ over the K walkers, with their states kept fixed. Then we propose to update α according to

$$\alpha_{n+1}=\alpha_n-\eta_n\nabla_{\alpha_n}\mathrm{Var}\,\omega(\alpha_n). \tag{2}$$

The sequence $\{\eta_n\}$ controls the speed of the adaptive process and vanishes as $n\to\infty$, typically like n^{-1}. The novelty of the procedure is that MC sampling and trial wave function optimization are coupled. A change in α induces a change in the walker dynamical distribution, which in turn determines the next evolution of α. The whole process is nonlinear and an explicit numerical investigation is required to assess its stability.

As a specific nontrivial application, we consider the two dimensional Gross-Neveu model [17] described by the Hamiltonian

$$H=\int dx\left[-i\psi^{a\dagger}\sigma_x\partial_x\psi^a-\frac{g^2}{2N_f}(\psi^{a\dagger}\sigma_z\psi^a)^2\right], \tag{3}$$

where ψ^a are N_f Dirac fermions and we sum over the repeated flavor index $a=1,\ldots,N_f$. The model is asymptotically free, admits a $1/N_f$ expansion, and spontaneously breaks the discrete chiral Z_2 symmetry $\psi\to\gamma_5\psi$.

Following Ref. [18], a lattice formulation with staggered Kogut-Susskind fermions [19] is based on

$$H=-\sum_{n=0}^{L-1}\left\{\frac{1}{2}(c_n^{a\dagger}c_{n+1}^a+\mathrm{H.c.})+\frac{g^2}{8N_f}(c_n^{a\dagger}c_n^a-c_{n+1}^{a\dagger}c_{n+1}^a)^2\right\},$$

where $\{c_n^a,c_m^b\}=0$, $\{c_n^a,c_m^{b\dagger}\}=\delta_{n,m}\delta_{a,b}$, and periodic boundary conditions are assumed. The state space is the set

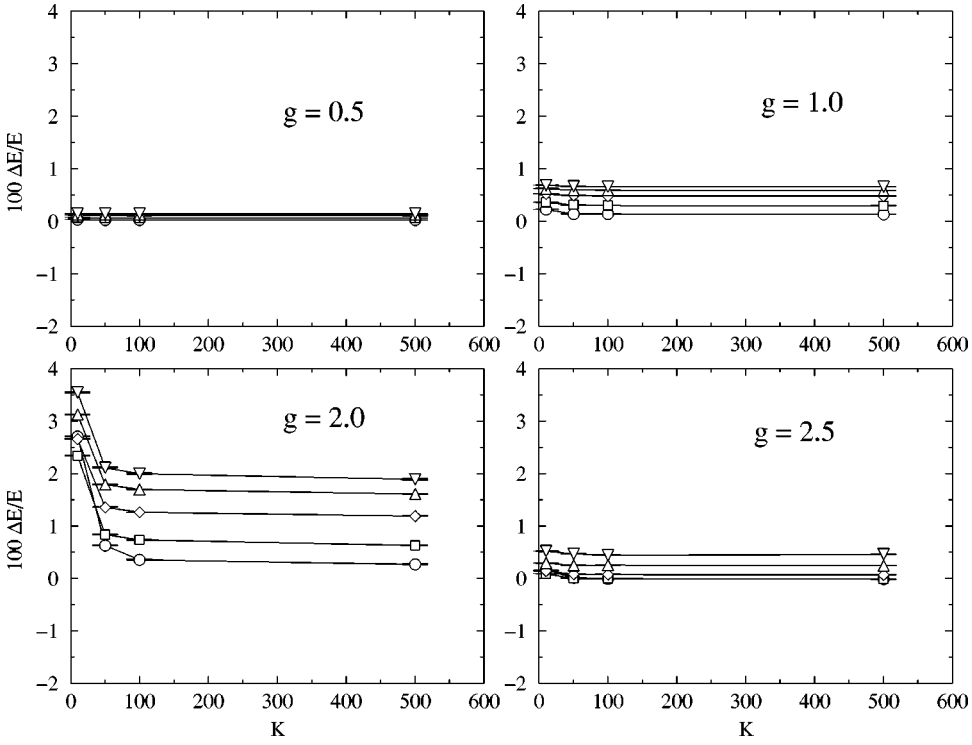

FIG. 2. $L=10$ and $N_f=2$. Relative percentual error on the ground state energy. The various lines correspond to $\beta=0.1$ (circles), $\beta=0.25$ (squares), $\beta=0.5$ (diamonds), $\beta=0.75$ (triangles up), and $\beta=1.0$ (triangles down).

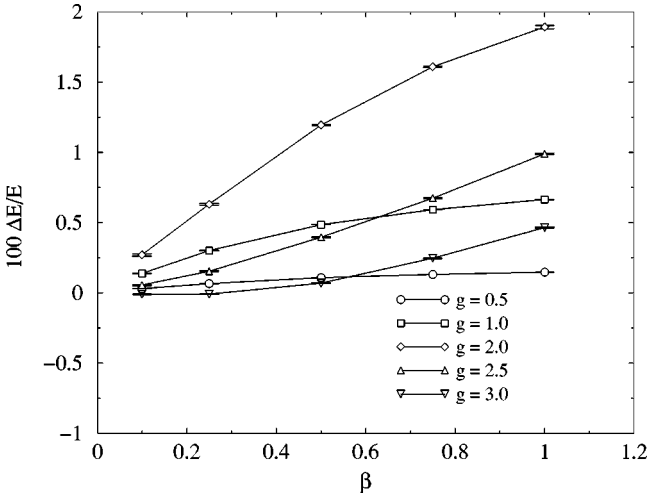

FIG. 3. $L=10$, $N_f=2$, and $K=500$. Relative percentual error on the energy obtained from data at large K at several β.

of eigenstates of the occupation number operators $n_i^a = c_i^{a\dagger} c_i^a$ denoted by $|\mathbf{n}\rangle$. The fermion number is conserved, and we focus on the half-filled sector with $\Sigma_i n_i^a = L/2$. The Z_2 symmetry corresponds to translations by two lattice sites. To avoid sign problems related to the boundary crossing, in the following we choose $L \bmod 4 = 2$ (the ground state is then nondegenerate).

We adopt the one parameter trial wave function

$$\langle \mathbf{n}|\Phi(\alpha)\rangle = \exp\left[\alpha \sum_{i=0}^{L-1} \left(\sum_{a=1}^{N_f} (n_i^a - n_{i+1}^a)\right)^2\right] \langle \mathbf{n}|g=0\rangle,$$

where $|g=0\rangle$ is the exact ground state at $g=0$. The algorithm requires an explicit formula for the ratio $\langle \mathbf{n}'|\Phi\rangle/\langle \mathbf{n}|\Phi\rangle$, where $|\mathbf{n}\rangle$ and $|\mathbf{n}'\rangle$ are states that differ by one fermion hopping. If $\{x_i\}$ and $\{x_i'\}$ are the $L/2$ fermion positions in the two states and if $x_i = x_i'$ for $i \neq p$, then the following formula can be derived:

$$\frac{\langle \mathbf{n}'|\Phi\rangle}{\langle \mathbf{n}|\Phi\rangle} = \exp\left(\frac{2\pi i}{L}\,\frac{L/2-1}{2}\,(x_p - x_p')\right) \times \frac{\prod_{k\neq p}\left(\exp\frac{2\pi i x_p'}{L} - \exp\frac{2\pi i x_k}{L}\right)}{\prod_{k\neq p}\left(\exp\frac{2\pi i x_p}{L} - \exp\frac{2\pi i x_k}{L}\right)}.$$

TABLE I. E_0/N_f for the $L=10$ model with $N_f=2$ flavors. $\Delta E = E_0^{Lanczos} - E_0^{MC}$.

g	$\alpha^*(500,0.1)$	Exact Lanczos diagonalization	Polynomial extrapolation	$1000\,\lvert\Delta E/E\rvert$
0.5	0.07638(1)	−3.34904	−3.34908(5)	0.012
1.0	0.31347(5)	−3.71687	−3.71689(5)	0.005
2.0	1.4044(3)	−5.99265	−5.9929(5)	0.03
2.5	2.0575(2)	−8.4526	−8.4524(3)	0.02
3.0	2.6198(2)	−11.6949	−11.6927(3)	0.2

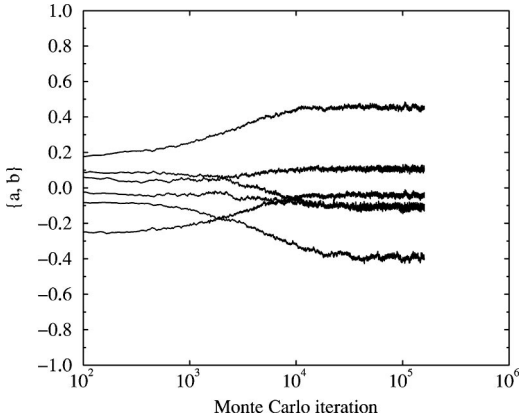

FIG. 4. $L=10$, $N_f=2$, and $g=2.0$. MC evolution of the six parameters $\{a,b\}$. From top to bottom, on the right of the plot, the parameters are a_1, a_3, a_2, b_1, b_2, and b_0.

We compute the ground state energy on a lattice with $L=10$ sites, and begin our analysis with the case $N_f=2$. We consider several ensemble sizes and evolution times: $K=10$, 50, 100, and 500 and $\beta=0.1$, 0.25, 0.5, and 1.0. For each pair (K,β) we determine, by the adaptive algorithm, the best α, and estimate the ground state energy. For comparison, we also determine E_0 by exact Lanczos diagonalization.

Figure 1 shows the typical initial steps of a run. The parameter α and the energy measurements evolve and fluctuate around (K,β) dependent definite average values $\alpha^*(K,\beta)$ and $\hat{E}_0(K,\beta,\alpha^*(K,\beta))$. For large K, the statistical error on $\hat{E}_0$ decreases like $K^{-1/2}$. For $K\to\infty$, the results are expected to be β independent. However, for moderate ensemble sizes, like those considered ($K\sim 500$), a residual β dependence can be observed, particularly at intermediate coupling, as shown in Fig. 2. This effect is due to the process of walker selection associated with SR. The correct approach is to take the $\beta\to 0$ limit where this effect is expected to be negligible. In Fig. 3, we plot $\hat{E}_0[\beta,500,\alpha^*(500,\beta)]$ as β and g are varied.

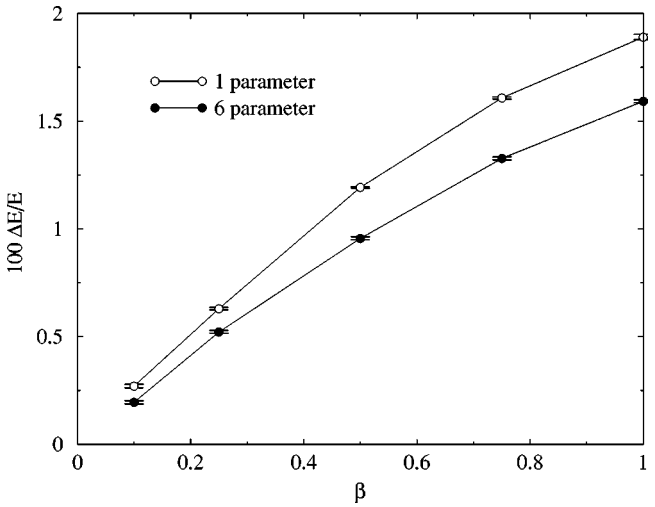

FIG. 5. $L=10$, $N_f=2$, $g=2.0$, and $K=500$. Improvement in the energy estimate with the six-parameter trial wave function.

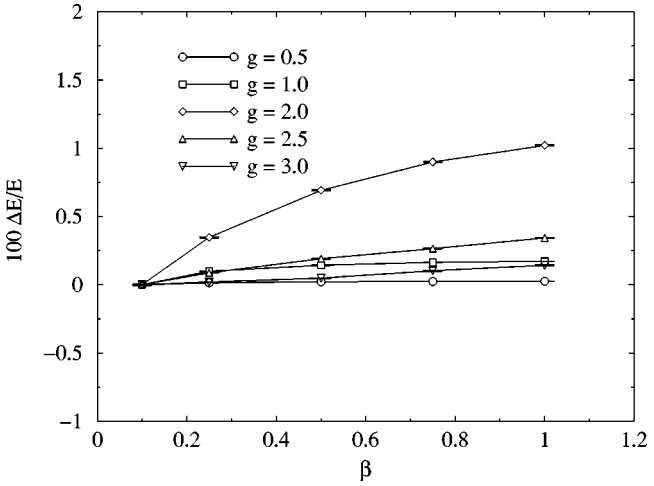

FIG. 6. $L=10$, $N_f=6$, and $K=500$. Relative percentual error on the energy estimate obtained from data at large K at several β.

All the curves converge to zero and, in fact, can be smoothly extrapolated to $\beta\to 0$. The resulting percentual relative error $100|E_0-\hat{E}_0|/|E_0|$ is very small, well below the permille level (see Table I for numerical results with fourth-order polynomial extrapolation).

For large coupling g, the convergence is quite fast. The one-parameter trial wave function is accurate because the ground state is dominated by states with low potential that are easily selected by $|\Phi(\alpha)\rangle$. Relatively small K's are then already in the asymptotic regime. For intermediate couplings, $g\sim 2.0$, the convergence is again smooth, but less than linear. For smaller couplings, a good convergence is observed, and in fact a precise wave function can obtained with $\alpha^*\simeq 0$. The optimal α^* at $K=500$, $\beta=0.1$ is shown in Table I.

For $g=2.0$, we explore a tentative six-parameter trial wave function. Denoting the two fermion flavors by $\uparrow$ and $\downarrow$, we use $\langle \mathbf{n}|\Phi\rangle = e^{\Sigma_i F_i}\langle \mathbf{n}|g=0\rangle$ with $F_i=\Sigma_{k=1}^{3} a_k(n_i^{\uparrow}n_{i+k}^{\uparrow} + \uparrow\leftrightarrow\downarrow) + \Sigma_{k=0}^{2} b_k(n_i^{\uparrow}n_{i+k}^{\downarrow} + \uparrow\leftrightarrow\downarrow)$. The MC automatic determination of the six parameters is shown in Fig. 4. The algorithm converges to definite coefficients $\{a,b\}$, but the behavior of $\hat{E}_0$ does not dramatically improve (Fig. 5). Nonetheless, some qualitative remarks can be stressed, as the presence of long range correlations between next to neighbor fermions with the same spin and anticorrelations between fermions with opposite spin.

Since the Gross-Neveu model can be studied nonperturbatively in the framework of the $1/N_f$ expansion, it is interesting to analyze the algorithm performance with a larger number of flavors. In Fig. 6, we show the results for $N_f=6$. The exact value is beyond Lanczos diagonalization, and we choose to normalize errors at the $\beta=0.1$, $K=500$ value. A comparison with Fig. 3 reveals that the error as well as its β dependence are rather reduced with respect to the previous $N_f=2$ case.

In summary, our data show that a clever extrapolation in the algorithm free parameters K and β allows accurate results even with small walker ensembles. This is an important feature for realistic large scale simulations aimed at reaching the continuum limit. Results with large N_f suggest that the present algorithm can be a viable numerical technique for other fermionic two-dimensional models where the $1/N_f$ expansion applies, like the important case of models with dynamical supersymmetric breaking [20]. In principle, extensions to models with sign problems are possible and, in fact, progress in the optimization issue was recently proposed [21] within the considered class of MC algorithms.

[1] K. Wilson, Phys. Rev. D **10**, 2445 (1974).

[2] J. B. Kogut and L. I. Susskind, Phys. Rev. D **11**, 395 (1975); J. B. Kogut, Rev. Mod. Phys. **51**, 659 (1979).

[3] E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar, Phys. Rev. B **41**, 9301 (1990).

[4] W. von der Linden, Phys. Rep. **220**, 53 (1992).

[5] M. Beccaria, Phys. Rev. D **61**, 114503 (2000); C. Long, D. Robson, and S. Chin, *ibid.* **37**, 3006 (1988).

[6] M. Beccaria, Phys. Rev. D **62**, 034510 (2000).

[7] D. M. Ceperley and M. H. Kalos, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer-Verlag, Berlin, 1992).

[8] D. Ceperley, G. Chester, and M. Kalos, Phys. Rev. B **16**, 3081 (1977); E. Koch, O. Gunnarsson, and R. M. Martin, Phys. Rev. B **59**, 15 632 (1999).

[9] G. F. De Angelis, G. Jona-Lasinio, and M. Sirugue, J. Phys. A **16**, 2433 (1983).

[10] M. Beccaria, G. F. De Angelis, C. Presilla, and G. Jona-Lasinio, Nucl. Phys. **B83-84**, 911 (2000); M. Beccaria, G. F. De Angelis, C. Presilla, and G. Jona-Lasinio, Europhys. Lett. **48**, 243 (1999).

[11] J. H. Hetherington, Phys. Rev. A **30**, 2713 (1984).

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

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

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

[15] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988); P. R. C. Kent, R. J. Needs, and G. Rajagopal, Phys. Rev. B **59**, 12 344 (1999).

[16] E. Koch, O. Gunnarsson, and R. M. Martin, Phys. Rev. B **59**, 15 632 (1999).

[17] D. J. Gross and A. Neveu, Phys. Rev. D **10**, 5235 (1974).

[18] A. Zee, Phys. Rev. D **12**, 3251 (1975); J. Shigemitsu and S. Elitzur, *ibid.* **14**, 1988 (1976); J. E. Hirsch, R. L. Sugar, D. J. Scalapino, and R. Blankenbecler, Phys. Rev. B **26**, 5033 (1982); Y. Cohen, S. Elitzur, and E. Rabinovici, Phys. Lett. **104B**, 289 (1981); Nucl. Phys. **B220**, 102 (1983).

[19] L. Susskind, Phys. Rev. D **16**, 3031 (1977).

[20] J. L. Goity, Nucl. Phys. **B269**, 587 (1986); T. Uematsu, Phys. Lett. **123B**, 209 (1983); **126B**, 459 (1983).

[21] S. Sorella, Phys. Rev. B **64**, 024572 (2001).