Computer algorithm for many-body wave functions

J. B. Bronzan and Timothy E. Vaughan
Department of Physics and Astronomy, Rutgers University, P.O. Box 849, Piscataway, New Jersey 08855
(Received 14 November 1988)

We develop and test an algorithm for the computation of a fixed number of the lowest eigenstates of a Hermitian operator. The algorithm is iterative and converges exponentially. We test the algorithm on the strong-coupled Hubbard model, which has a finite number of basis states, and we reproduce known results. We also compute with a partial basis, a subset of a complete basis. We show how to extrapolate the eigenvalues to the full basis result. This opens the possibility of applying the algorithm to problems having an infinite number of basis states.

I. INTRODUCTION

In this paper we present algorithms for the construction of the low-lying wave functions and eigenvalues of an Hermitian operator. The algorithm can be applied to systems having many degrees of freedom, such as lattice field theory or condensed-matter theories involving interacting electrons.

Our algorithms for wave functions are iterative, and we may describe the simplest or primary algorithm as follows. Imagine we are equipped with a set of basis states, and a linear combination of them which approximates the ground state of the Hamiltonian. One iterative step consists of taking one basis state and combining it with the current approximate ground state to make an "optimal" improvement. A desirable feature of the algorithm, which it shares with Monte Carlo procedures,[1] is that it has fixed complexity. The algorithm converges exponentially with n, the number of iterations. This is much faster than Monte Carlo convergence, which goes like $n^{-1/2}$, where now n is the number of uncorrelated configurations that have been generated. An important feature of our algorithms is that fermions pose no special problems.

The main unfavorable characteristic of the wave-function algorithms is that they converge to exact eigenvectors and eigenvalues only if the complete set of basis states is small enough to allow vectors expressed as linear combinations of them to be stored in the computer. This requirement is not usually met in the case of systems having bosonic degrees of freedom; then the number of basis states often is infinite. Even in purely fermionic or bosonic spin systems, the number of states in a small system can overwhelm computer storage capacity. If a system has F degrees of freedom, each of which can occupy K states, there are K^F states in the basis. Thus, for the spin-$\frac{1}{2}$ Heisenberg model on a 6×6 square lattice, $K^F=2^{36}\simeq7\times10^{10}$.

In light of this problem, a conservative way to view the algorithms is as a variational method incorporating enormous numbers of states. However, we wish to do better than that, so a major part of our effort has been devoted to the problem of obtaining good results with a finite number of basis states. The practical utility of our algorithms is greatly increased if we succeed. Our strategy begins with the observation that it may be possible to choose a finite basis whose states nearly saturate the norms of the low-lying exact eigenstates. This will be the case when we have physical or mathematical reasons to believe that states in a certain class are dominant. Let us suppose we have chosen a basis set following such a motivation, or, failing that, because the set simplifies the computation of matrix elements. The next task is to identify the subset of the basis that is optimal for the representation of low-lying states (and will fit into memory). In doing this we should proceed in an automatic and bias-free way, abandoning the physical or mathematical prejudices that led to the choice of basis.

These ideas are implemented as follows. Starting with a finite, or partial basis, we use the primary algorithms described above to obtain the best approximations to the low-lying state vectors that can be formed from the current partial basis. Then we use secondary algorithms to "weed" out some of the least important states in our basis, which we replace by randomly chosen new states from the complete basis. We then use the primary algorithms to obtain the best approximations possible with the modified partial basis. The process is repeated until the partial basis shows no further secular change. If our original prejudice about which states are dominant was correct, the partial basis we arrive at will include those putatively dominant states.

At this point, we have done as well as we can using a subset of N_0 basis states. Let $E(N_0)$ be the resulting energy of some low-lying state. In general, $E(N_0)$ will depend on N_0, and we will need to extrapolate to $N_0=N$, where N is the number of states in the complete basis. ($N=\infty$ for some problems with bosonic degrees of freedom.) In making the extrapolation we must know how $E(N_0)$ varies with N_0, and in this we are aided by a remarkable coincidence. As we stated above, the convergence of the primary algorithm is exponential. If we let $\mathcal{E}(n)$ be the deviation of the state energy from its final value after n iterations,

$$\mathcal{E}(n)=\mathcal{E}_0 e^{-nC/N_0}, \tag{1.1}$$

as we will show in Sec. IV. Here C is a constant that depends upon the global properties of the Hamiltonian. C can be measured by observing the convergence of the primary algorithm using the subset of N_0 optimal states.

Now consider the extrapolation in N_0. In Sec. V we

will show that the state energy has the asymptotic dependence

$$E(N_0)=A+B(N_0)^{-C}\,, \tag{1.2}$$

where C is the *same* constant that appears in Eq. (1.1). This means that the exponent controlling the extrapolation can be measured within the subset calculation.

The layout of the rest of the paper is as follows. In Sec. II we present the primary algorithm for wave functions. This algorithm is a generalization of the algorithm described above in that it permits us to compute a fixed number of the lowest eigenvectors and eigenvalues iteratively.

In Sec. III we introduce the Hubbard model, which we use as a test of our algorithm. The Hubbard model has the desirable features of a finite but large basis and partially known energy levels. We can therefore test our algorithms against known results. Since we know the eigenvalues to high accuracy, we can test the extrapolation of results obtained with a partial basis. In Sec. III we also discuss the problems posed to our algorithm by degenerate states. We argue that the use of a basis belonging to a row of an irreducible representation of the symmetry group of the Hamiltonian will eliminate degenerate states. We compare our results with previous work.

Section IV is devoted to the convergence of the algorithm. Simple theoretical considerations suggest the convergence is given by Eq. (1.1), and we verify that the Hubbard model computations follow this law over seven decades of variation.

In Sec. V we discuss the choice and use of a finite subset of basis states. We give a derivation of the criterion used in the weeding process, where we change the states in the subset retained in storage. We apply these ideas to the Hubbard model, where we recompute the lowest four eigenvectors and eigenvalues for several values of N_0. We then derive Eq. (1.2) and test the extrapolation to the exact energies.

A summary and conclusion are presented in Sec. VI.

II. PRIMARY ALGORITHM

Our algorithms are based on an extension of the Rayleigh-Ritz variational principle. Let H be an Hermitian operator, and let $\{\phi_a\}$ be a set of K orthonormal states. Order the eigenvalues of the $K\times K$ Hermitian matrix $H_{a,b}=\langle\phi_a|H|\phi_b\rangle$ in increasing order: $\lambda_1\le\cdots\le\lambda_K$. Then each of these eigenvalues is an upper bound to the corresponding ordered eigenvalue of H. In addition, let

$$\begin{bmatrix} c_1 \\ \vdots \\ c_K \end{bmatrix} \tag{2.1}$$

be an eigenvector of $H_{a,b}$. The corresponding approximate eigenvector of H is

$$\psi=\sum_{a=1}^{K} c_a\phi_a\,. \tag{2.2}$$

We start the computation by using the variational principle for $K=1$ with one of the basis states discussed in Sec. I. This gives an upper bound to the ground-state energy. We then add a different basis state and proceed with $K=2$. We continue adding states and diagonalizing $H_{a,b}$ until we reach $K=r$, where r is the preassigned number of eigenstates we wish to construct. In the subsequent iterations we retain the r lowest eigenvalues and eigenvectors, and discard the $(r+1)$st entities.

As we range through the basis in some fashion (to be discussed below), we will eventually choose a basis state $\bar{\phi}$ encountered earlier. $\bar{\phi}$ is not orthogonal to the current approximate eigenvectors ψ^a $(a=1,\ldots,K)$, as required by the variational principle. We therefore replace $\bar{\phi}$ by its orthonormal projection

$$\bar{\phi}_\perp=\left[1-\sum_{a=1}^{K}|\langle\psi^a|\bar{\phi}\rangle|^2\right]^{-1/2}\left[\bar{\phi}-\sum_{b=1}^{K}\psi^b\langle\psi^b|\bar{\phi}\rangle\right], \tag{2.3}$$

and use this state to continue the process.

An important feature of our algorithm is that as we add states it is easy to diagonalize $H_{a,b}$ because it is the sum of diagonal and rank one matrices. $H_{a,b}$ has the form

$$H_{a,b}=\begin{bmatrix} E_1 & 0 & \cdots & H_{1,K+1} \\ 0 & E_2 & \cdots & H_{2,K+1} \\ \vdots & \vdots & \ddots & \vdots \\ H_{K+1,1} & H_{K+1,2} & \cdots & H_{K+1,K+1} \end{bmatrix}. \tag{2.4}$$

Here we denote the added state by ψ^{K+1}, and $H_{a,K+1}=\langle\psi^a|H|\psi^{K+1}\rangle$. It follows that

$$c_a=\frac{H_{a,K+1}}{\lambda-E_a}c_{K+1}\quad(a=1,\ldots,K)\,. \tag{2.5}$$

The $(K+1)$st row of $H_{a,b}$ leads to the characteristic equation

$$\sum_{a=1}^{K}\frac{|H_{a,K+1}|^2}{\lambda-E_a}+E_{K+1}-\lambda=0\,, \tag{2.6}$$

where $E_{K+1}=H_{K+1,K+1}$. This can be converted into a polynomial of degree $K+1$, and so has $K+1$ solutions. However, Eq. (2.6) is illuminating, for inspection shows that one of the E's lies between successive solutions,

$$\lambda_1\le E_1\le\lambda_2\le\cdots\le E_K\le\lambda_{K+1}\,. \tag{2.7}$$

The fact that roots are bracketed in this way makes it easy to devise root-finding routines for them.[2]

We see from Eq. (2.7) that the lowest r roots decrease monotonically as we proceed. As we iterate, the r roots must approach eigenvalues of H. The argument is that the roots are monotonically decreasing and bounded from below, so they must approach limits. At the limits, $H_{a,K+1}=0$ for any state we try to add, so ψ^a is an eigenvector of H. In principle, the roots need not approach the r lowest eigenvalues, but we have never encountered anything *other* than the r lowest states. In many distinct runs with the Hubbard-Heisenberg model we always found the same roots, and comparison with previous

work shows that λ_1 is indeed the ground-state energy.

The new mth eigenstate is

$$\tilde{\psi}^m = \left[1+\sum_{a=1}^{K}\frac{|H_{a,K+1}|^2}{(\lambda_m-E_a)^2}\right]^{-1/2} \times\left[\psi^{K+1}+\sum_{b=1}^{K}\frac{\psi^b H_{b,K+1}}{\lambda_m-E_b}\right]. \tag{2.8}$$

This formula can lead to computational round-off errors in later iterations, where the approximate eigenvectors are close to true eigenvectors. Then $|H_{a,K+1}| \ll \min(E_{b+1}-E_b)$, and by Eq. (2.6), $\lambda_m = E_q + O(|H|^2)$. q takes on values m, $K+1$, or $m-1$ depending on the relation between E_{K+1} and the other E_a's. If $q=K+1$, Eq. (2.8) has no small denominators, and is suitable for computation. For the other m's, denominators in the qth terms in Eq. (2.8) are dangerously small, and must be recovered. This can be done by using Eq. (2.6) to eliminate $(\lambda_m - E_q)^{-1}$. The improved formula is[3]

$$\tilde{\psi}^m = N\left[\psi^q+\frac{H_{K+1,q}}{P}\left[\psi^{K+1}+\sum_{\substack{a=1\\a\neq q}}^{K}\frac{\psi^a H_{a,K+1}}{\lambda_m-E_a}\right]\right],$$

$$P=\lambda_m-E_{K+1}-\sum_{\substack{a=1\\a\neq q}}^{K}\frac{|H_{a,K+1}|^2}{\lambda_m-E_a}, \tag{2.9}$$

$$N=\left[1+\frac{|H_{q,K+1}|^2}{P^2}\left[1+\sum_{\substack{a=1\\a\neq q}}^{K}\frac{|H_{a,K+1}|^2}{(\lambda_m-E_a)^2}\right]\right]^{-1/2}.$$

III. APPLICATION TO THE HUBBARD MODEL

The Hamiltonian of the Hubbard model[4] is

$$H = tK + U\sum_i c_{i+}^\dagger c_{i+} c_{i-}^\dagger c_{i-},$$

$$K = -\sum_{\langle ij\rangle,\sigma}(c_{i\sigma}^\dagger c_{j\sigma} + c_{j\sigma}^\dagger c_{i\sigma}), \tag{3.1}$$

where $\langle ij\rangle$ denotes a sum over nearest-neighbor electrons. Apart from an additive constant, the hopping term K is the kinetic energy, and the term U is a repulsive Coulomb interaction between electrons on the same site.

We will study this model on a 4×4 square lattice with periodic boundary conditions. On the lattice we place eight spin-up and eight spin-down electrons (half-filling). The number of spin-up (or -down) electrons is conserved by H, and each group can be distributed in $16!/(8!)^2 = 12\,870$ ways. Therefore, this subspace of Hilbert space is spanned by 165 636 900 states. However, we will study the Hubbard model in the limit $t \ll U$, where K can be treated as a perturbation. To zeroth order in K, the ground state is degenerate, with every degenerate state having exactly one electron per site. There are 12 870 of these degenerate states, which have zero energy, and which we denote by $|a\rangle$. The basis set is of a convenient size to test our algorithms when we use either complete or truncated bases.

Application of K to $|a\rangle$ produces a state having two electrons on one site and, therefore, energy U. This means that K lifts the degeneracy of the perturbative ground state in the *second* order of perturbation theory, not the first. To leading order in t/U, the effective Hamiltonian is the $12\,870\times12\,870$ matrix

$$\tilde{H}_{a,b} = -\frac{t^2}{U}\langle a|K^2|b\rangle. \tag{3.2}$$

Equation (3.2) is convenient for programming, but in order to compare our results with previous work, we note that it is equivalent to the spin-$\frac{1}{2}$ Heisenberg model.[5] The terms in K^2 contributing to $\tilde{H}_{a,b}$ are

$$K^2 \simeq \sum_{\langle ij\rangle,\sigma}(c_{i\sigma}^\dagger c_{j\sigma} c_{j\sigma}^\dagger c_{i\sigma} + c_{i,-\sigma}^\dagger c_{j,-\sigma} c_{j\sigma}^\dagger c_{i\sigma} + c_{j\sigma}^\dagger c_{i\sigma} c_{i\sigma}^\dagger c_{j\sigma} + c_{j,-\sigma}^\dagger c_{i,-\sigma} c_{i\sigma}^\dagger c_{j\sigma}). \tag{3.3}$$

Introduce spin operators for each site

$$s_{ix} \pm s_{iy} = c_{i\mp} c_{i\pm}^\dagger,$$
$$s_{iz} = (\tfrac{1}{2})(c_{i+}^\dagger c_{i+} - c_{i-}^\dagger c_{i-}). \tag{3.4}$$

These operators satisfy angular momentum commutation rules. On our degenerate ground-state subspace, where there is one electron per site, $s_{ix}^2 = s_{iy}^2 = s_{iz}^2 = \frac{1}{4}$, and matrix elements of K^2 may be replaced by matrix elements of spin operators. In this way we obtain the Hamiltonian of the antiferromagnetic Heisenberg model

$$\tilde{H} = \frac{4t^2}{U}\sum_{\langle ij\rangle}(\mathbf{s}_i\cdot\mathbf{s}_j - \tfrac{1}{4}). \tag{3.5}$$

The strong-coupling Hubbard model, which we call the Hubbard-Heisenberg model, describes an antiferromagnetic insulator. It is interesting in its own right because it is believed that materials exhibiting high-T_c superconductivity may possess an antiferromagnetic insulator phase.

The simplest application of our algorithm is to obtain the ground state of the Hubbard-Heisenberg model using the full basis of 12 870 states. Some features of the program are worthy of comment. First, we compute matrix elements of $\tilde{H}$ as required, and they are simple: $\tilde{H}_{a,b} = 2t^2/U$ if $|a\rangle$ and $|b\rangle$ differ by the exchange of one pair of neighboring spin-up and spin-down electrons; $\tilde{H}_{a,a} = -2t^2 n/U$, where n is the number of states permitted by the Pauli principle that arise from $|a\rangle$ when one of the spin-up electrons is displaced by one site; otherwise $\tilde{H}_{a,b} = 0$.

Second, when we add a basis state, we choose it as follows. We randomly select a basis state *already* contributing to our approximate eigenvectors. Within this state we randomly select one of the eight spin-up electrons, and examine the nearest-neighbor site in a random direction. If we find a spin-down electron there, we interchange the pair; this is the state we add. If we find a spin-up electron there, we start the selection process over. This protocol has the advantage that $\langle\psi^a|\tilde{H}|\psi^{K+1}\rangle \neq 0$ (except accidentally). As a result, each

iteration advances the computation even though $\tilde{H}_{a,b}$ is a sparse matrix.

The third point is that we periodically update our work by renormalizing the approximate eigenvector and recomputing the energy as $\langle\psi|\tilde{H}|\psi\rangle$. Updating interrupts the round-off errors that accumulate in an iterative computation. It is expensive in computer time: In the course of an iteration we compute a column of $\tilde{H}$, but updating requires the whole matrix. If we update every 10 000 iterations, we spend about half the CPU time updating. We have done this in our work, but the corrections we find indicate that we could update considerably less frequently except near the beginning of a run.

Using the setup discussed above, we ran for 290 000 iterations, at which point the approximate ground-state energy was -76.80 in units of t^2/U. The energy was still changing, and we will see later that the true ground-state energy is $-76.913\,932\,8337$. For reference, the approximate energies after 10 000, 66 000, and 183 000 iterations were -71.25, -75.31, and -76.47, respectively. The computation took about 37 h of CPU time on our Sun 4/280 computer.

If one tries to repeat the calculation for several states, one encounters degenerate states which cause the program to crash. We have not attempted to develop a formalism to cope with this directly; rather, we adopt the view that degeneracy arises because of symmetries of the Hamiltonian. When we take these symmetries into account, we can reduce the Hamiltonian to block diagonal form. Within the blocks, the states are nondegenerate. This approach has the additional advantage that each block is spanned by a basis of reduced size.

We construct the linear combinations of basis states $|a\rangle$ that span the blocks using group theory. Let G be the symmetry group of the Hamiltonian, in this case the space group of a square lattice with periodic boundary conditions. Let $D_{ij}^{(l)}(G)$ be an irreducible representation, and $|a\rangle$ one of the basis states. Then the linear combination

$$|l,i;a\rangle = N\sum_{G} D_{ii}^{(l)}(G)^*\hat{G}|a\rangle \tag{3.6}$$

belongs to the ith row of representation l. When we let $|a\rangle$ range through the basis, we generate a complete basis of states for the block labeled by (l,i). Blocks having the same l but different i have the same spectra; therein lies the degeneracy. On the other hand, *within* a block there can be only accidental degeneracy.

It is not difficult to implement Eq. (3.6),[6] but for our purposes it suffices to remove most of the degeneracy by replacing G by the translation subgroup of G. In this case, we use the linear combinations

$$|p,q;a\rangle = N\sum_{j=0}^{3}\sum_{k=0}^{3}\exp\left[-\frac{i\pi}{2}(jp+kq)\right]\times(\hat{T}_x)^j(\hat{T}_y)^k|a\rangle\ . \tag{3.7}$$

Here $0\le p, q\le 3$, and $\hat{T}_i$ translates the electrons in $|a\rangle$ by one site in direction i. The point group operations in the space group have been ignored here, and many of the momentum states of Eq. (3.7) are transformed into each other by these operations. As a result, these states are degenerate in energy, and it suffices to consider only one member of each coset. The independent momentum channels can be taken to be those labeled by $(p,q)=(0,0)$, $(0,1)$, $(0,2)$, $(1,1)$, $(1,2)$, and $(2,2)$. The numbers of states in these channels are 822, 800, 816, 800, 800, and 816, respectively. The ground state is in the (0,0) channel.

The point subgroup of the space group is eight dimensional, and has four one-dimensional irreducible representations and one two-dimensional representation. This opens the possibility of a residual twofold degeneracy within the independent momentum channel blocks. Actually, this can occur only within the (0,0) and (2,2) blocks. The reason is that it is a *subgroup* of the point group—the group of operations that leave the momentum vector invariant up to a multiple of the reciprocal lattice vector—that appears in the irreducible representations of G.[6] This subgroup has a two-dimensional representation only for the (0,0) and (2,2) blocks; in those blocks the "subgroup" is the entire point group.

We have used our algorithm to recompute the ground-state energy in the (0,0) block, which has a basis of 822 states. This leads to the 12 significant figure energies cited above. We ran the program for 100 000 iterations, but the energy did not change after 78 000 iterations. At that point, every state had been visited 95 times on the average. The energy was accurate to three significant figures after 18 000 iterations, and to five significant figures after 30 000 iterations. In Sec. IV we discuss the convergence of the algorithm more thoroughly.

The energy we find for the Hubbard model implies that the energy per spin of the Heisenberg model is $1.403\,560\,401$ in units of $-2t^2/U$. This agrees with the value computed by Oitmaa and Betts,[7] and we also reproduce the ground-state pair correlations listed in Table 2 of Ref. 7.

We next calculated the energies of the four lowest states in all blocks. None of the states in the (0,0) and (2,2) channels are degenerate, which means that the dangerous representations of the point group are not found among these first few states. The results are given in Table I, where the number of significant figures reflect the accuracy attained after 10 000 iterations. In these calculations we have updated after every 1000 iterations by rediagonalizing $\tilde{H}$ in the subspace spanned by the four approximate eigenvectors. It took a little more than 1 h

TABLE I. Energies of the lowest four states of the Hubbard-Heisenberg model in momentum channel (p,q) after 10 000 iterations. Energies are given through the significant figure that will *not* change in the next 100 iterations.

Momentum	E_1	E_2	E_3	E_4
(0,0)	−76.913	−70.063	−64.94	−61.75
(0,1)	−67.174	−63.49	−63.060	−61.549
(0,2)	−66.096	−65.00	−61.84	−61.496
(1,1)	−66.0542	−63.23	−61.83	−61.70
(1,2)	−67.536	−63.1	−62.50	−60.2
(2,2)	−74.5995	−63.329	−63.222	−52.45

of CPU time on the Sun to perform the 10 000 iterations in each channel.

IV. CONVERGENCE

Here we consider the convergence of the primary algorithm introduced in Sec. II. We work with a basis of N states, which can be either the complete basis for a block of H, or a partial basis of the type discussed in Sec. V. At any point in the iteration, we have K states which diagonalize the matrix $H_{a,b}$ introduced in Sec. II,

$$\langle \psi^a|H|\psi^b\rangle=\delta_{a,b}E_a \ . \tag{4.1}$$

From our basis we now choose a new state ψ^{K+1} orthonormal to these states, following which the new eigenvalues λ satisfy Eq. (2.6). We suppose that we have iterated enough to assure that adding the new state provides only a small shift in the energy levels. This will be the case when $\langle \psi^{K+1}|H|\psi^{K+1}\rangle=E_{K+1}>E_p$, and $|H_{a,K+1}|\ll\min(E_{p+1}-E_p)$, where $p=1,2,\ldots,K$. Then we can solve Eq. (2.6) perturbatively for the shift in the pth energy level,

$$\lambda_p-E_p=\Delta\mathcal{E}=-\frac{|H_{p,K+1}|^2}{E_{K+1}-E_p} \ . \tag{4.2}$$

$\mathcal{E}\equiv E_p-\bar{E}_p$ is the deviation of the current approximate energy from the true eigenvalue $\bar{E}_p$ on the space spanned by the N state basis.

The pth state wave function used in the matrix elements above may be written

$$\psi^p=(1-\epsilon^2)^{1/2}\bar{\psi}^p+\epsilon\chi^p \quad (0<\epsilon<1) \ , \tag{4.3}$$

where $\bar{\psi}^p$ is the true pth state wave function, and χ^p is orthonormal to $\bar{\psi}^p$. ϵ is assumed to be small already, and in the following we work to lowest order in ϵ. When we iterate, the added state may be decomposed similarly

$$\psi^{K+1}=\epsilon^{i\phi}(1-\alpha^2)^{1/2}\bar{\psi}^p+\alpha\chi^{K+1} \quad (0<\alpha<1) \ . \tag{4.4}$$

Using these expressions, to lowest order in ϵ,

$$\begin{aligned}\Delta\mathcal{E}&=-\epsilon^2\frac{|\langle\chi^p|(H-\bar{E}_p)|\chi^{K+1}\rangle|^2}{\langle\chi^{K+1}|(H-\bar{E}_p)|\chi^{K+1}\rangle} \ ,\\ \mathcal{E}&=\epsilon^2\langle\chi^p|(H-\bar{E}_p)|\chi^p\rangle \ .\end{aligned} \tag{4.5}$$

Define

$$\phi^q=(H-\bar{E}_p)\chi^q, \quad \hat{\phi}^q=\frac{\phi^q}{\langle\phi^q|\phi^q\rangle^{1/2}} \ . \tag{4.6}$$

Then we may write

$$\frac{\Delta\mathcal{E}}{\mathcal{E}}=-\frac{|\langle\chi^p|\hat{\phi}^{K+1}\rangle\langle\chi^{K+1}|\hat{\phi}^p\rangle|}{\langle\chi^p|\hat{\phi}^p\rangle\langle\chi^{K+1}|\hat{\phi}^{K+1}\rangle} \ . \tag{4.7}$$

The inner products in Eq. (4.7) may be written

$$\begin{aligned}&|\langle\chi^p|\hat{\phi}^{K+1}\rangle|=C_1/\sqrt{N}, \quad |\langle\chi^{K+1}|\hat{\phi}^p\rangle|=C_2/\sqrt{N} \ ,\\ &\langle\chi^p|\hat{\phi}^p\rangle=C_3, \quad \langle\chi^{K+1}|\hat{\phi}^{K+1}\rangle=C_4 \ .\end{aligned} \tag{4.8}$$

Consider $|\langle\chi^p|\hat{\phi}^{K+1}\rangle|$ first. If $\hat{\phi}^{K+1}$ were a random unit vector, we would expect C_1 to be one. This statement means that, while C_1 would vary from iteration to iteration, its average over many iterations would be one. However, the operator $H-\bar{E}_p$ in Eq. (4.6) has the effect of aligning $\hat{\phi}^{K+1}$ along the highly excited eigenstates of H, so that while χ^{K+1} varies randomly, $\hat{\phi}^{K+1}$ does not. The behavior of χ^p is quite different: When the iteration process has been under way for some time, χ^p should be aligned along the low-lying eigenstates of H. Therefore, we expect the scalar product to be less than the random value, or $C_1<1$. If H is ill conditioned, we might expect C_1 to be especially small.[8] The same considerations apply to C_2. C_3 and C_4 are less than one also.

Equation (4.7) leads to the differential equation

$$\frac{d\mathcal{E}}{\mathcal{E}}=-\frac{C\,dn}{N} \ , \tag{4.9}$$

where $C=C_1C_2/C_3C_4$, and n is the number of iterations. We have already noted that C_1 to C_4 can vary significantly from iteration to iteration, so that C will fluctuate with n. We ignore these fluctuations, so Eq. (4.9) gives the evolution of $\mathcal{E}(n)$ averaged over many iterations. In principle, C can also depend on n in a secular fashion because as the iteration proceeds, the vectors χ^p and $\hat{\phi}^p$ become more closely aligned along the low-lying excited states. Note that C is an entity that depends on the structure of the Hamiltonian.

In practice, C has little or no significant variation while $\mathcal{E}$ is changing over many decades. (In Sec. V we discuss one exception to this finding.) Therefore, we are justified in treating C as a constant when we solve Eq. (4.9),

$$\mathcal{E}(n)=\mathcal{E}_0e^{-nC/N} \ . \tag{4.10}$$

Equation (4.10) fits the data well. In the long run of 100 000 iterations ($N=822$) mentioned in Sec. III, $\mathcal{E}$ varied over more than seven decades. The comparison with Eq. (4.10) is given in Table II. [The choice $C=0.23$ was made by a least-squares fit of a straight line to $\ln\mathcal{E}(n)$.] The value of C increased to 0.63 when we ran the algorithm to compute the four lowest states. The increase can be understood in light of the fact that in this case the ground state is required to be orthogonal to the next three excited states. This constraint apparently makes each iteration more efficient, but there is a cost in

TABLE II. Comparison of measurements $\mathcal{E}$ (data), with Eq. (4.10), $\mathcal{E}$ (formula). The data is for the ground state in the (0,0) channel, using the algorithm that calculates only the ground state. Parameters are $\mathcal{E}_0=6.1$, $C=0.23$, and $N=822$. $x[y]=x\times10^y$.

n (thousands)	$\mathcal{E}$ (data)	$\mathcal{E}$ (formula)	$\mathcal{E}_d/\mathcal{E}_f$
5	9.39[−1]	1.51	0.62
15	1.05[−1]	9.17[−2]	1.1
25	6.24[−3]	5.59[−3]	1.1
35	6.87[−4]	3.41[−4]	2.0
45	2.04[−5]	2.08[−5]	0.98
55	6.31[−7]	1.26[−6]	0.50
65	4.65[−8]	7.70[−8]	0.60

computer time and storage to deal with more states. It should be mentioned that the values of C for the excited states were smaller. For example, for the first excited state, C was 0.44.

When we used the full basis of $N=12\,870$ states to compute the ground state, $\mathcal{E}$ varied over about two decades, and the fit was with $C=0.21$. Altogether, these experiments verify that the convergence of the algorithm is exponential. The number of iterations for e folding is several times N.

V. PARTIAL BASES

The number of states in the Hubbard-Heisenberg model is small enough that we can store them all in our computer. We mentioned in Sec. I that complete storage is usually impossible, and in this section we discuss modified algorithms that store only a partial basis. When we deal with a subset of the basis states, we do no harm if the states we retain dominate the eigenvector we wish to compute. For example, if we sum the probability amplitudes of the partial basis states in the normalized eigenvector and find a result that is very close to one, the partial basis is adequate for many purposes. It is an art to choose a basis that has the feature that a subset saturates the norm. One must use all available information, such as the presence of smallish parameters or topological excitations to choose the basis.

After a "good" basis has been chosen, we must decide how to select the subset that constitutes the partial basis, which consists, let us suppose, of N_0 orthonormal states. We want to make this selection on the basis of dynamics rather than preconception, so we adopt the following scheme. First we apply our iterative algorithm until $N_0+\Delta$ basis states have been included. We then continue the iteration using only these $N_0+\Delta$ basis states until the eigenvectors and eigenvalues are stable. At this point, the approximate eigenvectors are the optimal ones that can be constructed using the basis at hand. However, there is no guarantee that the $N_0+\Delta$ states in our basis are the most favorable ones. We therefore determine which Δ of the $N_0+\Delta$ basis vectors are least important, and we throw them away. Below we discuss how these unimportant basis vectors are identified. The number Δ of discarded states is at our disposal, but in any case we make room for Δ new basis vectors, which we select from the full basis. In this way, we weed our basis of unimportant states, replacing them, we hope, with states that are important. We continue this process until further weeding proves futile.

To see how to weed, suppose that we have determined the K lowest eigenstates of H on the subspace spanned by the partial basis of $N_0+\Delta$ states. Let the K eigenstates and eigenvalues be ψ^a and E_a. We now determine the effect of removing the basis state ϕ. To do this, form the $K-1$ linear combinations

$$\tilde{\psi}^m=\left[\sum_{a=1}^{K}\frac{|\langle\phi|\psi^a\rangle|^2}{(\alpha_m-E_a)^2}\right]^{-1/2}\left[\sum_{b=1}^{K}\frac{\langle\phi|\psi^b\rangle\psi^b}{\alpha_m-E_b}\right], \tag{5.1}$$

where α_m is the mth ordered solution of the equation

$$\sum_{a=1}^{K}\frac{|\langle\phi|\psi^a\rangle|^2}{\alpha-E_a}=0\ . \tag{5.2}$$

This equation is equivalent to an order $K-1$ polynomial, and so has $K-1$ solutions. Examination shows the solutions to be bracketed by the previous eigenvalues,

$$E_m\le\alpha_m\le E_{m+1}\ . \tag{5.3}$$

The vectors $\tilde{\psi}^m$ are orthonormal, as can be shown using Eq. (5.2). Therefore, these vectors span a $K-1$ dimensional subspace. Equation (5.2) also implies $\langle\phi|\tilde{\psi}^m\rangle=0$, so this subspace is the orthogonal complement of ϕ. Finally, these states diagonalize H on the $K-1$ dimensional subspace,

$$\langle\tilde{\psi}^p|H|\tilde{\psi}^q\rangle=\delta_{pq}\alpha_p\ . \tag{5.4}$$

We conclude that the exclusion of basis state ϕ raises the energy of state p from E_p to α_p. In order to decide which state to throw out, we must see which choice of ϕ minimizes the shift; this is the least important state for the computation of the eigenvalue p. This can be done by evaluating Eq. (5.2) for each basis state successively, but there is a much simpler scheme that is roughly equivalent. When $N_0\gg1$, and p is small, the perturbative solution of Eq. (5.2) is a reliable guide,

$$\alpha_p-E_p=|\langle\phi|\psi_p\rangle|^2\left[\sum_{\substack{a=1,\\ a\neq p}}^{K}\frac{|\langle\phi|\psi_a\rangle|^2}{E_a-E_p}\right]^{-1}\ . \tag{5.5}$$

We see that α_p is close to E_p when the probability amplitude $|\langle\phi|\psi^p\rangle|$ is small. We therefore choose ϕ to be the state that has the smallest probability amplitude in state p. This is the criterion one would have chosen intuitively. Note that in weeding a basis, the states ultimately retained will depend on the level p used during state selection.

In practice, we eliminate the Δ states having the smallest probability amplitudes $|\langle\phi|\psi^p\rangle|$ all at once, without recomputing the probability amplitude $|\langle\phi|\psi^p\rangle|$ after each state is removed. Then we add Δ more randomly chosen states from the full basis, as described above. We have used various criteria for ending the weeding process, and various choices of N_0 an Δ. Several general features are apparent. First, when the primary algorithm is used to find the eigenstates and eigenvalues, $\ln\mathcal{E}(n)$ shows significant fluctuation with n when the partial basis is, say, 150 out of 822 states. That is, the stochastic character of the process, which was mentioned in Sec. IV, becomes quite prominent. $\ln\mathcal{E}(n)$ is globally linear, but its rise is by sudden jumps, separated by plateaus on which $\ln\mathcal{E}(n)$ changes very slowly. The features are shown in Fig. 1. Since $\ln\mathcal{E}(n)$ is approximately linear over many iterations, it is still possible to determine an effective C governing convergence.

The values of C we obtain from the convergence with partial bases are close to those cited in Sec. IV. This means that C is only weakly dependent on N_0, which is important in extrapolating our partial basis results to obtain corrected eigenvalues. The specific values of C for

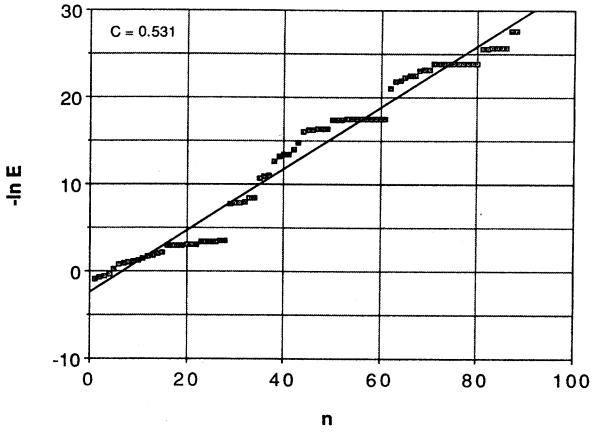

FIG. 1. Behavior of $-\ln\mathcal{E}$ for $N_0=150$ and $\Delta=30$. The straight line corresponds to $C=0.531$. The number of iterations n is in hundreds.

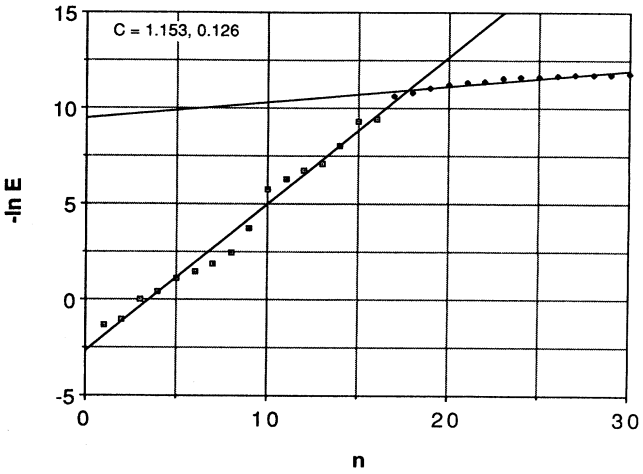

FIG. 2. Behavior of $-\ln\mathcal{E}$ for $N_0=150$ and $\Delta=45$. The two lines correspond to $C=1.153$ and $C=0.126$. n is in hundreds.

the convergence of the ground state are given for various values of N_0 in Table III. Each entry in Table III required a few hours of CPU time to compute.

The second feature worth noting is that when Δ is as large as, say, $0.3N_0$, $\ln\mathcal{E}$ sometimes shows a break in its rise. An example is shown in Fig. 2. The break does not occur when Δ is small, even though partial bases showing a break and no break differ by only a few states. In Fig. 2, when n is small, the value of C is somewhat larger than that stated in Sec. IV; for large n, C is tiny. The convergence behaves as if it is a superposition of two exponentials, with the slowly falling exponential exposed only when the rapidly falling exponential has died away. Such double exponential behavior shows that good agreement with Eq. (4.10) is not inevitable, and the question is how to explain the double exponential.

Our hypothesis is that double exponential behavior is associated with the sparse nature of H. Even with a full basis, most matrix elements of H are zero. When we remove Δ states, it may be that there are subsets of the partial basis that have very few (or even no) nonzero matrix elements of H between them. Such a disconnection is more likely when Δ is large. If there is a disconnection, one can imagine that when we iterate we are working on two nearly uncoupled problems, with convergence controlled by separate exponentials. We have tested this hypothesis by a simplified model. We take H to be block diagonal on two subspaces of N_a and N_b states: $N_a+N_b=N_0$. For simplicity we consider the primary algorithm for the ground state alone. An extension of the argument in Sec. IV shows that the error $\mathcal{E}$ is a sum of two exponentials, and if $N_a \gg N_b$, the ratio of the two exponents differs greatly from one, as seen in Fig. 2. Our hypothesis therefore explains Fig. 2, and it suggests using a sufficiently small Δ to evade the double exponentials.

TABLE III. Extrapolation of partial basis data on ground-state energies in the (0,0) channel. Equation (5.7) is used to extrapolate data from two choices of N_0 to $N=822$. The ground-state energy is actually -76.91, and at $N=822$, $C=0.63$ for the four-state algorithm.

N_0	$E_1(N_0)$	C	E_1 (extrapolated)
150	−74.401 641	0.51	
200	−74.674 341	0.64	
		0.575(av)	−75.52
200	−74.674 341	0.64	
250	−74.937 475	0.64	
		0.64(av)	−75.85
250	−74.937 475	0.64	
300	−75.214 085	0.59	
		0.615(av)	−76.29
300	−75.214 085	0.59	
350	−75.512 496	0.64	
		0.615(av)	−76.74
350	−75.512 496	0.64	
400	−75.754 590	0.45	
		0.545(av)	−76.80

Table III shows that the ground-state energies obtained with partial bases are not close to the energy obtained with the full basis, and we now show how to extrapolate the results to obtain improved energies. To do this, reconsider the argument used in Sec. IV to study convergence, but this time let $K=N=N_0$, so that, in principle, we have completely diagonalized H on the partial basis. Then E_p is the energy of the pth state that results from applying the primary algorithm until stability results. Now enlarge the basis by one state: $N_0 \to N_0+1$. The argument leading to Eq. (4.7) still applied, except that now $\mathcal{E}=E_p-\bar{E}_p$, with $\bar{E}_p$ the state energy using the complete basis,

$$\frac{\Delta\mathcal{E}}{\mathcal{E}}=-\frac{C}{N}\,. \tag{5.6}$$

However, this time the resulting differential equation and its solution are

$$\frac{d\mathcal{E}}{\mathcal{E}} = -\frac{C\,dN}{N},$$
$$\mathcal{E}(N) = B(N)^{-C}, \tag{5.7}$$
$$E_p(N) = A + B(N)^{-C}.$$

Using this extrapolation, we have used data from pairs of partial bases to compute A and B. For C we used the average of the values of C for each partial basis. The extrapolations are given in Table III. Clearly, if N_0 is as large as 400, we can extrapolate fairly well to the ground-state energy at $N=822$. Table III shows that the extrapolants are themselves approaching a limit, so within the context of calculations relying solely on partial bases it is possible to estimate how well we do.

VI. CONCLUSION

The antiferromagnetic Heisenberg model we have studied poses strong demands on any computational scheme. It has no adjustable parameters, and no systematic approximation scheme. Accordingly, we have not been able to introduce the sort of good basis contemplated in Sec. I. Instead, for our basis we have simply used plane-wave superpositions of states having definite configurations of spin-up and spin-down electrons. The dominant amplitude in the ground state is the antiferromagnetic Néel state, but the probability of this state in the ground state is only 16.5%.

Most of the data we have presented have been eigenvalues in the (0,0) momentum channel. When we used the full basis of 822 states in this channel, the accuracy of the eigenvalues we obtained was limited only by the computer. Our iterative algorithm converged exponentially, as predicted by simple theoretical considerations. The computation requirements are modest.

When we used partial bases, the accuracy was poorer, and it is useful to understand how much so. The maximum expectation of $\tilde{H}_{a,b}$ in one of our basis states is -8 (in units of t^2/U), and the minimum is -64. Of course, the maximum and minimum eigenvalues are somewhat larger and smaller than these limits, respectively. The 822 eigenvalues are presumably concentrated in the middle of this range, because the ground and first excited states are at -77 and -70 (to the nearest integer). From this perspective the approximate ground-state energy of $-74\frac{1}{2}$ obtained with a partial basis of 150 or 200 states is not so bad: Many physics problems have been understood on the basis of this sort of accuracy. With 300 to 350 states and extrapolation, the error in the ground-state energy was reduced to 0.14. If we assume, for the sake of argument, that the mean of the 822 eigenvalues is -42, then the ground state lies 35 units below the mean, and the error in our computation of this deviation is 0.4%. Extrapolation reduced the error greatly, and the success of the extrapolation scheme is an encouraging finding.

The main point we have not touched in this paper is how well the algorithm will do with a problem having a much larger basis. One does not have to look far for a problem: If we drop the condition $t/U \ll 1$, then the (0,0) channel of the 4×4 half-filled Hubbard model has a basis of 10 353 252 states. Our speculation is that as the basis becomes larger, it becomes more important to find a good basis.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant No. PHY84-15534.

[1] *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer, New York, 1986).

[2] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, England, 1986), Chap. 9.

[3] On the computer, the eigenstates ψ^a and $\tilde{\psi}^a$ are expressed as linear combinations of basis states. We omit further elaborations this entails.

[4] J. Hubbard, Proc. R. Soc. London, Ser. A **281**, 402 (1964).

[5] V. Emery, Phys. Rev. B **14**, 2989 (1976).

[6] G. F. Koster, *Space Groups and Their Representations* (Academic, New York, 1957).

[7] J. Oitmaa and D. D. Betts, Can. J. Phys. **56**, 897 (1978).

[8] This would slow down the algorithm for a critical system.