Observations on the statistical iteration of matrices

J. H. Hetherington
Physics Department, Michigan State University, East Lansing, Michigan 48824
(Received 3 January 1984)

Monte Carlo calculations which rely on the statistical iteration of some operator can sometimes lead to results in which the variance grows as the iterations proceed. Alternatively the variance can be stable but the result incorrect. The effects are demonstrated by the Monte Carlo iteration of a 2×2 matrix and analyzed in detail for this simple case. In addition an algebraic formulation of a full Monte Carlo calculation with many simultaneous configurations including a method for keeping the number of configurations constant is given. With this formulation it is shown that the naive sampling of the wave function and naive estimate of the eigenvalue based on the growth in the number of configurations will be stable but biased. It is made plausible that for a sufficiently large number (M) of simultaneous configurations the naive method leads to a result which approaches the correct one as $M\rightarrow\infty$. It is shown that the correct average eigenvalue and eigenvector is a certain weighted average defined so that it avoids the problem of the growing variance and thus becomes more accurate as the chain is extended.

I. INTRODUCTION

The Green's-function Monte Carlo method of Kalos *et al.*,[1] the Monte Carlo path-integral method of Negele,[2] and other similar schemes rely on a statistical sampling of an iterated wave function. This investigation was undertaken to understand certain puzzling results observed for large dimensional Monte Carlo problems of this type. Often when the calculation breaks down, instead of finding more and more scatter, one finds systematic deviation from the exact result. It was not known if the problem was that the most probable result was very different from the average or if it was a matter of having an unknown bias. In fact, both effects are present. The article on the Green's-function Monte Carlo method by Kalos and Ceperly[1] addresses the question but does not fully resolve the difficulty.

These methods rely on the idea that an operator's eigenvector with largest eigenvalue can be found by multiplying an arbitrary starting vector repeatedly by the operator. Since these methods proceed by sampling, the scheme is very similar to that of the Markov process because the new sample is related through a probabilistic process to the preceding sample. However, as we shall see there is an important difference. Because the well-known mathematical results on Markov processes are relevant, however, we begin by a review of those results and in particular show how our problem does not fit exactly into that scheme.

II. DEFINITION OF A STOCHASTIC MATRIX

The theorems about Markov processes are based on the use of a stochastic matrix. Since the physical problems we are investigating cannot be formulated in terms of a stochastic matrix we need to examine the distinction carefully. A stochastic matrix is a non-negative matrix all of whose columns add to one,

$$\sum_i M_{ij}=1 \tag{1}$$

(mathematics texts require $\sum_j M_{ij}=1$ because they apply the matrix to the left rather than to the right. I prefer to keep the probability distribution as a column matrix and apply the probability matrix to the right). M_{ij} is to be interpreted as follows: If a system is in state j at time t (or at iteration n), then at time $t+1$ (or iteration $n+1$) it will be in state i with probability M_{ij}. The stochastic condition [Eq. (1)] merely states that the total probability that it be in some state i is unity.

There are several results of this definition: (1) A left eigenvector of M is (1,1,1, ...) and has eigenvalue 1. (2) 1 is the maximum eigenvalue. All other eigenvalues are inside the unit circle. (3) If the matrix is nondecomposable (i.e., cannot be put in block diagonal form by row and column permutations) then the maximum eigenvalue is not degenerate. These results follow from Frobenius's theorem and the theory of Gersgoren circles.[3]

III. MARKOV CHAINS

Suppose a system has a finite number of states j. Suppose it moves from one state to another with probability M_{ij} as described above. The succession of states so generated is called a Markov chain. In particular, one must have a finite number of states, state i can only depend on the existing state j, and the matrix M_{ij} must be stochastic for it to be a Markov chain.

After two steps the probability is

$$\sum_k M_{ik}M_{kj}=(M^2)_{ij}\ ,$$

and after three steps, $(M^3)_{ij}$, etc. Therefore after many steps the probability distribution will be stable and independent of initial condition because the matrix $(M^N)_{ij}$ will be dominated by the (right) eigenvector of the highest

eigenvalue (here assumed to be not degenerate) and we have

$$p_i = M_{ij}M_{jk}\cdots M_{lm}$$

(summation convention assumed). Thus p_i is the probability of being at i if it started at m, but this is independent of m after enough steps, N. Thus for large N, $P_i = \phi_i$, where

$$\phi_i = M_{ij}\phi_j \ ,$$

i.e., ϕ_i is the right eigenvector corresponding to eigenvalue 1 and which corresponds to the left eigenvector $(1,1,1,\ldots)$.

IV. RELATION TO NON-NEGATIVE MATRICES

Some matrices can be "made stochastic" as follows: A_{ij} has left eigenvector Z_i with (maximum) eigenvalue λ and an unknown right eigenvector with eigenvalue λ. Then, if we take the modified matrix,

$$M_{ij} = \frac{1}{\lambda} Z_i M_{ij} Z_j^{-1} \tag{2}$$

(no summation), we find that it has column sums of unity (because $\sum_i Z_i A_{ij} = \lambda Z_j$), and therefore M_{ij} is stochastic. Now we can use a stochastic iteration of M_{ij} to find the unknown right eigenvector of A_{ij}.

First iterate M_{ij} to find its right eigenvector ϕ_i,

$$\phi_i = \lim_{N\to\infty} (M^N)_{ij} \ .$$

That is,

$$\sum_j \frac{1}{\lambda} Z_i A_{ij} \frac{1}{Z_j} \phi_j = \phi_i \ .$$

Then by inspection, the right eigenvector of A is

$$\psi_i = \frac{1}{Z_i}\phi_i \ . \tag{3}$$

(The introduction of Z_i into the problem is a kind of importance sampling, in fact the optimum sampling because it produces a stochastic matrix.)

V. THE DILEMMA OF THE SYMMETRIC MATRIX

Suppose A in the preceding problem is symmetric. Then the left eigenvectors are identical to the right eigenvectors and the process of obtaining a stochastic matrix requires knowledge of the very eigenvector we wish to calculate! Of course approximations may be used and they will improve the behavior by making the matrix more nearly stochastic. However the fundamental problem remains that the matrix is not stochastic unless the exact eigenvector can be introduced.

VI. USING NONSTOCHASTIC MATRICES

The limitation to stochastic matrices does not need to stop us. If the total probability of the final states is greater than the initial, one merely takes more final configurations or takes a weighted final state. In practice one cannot tolerate either a net exponential increase or decrease in the number of configurations because one would eventually have too many for any computer in one case or none in the other. Thus some kind of weighted configuration will be necessary. Because of these practical considerations we will assume that a weight will be carried to make up for the fact that the matrix is not stochastic.

Any non-negative matrix A_{ij} which has nonzero column sums can be written as a product of a stochastic matrix and a diagonal matrix

$$A_{ij} = M_{ij}w_j \ , \tag{4}$$

where $\sum_i M_{ij} = 1$ and $w_j = \sum_i A_{ij}$.

We can proceed to sample the iterate of A by deciding on the next state using probability M_{ij} and multiplying the weight carried by w_j. It is true and easily proved that on the *average* the result is a correct sampling of the eigenvector of A_{ij}. In particular if many samples are carried we have many configurations defined by a state i and a weight w. Label these configurations by the index α and a state can be indicated by (i_α, w_α). Then,

$$\psi_i = \sum_\alpha \delta_{ii_\alpha} w_\alpha \tag{5}$$

estimates the eigenvector of A_{ij}.

To illustrate this fully we will take a particular 2×2 matrix (which has weights which are reciprocals of each other for convenience)

$$A_{ij} = \begin{vmatrix} 5 & 1 \\ 1 & 3 \end{vmatrix} (6\times 4)^{-1/2}$$
$$= \begin{vmatrix} \frac{5}{6} & \frac{1}{4} \\ \frac{1}{6} & \frac{3}{4} \end{vmatrix} \times \begin{vmatrix} \frac{6}{4} & 0 \\ 0 & \frac{4}{6} \end{vmatrix}^{1/2} \ . \tag{6}$$

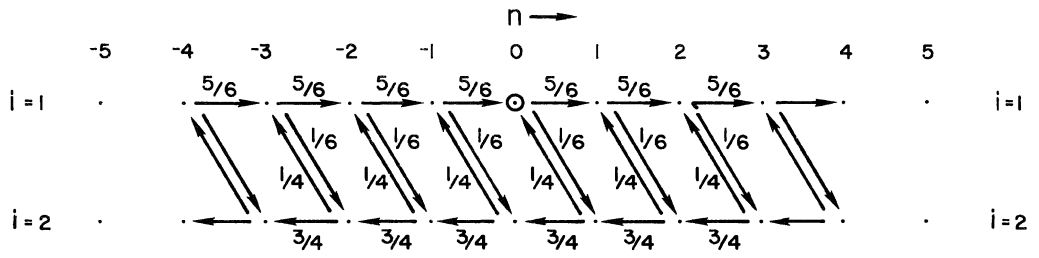

FIG. 1. Mapping of the problem of the 2×2 matrix. Note that each state, represented by a dot, makes a transition to another state according to the arrows with probability attached.

With this simplification all possible weights are of the form $(\sqrt{6/4})^n$ where n can be any integer positive or negative.

Now the final state after operating several times by the above process is described by the integers i,n where i can have the values (1,2) and n can have any integer value $\ldots,-2,-1,0,1,2,\ldots$. This space can be mapped as shown in Fig. 1. It looks a bit like a kind of distillation column. We can begin at $i=1$, $n=0$ as marked and proceed according to the probabilities shown. If this is done many times the final weighted result will indeed be the major eigenvector of the above matrix. If the final probability is p_{in} then the eigenstate of the matrix (6) will be, according to Eq. (5),

$$\phi_i = \sum_n p_{in} (\tfrac{6}{4})^{n/2} . \tag{7}$$

VII. DIFFICULTIES ARISE

Now the preceding is true for the average but consider: We might guess that after many steps, the value of n in the above problem will be normally distributed as a result of the central limit theorem. However, we must weight the results by $(\frac{6}{4})^{n/2} \equiv w_0^n$.

Therefore after N steps we expect

$$p_{in} \sim C_i \exp[-A(n-n_{0i})^2/N] , \tag{8}$$

where n_{0i}, A, and C_i are constants. Then,

$$\phi_i \propto \sum_n p_{in} e^{n \ln w_0} . \tag{9}$$

Now we find that although P_{in} (8) spreads out like $\sqrt{N}$, the maximum of the summand in (9) moves out like N and thus samples the tails of the distributions (tails which in fact probably do not obey the central limit theorem). Calculated probabilities are shown in Fig. 2. Indeed they behave as described above.

Using the assumption in Eqs. (8) and (9) we calculate the variance of ϕ and we find

$$V = (e^{N\alpha} - e^{N\alpha/2})^{1/2} / \sqrt{M} , \tag{10}$$

where

$$\alpha = \frac{(\ln w_0)^2}{A}$$

$(=1.526$ for our case). Here N is the iteration number and M the total number of samples followed. It is clear that the variance increases exponentially with the iterations and only decreases by the usual $1/\sqrt{M}$ with a large number of samples. The error will actually go through a minimum as a function of N. At first the error will decrease because the iteration eliminates components of alternate eigenvectors but eventually the error will grow because of the peculiar increase in the variance. This behavior is illustrated in Fig. 3. The estimate of the wave function and its error will behave as

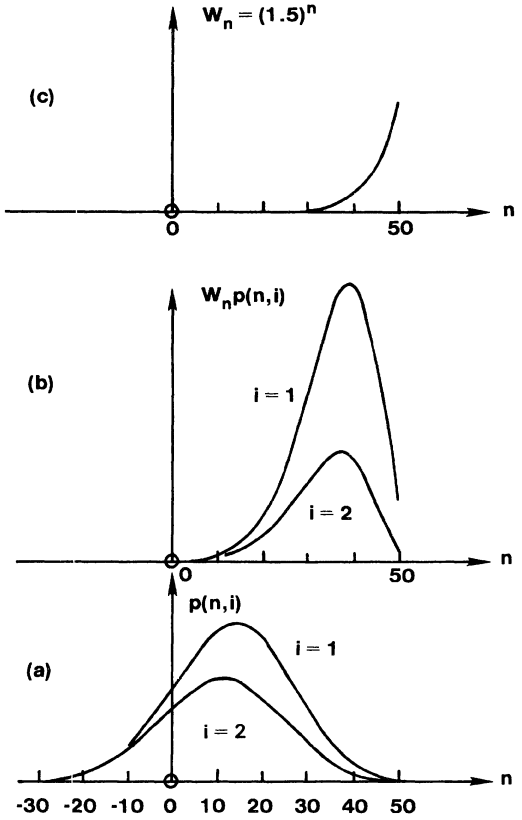

FIG. 2. (a) Probability distribution for the outcome of the procedure illustrated in Fig. 1 after 50 iterations. (b) The result of multiplying the weight shown in (c) by the probabilities in (a). Note that the peak is much farther out in (b) than in (a). In fact, the peaks in (b) are separated from the peaks of (a) by an amount proportional to N the number of steps while the width of (a) is proportional to $\sqrt{N}$. As a result, the most probable answer does not equal the average answer and the variance grows as N gets larger.

$$\phi_i = (C_0 \phi_i^0 + C_1 e^{-N\Delta} \phi_i^1 + \cdots) \pm \frac{C e^{N\alpha}}{\sqrt{M}} , \tag{11}$$

where $e^{-\Delta}$ is the ratio of the largest and next largest eigenvalues and α characterizes the exponential spread in variance. Although not mathematically related to the idea of the asymptotic series, the error is similar in that it first decreases systematically as N increases and then increases chaotically beyond some optimum N. The minimum error then depends on M as

$$M^{-[\alpha/(\alpha+\Delta)]/2} ,$$

or in other words like $M^{-\beta}$ where $0 < \beta < \frac{1}{2}$. Little is gained in proceeding further than would be necessary to achieve convergence of the iterated vector to the eigenvector in an exact scheme.

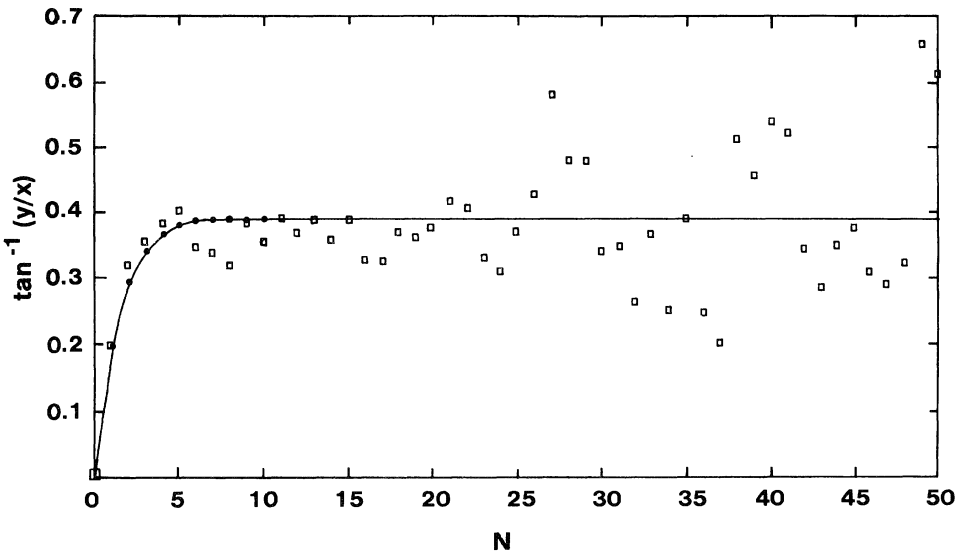

FIG. 3. An example of the iteration of the problem shown in Fig. 1. Plotted is the angle $\phi=\tan^{-1}(X_2/X_1)$, where $X_i=\sum_\alpha \delta_{ij_\alpha}(w_0)^{n_\alpha}$, where α runs over all steps in the chain and j_α, n_α are the various states in the chain. Shown are the exact result, calculated by matrix multiplication, and the statistical results when 512 samples were followed and averaged. Note that as a result of the increasing variance the calculation becomes less reliable with higher iterations.

VIII. THE WEIGHTED AVERAGE

Once one knows how many steps are necessary for the convergence of an exact iteration of the matrix one can keep a product of that many weights rather than continuing to multiply more and more weights together. In this way the variance can be controlled while the quantity in question can be averaged by taking a large series of steps. For example, the eigenvalue is then given by

$$\lambda(L)=\sum_{n=1}^{N} w^n G_n^{(L)} \Big/ \sum_{n=1}^{N} G_n^{(L)}, \tag{12a}$$

where w^n indicates the w appearing on the nth iteration (i.e., $w^n \equiv w_{i(n)}$), and where

$$G_n^{(L)}=\prod_{i=1}^{L} w^{n-i}. \tag{12b}$$

In other words, $G_n^{(L)}$ is the product of the previous L weights, where L is assumed large enough for convergence of the matrix problem. The eigenvector is given by

$$\psi_i=\sum_n p_{in} G_n^{(L)} \Big/ \sum_n G_n^{(L)},$$

where P_{in} is the probability of occupation of the state i on the nth iteration.

An example can be given based on the 2×2 matrix of Sec. VII. We consider several L values and look at the average eigenvalue and its variance as a function of L.

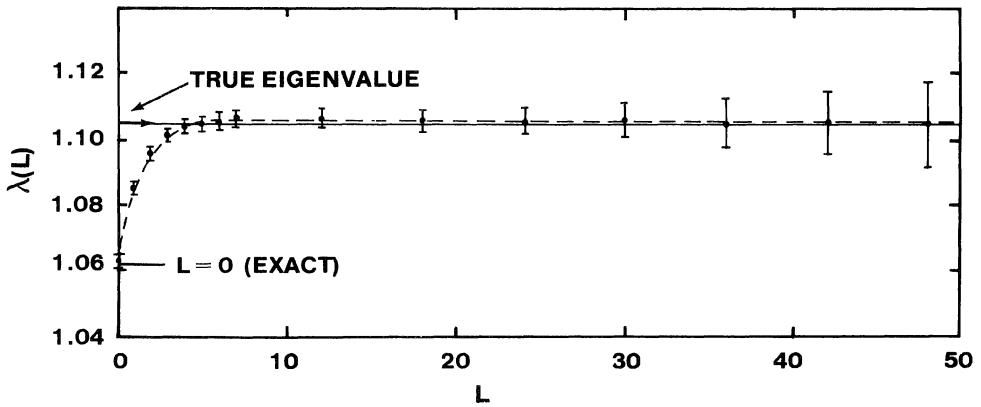

FIG. 4. A plot of the "eigenvalues" $\lambda(L)$ for the 2×2 matrix problem of Sec. VII. Note that they drift from the value for $\sum w_i p_i$ to the correct value as L increases, but that eventually the error term exponential in L dominates the result. The numerical calculation was a single configuration iterated 20 000 times. It was repeated five times to obtain error statistics. The errors are exaggerated to make them more visible but the apparent consistency of the error centers is also partly a result of correlations.

For $L=0$ (i.e., just averaging w's without weight) we obtain the result

$$\bar{w}=\sum_i w_i p_i ,$$

where p_i is the eigenvector of the stochastic part of the matrix. For the matrix of Sec. VII, $\bar{w}=1.061\,45$ while the true eigenvalue is $\lambda=1.105\,17$. Therefore as L increases we expect the average defined by Eqs. (12) to change. It should approach the correct value for large L. That is $\lambda(0)=\bar{w}=1.061\,45$, while for large L, $\lambda(L)\rightarrow 1.105\,17$. At the same time we expect the variance to increase exponentially as L becomes large. Figure 4 illustrates this behavior. There a single Markov chain is followed. The chain is generated from the matrix of Sec. VII. The eigenvalues are determined by averages of the form (12a) where $N=20\,000$. The process was repeated five times to obtain an average and standard deviation for the quantities $\lambda(L)$. Notice that the curve approaches the true eigenvalue for all L's larger than L_0, the number of steps needed to converge to the correct eigenvalue by actual matrix multiplication as shown by the dashed line in Fig. 4.

IX. CARRYING MANY CONFIGURATIONS SIMULTANEOUSLY

In practice one quickly finds that the idea of carrying weights and taking the sample according to the stochastic part of the true matrix does not work well because of the preceding difficulties. A very intuitive approach is to apply the above procedure to several configurations, and then with probability proportional to the weights, to select the same number of new configurations. Variants of this are also possible; for example, several applications of the procedure can occur before reconfiguration and indeed reconfiguration itself can proceed in different ways which have different variances. Sometimes "reconfiguration" is done by changing a trial eigenvalue which is factored from the problem.[1] We will analyze one method as an example.

Assume the matrix to be iterated, B_{ij}, is symmetric and non-negative. It can be factored into a stochastic and diagonal part

$$B_{ij}=a_{ij}w_j , \tag{13}$$

where $w_j=\sum_i B_{ij}$ and $\sum_i a_{ij}=1$. We will follow M configurations simultaneously and therefore the product space

$$\alpha=\{\alpha_1,\alpha_2,\ldots,\alpha_M\}$$

must be considered. Here α_1,α_2, etc., cover the same space as i,j, etc. Therefore if the original space has dimension N the product space has dimension N^M.

To condense the product space into the original space one asks how many of the numbers $\alpha_1\cdots\alpha_M$ are equal to i and this gives the estimate of the ith component of the vector in the reduced space.

This process is represented by the nonsquare matrix

$$n_{i\alpha}\equiv\delta_{i\alpha_1}+\delta_{i\alpha_2}+\cdots+\delta_{i\alpha_M} , \tag{14}$$

which might be termed the "condensation matrix."

The process of reconfiguring is represented by the operator

$$P_{\alpha\beta}=\prod_{i=1}^{M}\left[\frac{w_{\alpha_i}}{\sum_i w_{\beta_i}}(\delta_{\alpha_i\beta_1}+\delta_{\alpha_i\beta_2}+\cdots+\delta_{\alpha_i\beta_M})\right] , \tag{15}$$

which will be called the "reconfiguration matrix." Operationally this stochastic reconfiguration matrix, $P_{\alpha\beta}$, is defined by dividing the interval (0,1] into subintervals for each configuration in proportion to their weight; then configurations to be used in the next iteration are selected by the subintervals in which M random numbers fall.

With these definitions the following facts are easily shown:

$$P_{\alpha\beta} \text{ is stochastic} , \tag{16a}$$

$$n_{i\alpha}P_{\alpha\beta}=w_i n_{i\beta}(W_\beta)^{-1} , \tag{16b}$$

where

$$W_\beta\equiv\frac{1}{M}(w_{\beta_1}+w_{\beta_2}+\cdots+w_{\beta_M}) . \tag{17}$$

We define a stochastic matrix in the product space as

$$A_{\alpha\beta}\equiv a_{\alpha_1\beta_1}a_{\alpha_2\beta_2}\cdots a_{\alpha_M\beta_M} . \tag{18}$$

Then in addition we have

$$n_{i\alpha}A_{\alpha\beta}=a_{ij}n_{j\beta} . \tag{19}$$

[Note that A and a are square matrices while n is not. Equations of the form (19) are treated in Gantmacher.[3] In particular, nonzero n implies equality of some eigenvalues of A and a.]

Finally we come to the following result: If

$$A_{\alpha\beta}P_{\beta\gamma}W_\gamma Z_\gamma=\lambda Z_\alpha \tag{20}$$

using the summation convention, then defining

$$y_i\equiv n_{i\alpha}Z_\alpha , \tag{21}$$

we find that

$$a_{ij}w_j y_j=\lambda y_i . \tag{22}$$

This is easily proved by applying $n_{i\alpha}$ from the left to Eq. (20) using Eqs. (18) and (19) together with the definition Eq. (21).

In other words the result means that the solution of the larger problem (which is written as a product of a stochastic matrix $A_{\alpha\beta}P_{\beta\gamma}$ and a weight W_γ) leads via the condensation matrix $n_{i\alpha}$ to the solution of the desired problem, $a_{ij}w_j$. Furthermore, it seems intuitive that this larger problem suffers less from the difficulties of divergence of the variance.

X. THE LIMIT OF LARGE M

If the weight W_γ in Eq. (20) is omitted then one iterates the stochastic combination

$$A_{\alpha\beta}P_{\beta\gamma}$$

and the limiting probability Q_α, which solves the resulting equation

$$A_{\alpha\beta}P_{\beta\gamma}Q_\gamma = Q_\alpha \ , \tag{23}$$

which will eventually develop. Note that the eigenvalue is unity because the matrix is stochastic.

Now since Eq. (23) differs from the exact equation (20) by the omission of the factor W_γ we expect the maximum eigenvector of (23) to be smaller in the region of large W_γ than the eigenvector of Eq. (20). Therefore we can expect that the average weight W_γ for a solution will be too small. As the number M of configurations is increased the average $\sum_\gamma W_\gamma Z_\gamma$ should gradually increase from below and reach the limit λ.

Proceeding as in the proof above and multiplying Eq. (23) by $n_{i\alpha}$ on the left, one obtains

$$a_{ij}w_j n_{j\alpha} W_\alpha^{-1} Q_\alpha = n_{i\alpha} Q_\alpha \ . \tag{24}$$

We define $\phi_i \equiv n_{i\alpha}Q_\alpha$ and $\psi_i \equiv n_{i\alpha}W_\alpha^{-1}Q_\alpha$. Then if we write

$$\begin{aligned}\psi_j &= n_{j\alpha}W_\alpha^{-1}Q_\alpha \\ &= (\delta_{j\alpha_1}+\delta_{j\alpha_2}+\cdots) \\ &\times \left[\frac{M}{W_{\alpha_1}+W_{\alpha_2}+\cdots+W_{\alpha_M}}\right] Q_{\alpha_1\alpha_2\cdots\alpha_M} \ ,\end{aligned}$$

on the average this expression is symmetric under interchange of the α_i and we have

$$\psi_j \cong M\delta_{j\alpha_1}W^{-1}_{\alpha_1\cdots\alpha_M}\left[1-\frac{w_{\alpha_1}}{M}W^{-1}_{\alpha_2\cdots\alpha_M}\right] \times Q_{\alpha_1\cdots\alpha_M} \ . \tag{25}$$

If the α's in Q are uncorrelated to order $1/M$, the ψ are related to the ϕ by

$$\psi_j = \left[1+w_j O\left[\frac{1}{M}\right]\right]\phi_j \ . \tag{26}$$

Then one can say that if Q solves Eq. (23) then ϕ_j solves

$$a_{ij}w_j[1+w_j O(M^{-1})]\phi_j = \lambda\phi_j \ . \tag{27}$$

Since (27) is a matrix problem which differs only by a part proportional to $1/M$ from the exact problem, we have in (27) a stochastic matrix the iteration of which leads to a result differing from the desired result by terms proportional to $1/M$. It must be noted that this result depends on the assumption about the degree of correlation of the configurations in Q and therefore cannot be considered proven. Some numerical experiments have indicated the same result, however.

XI. DISCUSSION

The discussion will be less cumbersome if several quantities are defined here: M is the number of simultaneous configurations which are treated. It is assumed that these configurations communicate in the sense that reconfiguration takes place after every iteration. $L_0 = 1/\ln(\lambda_0/\lambda_1)$, i.e., L_0 is the number of iterations needed to eliminate (by factor of e) the effects of the next significant eigenvector. L is the number of factors of W included in the weighting of averages. N is the total number of iterations. S is the number of terms in the averages.

In the Monte Carlo iteration of an operator which does not preserve the number of configurations (is not stochastic) where several (M) configurations are kept, we make the following statements.

(a) We are always left with a problem which is not stochastic in detail. Therefore the best measure of, say, the eigenvalue is to use the weighted average derived in Sec. VIII,

$$\lambda(L) = \sum_{s=0}^{S} W^{N-s}G_{N-s} \Big/ \sum_{s=0}^{S} G_{N-s} \ , \tag{28a}$$

where

$$G_N = \prod_{i=1}^{L} W^{N-i} \ . \tag{28b}$$

Here W^n is the average weight as in Eq. (17) occurring on the nth iteration. The average given by Kalos and Ceperly (Eq. 4.68 of Ref. 1) is not the same as this one but in fact is somewhere between this value and the unweighted value.

Figure 5 illustrates application of Eqs. (28) to a ten-dimensional coupled harmonic oscillator problem. The matrix iterated is $e^{-\beta T}e^{-\beta V}$. The factor $e^{-\beta T}$ is stochastic while the factor $e^{-\beta V}$ is the diagonal weight matrix. β

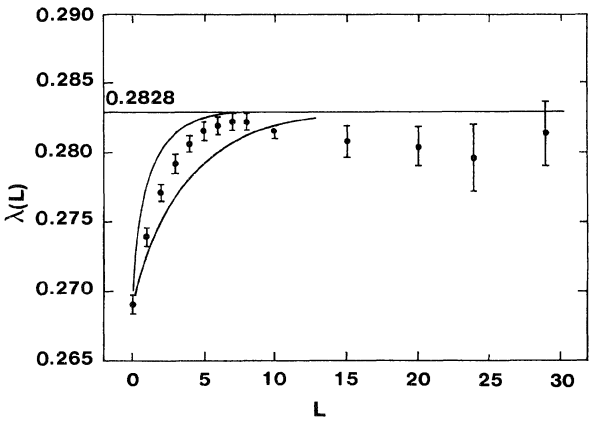

FIG. 5. Ten one-dimensional harmonic oscillators on a ring with nearest-neighbor coupling. The calculation was repeated four times and the average and expected error of the average of the four calculations are shown. Each calculation was carried out with 30 configurations ($M=30$) and 2000 iterations. Shown is the eigenvalue $\lambda(L)$ as a function of L. The exact value is within 1 to 2 standard deviations of the expected error after $L \sim 10$. Note that the error bars expand and the value $\lambda(L)$ wanders increasingly with larger L. Convergence is more rapid than would be expected from the lowest even excited state, but slower than from the even excited state based on the highest frequency mode as shown by the two exponential curves.

is taken small enough that little effect is expected from the noncommutivity of T and V.

(b) The wave function should be similarly weighted,

$$\Psi_i = \sum_{s=0}^{S} \phi_i^{N-s} G_{N-s} .$$

(c) If $L=N$ then averages will diverge with N as the iteration proceeds (see Sec. VII).

(d) As M is increased the stochastic part of the procedure approximates more closely the actual problem (see Sec. X).

(e) There are contributions to the error in this estimate of the following form:

$$\frac{C_1}{M} e^{-L/L_0} \pm C_2 \left[\frac{L}{SM}\right]^{1/2} \pm C_3 \left[\frac{L}{S}\right]^{1/2} \times \left[\exp\left[\frac{2L}{M}B^2\right] - \exp\left[\frac{L}{M}B^2\right]\right]^{1/2} . \quad (29)$$

Here, C_1, C_2, C_3, and B are constants. The first term is due to the incomplete elimination of effects of other eigenvectors of the exact problem in the soution of the stochastic part, $A_{\alpha\beta}P_{\beta\gamma}$, of Eq. (20). The first term also has the factor $1/M$ as suggested in Sec. IX. The second term is a simple statistical term which is inversely proportional to the square root of the total number of contributing terms in the average. The influence of the correlation between successive terms in the average has been represented by the $\sqrt{L}$ factor. The third term is due to the increasing variance as a result of keeping a long product of varying factors. It differs from Eq. (10) by the factor $(L/S)^{-1/2}$ and by the assumption that the statistical fluctuation in the logarithm of the W's is proportional to $1/\sqrt{M}$. The factor $(L/S)^{-1/2}$ results from the averaging over many correlated terms. Note that the first term in (29) is not statistical in nature and therefore cannot be averaged out but can only be removed by increasing M or L, or both. The third term is, for sufficiently large L/M, exponential in character and will eventually dominate if L/M is large enough. Therefore for given M and S we expect that increasing L will at first decrease the error because of a decrease in the systematic error of the first term but that the error will eventually increase chaotically because of the third term.

One expects from Eq. (29) that the number of simultaneous configurations M, needed for a successful statistical performance, is proportional to a constant times L_0 and therefore is manageable for many problems.

XII. CONCLUSIONS

Iteration of a matrix operator applied to a sampling of a vector is a Markov chain provided the matrix is stochastic. The matrix operators which lead to the ground state of a Hamiltonian are not stochastic. For this reason the number of configurations in the sample will either grow or decay as a result of iteration. Practical considerations then require that some method be adopted to control the number of configurations. Algebraic formulation of one method of control makes it possible to show that the overall problem in the product space of all configurations is also not stochastic. For this reason the growth factor in the number of configurations is not the eigenvalue of the operator.

Consideration of the general problem of iteration of nonstochastic matrices yields the correct form for the eigenvalue. It involves a product of such growth factors and therefore a variance which would be uncontrolled if too many factors were used. The error in the eigenvalue is shown to go through a minimum as a function of the number L of these factors somewhat in analogy with an asymptotic series. The solution of the stochastic part of the problem approaches the correct solution of the problem as the number of configurations increases. As a result the systematic part of the error in the eigenvalue or other quantities will be reduced when the number of configurations is increased.

The advantages of the use of the weighted average are that the error can be caused to be nonbiased and that the size of a problem solvable in a given computer memory can be increased because one does not need to depend solely on the $1/M$ reduction of the bias caused by the nonstochastic nature of the problem. Furthermore, analysis of the quantities calculated as functions of the number of factors in the weights as in Fig. 5 gives important information about both the systematic and random errors.

ACKNOWLEDGMENT

I wish to thank Dan Stump for many helpful conversations.

[1]D. M. Ceperly and M. H. Kalos, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer, New York, 1979).

[2]J. W. Negele, in *Recent Progress in Many Body Theory*, Vol. 198 of *Lecture Notes in Physics*, edited by H. Kümmel and M. L. Ristig (Springer, New York, 1983); B. Serot, S. E. Koonin, and J. W. Negele, Phys. Rev. C **28**, 1679 (1983).

[3]F. R. Gantmacher, *Matrix Theory* (Chelsea, New York, 1964).