Noise without Noise: A New Monte Carlo Method

A. D. Kennedy
Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

and

J. Kuti[a]
Department of Physics, University of California, La Jolla, California 92093
(Received 18 March 1985)

A new Monte Carlo method is introduced which generates configurations according to any desired probability distribution. Unlike previous techniques, which require the relative probability of any two configurations to be computed exactly, this method allows the prescence of large but unbiased noise in this computation. The method has important applications in including the effects of dynamical fermions in Monte Carlo calculations, amongst other problems.

PACS numbers: 02.70.+d, 02.50.+s, 11.15.Ha

In recent years Monte Carlo methods have been used with considerable success in the study of the properties of quantum field theories outside the domain of applicability of weak- or strong-coupling perturbative methods. Important results have been obtained recently in condensed-matter physics by the application of similar techniques.

One of the most outstanding problems in these studies has been the inclusion of the dynamical effects of fermion fields. A standard procedure is to derive an effective bosonic action by integrating out the fermions in the field-theoretic or condensed-matter model under investigation. The resultant effective bosonic action is highly nonlocal, since it contains the determinant of the fermion matrix associated with the model as a factor.

The major difficulty with this subject is the vast amount of computation required to perform any reliable numerical calculations: This is because the computation of the relative measure $\mu[A_{\text{new}}]/\mu[A_{\text{old}}]$ which is required to effect a step of the Markov process which generates bosonic configurations with the correct distribution becomes prohibitively costly when the action involved becomes highly nonlocal.

The symptom of this problem is that the stochastic part of the Monte Carlo computation (e.g., the generation of random numbers) becomes a negligible part of the computational cost as compared to the explicit evaluation of the relative measures. The purpose of this paper is to show that this time-consuming exact computation is unnecessary, and that a noisy (but unbiased) estimate of the relative measures gives completely equivalent results.

Our results show that by use of an acceptance algorithm different from the usual Metropolis one the effect of noise does not serve to disorder the system being studied, as one might at first expect. Instead, a noisy system provides an identical Markov process to one without noise, and hence the configurations sampled are distributed according to the correct measure, and the measured values (and variances) of all physical quantities will be unchanged.

The method may also be useful for "improved" bosonic actions in quantum field theories, because such actions are significantly more nonlocal than naive actions.

In the following we shall introduce our new algorithm for generating a valid Markov process in the presence of noise. For simplicity of exposition we shall discuss explicitly the case of dynamical fermions in quantum chromodynamics, but the general applicability of the method is obvious.

Markov process.—The basic requirement of a Monte Carlo updating algorithm in lattice QCD is that it generates boson configurations according to the probability distribution

$$P(U)(dU) = Z^{-1}\det M(U)\, e^{-\beta S_B(U)}(dU), \qquad (1)$$

where Z is chosen such that $\int (dU)P(U) = 1$, (dU) is the SU(3) Haar measure, S_B is the bosonic action, and $\det M$ is the fermion determinant. It is well known that if we construct a Markov process which is ergodic and which has P as a fixed point, then the distribution of configurations it generates will converge to P. A sufficient (although not necessary) condition for P to be a fixed point is that detailed balance is satisfied:

$$P(V \leftarrow U)P(U) = P(U \leftarrow V)P(V), \qquad (2)$$

where $P(V \leftarrow U)$ is the probability of generating configuration V from configuration U.

We may write this transition probability as the product of an *a priori* probability of selecting V, $P_s(V)$, and a probability $P_a(V \leftarrow U)$ of accepting the update. We can choose $P_s(V)$ to be the distribution for link V ignoring the fermion determinant, using a heat-bath algorithm:

$$P_s(V)(dV) \propto e^{-\beta S_B(V)}(dV), \qquad (3)$$

and the effects of the dynamical fermions are incorporated by the acceptance probability $P_a(V \leftarrow U)$.

Introduction of noise.—We now turn to the question of how to implement the acceptance step. The basic premise of our method is that all we know about the fermion determinant ratio is an unbiased estimator

$$\Delta_{V\leftarrow U}(x)=\frac{\det M(V)}{\det M(U)}x, \tag{4}$$

with x a random variable whose distribution $\rho(x)$ is arbitrary except that it is normalized and of mean unity:

$$\langle\Delta_{V\leftarrow U}(x)\rangle_x=\int dx\,\rho(x)x\frac{\det M(V)}{\det M(U)}=\frac{\det M(V)}{\det M(U)}. \tag{5}$$

If we require that the acceptance probability also be unbiased we must make P_a linear in this estimator,

$$P_a(V\leftarrow U)=\begin{cases}\lambda^+ +\lambda^-\langle\Delta_{V\leftarrow U}(x)\rangle_x & \text{if} \quad U>V,\\ \lambda^- +\lambda^+\langle\Delta_{V\leftarrow U}(x)\rangle_x & \text{if} \quad U\leq V,\end{cases} \tag{6}$$

with $\lambda^{\pm}$ being constants, that is they do not in any way depend upon the random variable x. Two points about Eq. (6) deserve further explanation. First, the condition $U>V$ refers to any pairwise ordering of the configurations U and V; a suitable ordering could be that $U>V$ if U is the state with larger bosonic measure, i.e., $e^{-\beta S_B(U)}>e^{-\beta S_B(V)}$, but this is totally *ad hoc*. Second, we should explain why the average $\langle\ldots\rangle_x$ appears in the equation: In any particular step in a practical calculation the random variable x will take on some particular value, and the probability of accepting the new configuration V will depend on this value. However, the probability which occurs in the detailed balance equation is not this one, but its convolution with the probability that x takes such a value, as we have used in Eq. (6).

It is now easy to verify that detailed balance is indeed satisfied; from Eqs. (2)–(6) we find that (for the case $U>V$)

$$\begin{aligned}&P_a(V\leftarrow U)P_s(V)P(U)-P_a(U\leftarrow V)P_s(U)P(V)\\ &\propto(\lambda^-\langle\Delta_{V\leftarrow U}(x)\rangle_x+\lambda^+)e^{-\beta S_B(V)}\det M(U)e^{-\beta S_B(U)}\\ &\quad-(\lambda^+\langle\Delta_{U\leftarrow V}(x)\rangle_x+\lambda^-)e^{-\beta S_B(U)}\det M(V)e^{-\beta S_B(V)}\\ &\propto\lambda^-\det M(V)+\lambda^+\det M(U)-\lambda^+\det M(U)-\lambda^-\det M(V)=0.\end{aligned} \tag{7}$$

It is worthwhile to compare this case with the Metropolis algorithm, for which

$$P_a(V\leftarrow U)=\theta\left(\frac{\det M(V)}{\det M(U)}-1\right)+\theta\left(1-\frac{\det M(V)}{\det M(U)}\right)\frac{\det M(V)}{\det M(U)}, \tag{8}$$

which cannot be constructed in an unbiased way from $\Delta_{V\leftarrow U}(x)$.

Choice of parameters.—It only remains to choose the parameter values $\lambda^{\pm}$ in Eq. (6). These must be chosen such that P_a always lies in the interval [0,1], as otherwise P_a would not be a probability. This is impossible in general for arbitrary $\rho(x)$, but if we assume that $0<\Delta_{U\leftarrow V}(x)<2$ then $\lambda^-=\frac{1}{2}$, $\lambda^+=0$ seems to be a favored choice, because the determinant ratio estimator Δ then has to be computed for only about half of the steps. This compensates for the fact that the intrinsic acceptance rate is about half of that of the Metropolis algorithm. There are, of course, many other choices for $\lambda^{\pm}$, e.g., ones which permit Δ to fluctuate over a wider range at the cost of a lower acceptance rate (the acceptance rate falls linearly with the allowed range of Δ values, as compared with the $1/\sqrt{n}$ variance for the stochastic computation of Δ).

In practice, of course, the actual range of the distribution is not going to be bounded however small we make the variance, but the number of times that the condition $0<\Delta<2$ is violated can be adjusted to be very small, and furthermore it is trivial to count the number of times such violations occur. It should be relatively easy to keep the systematic errors induced by these bound violations under control.

The simplest illustration of our Monte Carlo algorithm is to a system with a finite number of states. For the sake of definiteness we consider a system with five states, which we assign "energies" $E_i\equiv i/10$ $(i=0,\ldots,4)$, and by fiat we choose the probability of the states to be $P_i\equiv e^{-E_i}/\sum_{j=0}^{4}e^{-E_j}$. We shall generate a sequence of states with this distribution using a Markov process of the type described above.

For a given step we choose the candidate new state uniformly at random from the five possibilities, that is, in the language used before, $P_s(i)=\frac{1}{5}$ $(\forall\, i)$. The quantity which we called Δ before will in this case be taken to be $\Delta_{i\leftarrow k}(x)\equiv P_i/P_k+x$, where x is a random variable with mean zero. One may notice that we use additive noise here, whereas we used multiplicative noise in the theoretical argument: The conclusion, however, remains the same.

If we choose the arbitrary ordering of the states required in Eq. (6) to be that state i is less than state k if $i<k$, then we have

$$P_a(i\leftarrow k)=\begin{cases}\frac{1}{2}\langle\Delta_{i\leftarrow k}(x)\rangle_x=P_i/2P_k & \text{if} \quad k>i,\\ \frac{1}{2} & \text{if} \quad k\leq i.\end{cases} \tag{9}$$

This procedure leads to the following Markov transition matrix:

$$\begin{pmatrix} \frac{3}{5} & \frac{P_0}{10P_1} & \frac{P_0}{10P_2} & \frac{P_0}{10P_3} & \frac{P_0}{10P_4} \\ \frac{1}{10} & \frac{7P_1-P_0}{10P_1} & \frac{P_1}{10P_2} & \frac{P_1}{10P_3} & \frac{P_1}{10P_4} \\ \frac{1}{10} & \frac{1}{10} & \frac{8P_2-P_1-P_0}{10P_2} & \frac{P_2}{10P_3} & \frac{P_2}{10P_4} \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{9P_3-P_2-P_1-P_0}{10P_3} & \frac{P_3}{10P_4} \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{10P_4-P_3-P_2-P_1-P_0}{10P_4} \end{pmatrix} \tag{10}$$

provided only that the noise lies between the allowed bounds, in this case $|x| < 2 - P_0/P_4 \approx 0.508$. Iteration of the Markov process governed by Eq. (10) will produce the five states with relative frequencies approaching the desired probabilities.

In Fig. 1 we show the average energy measured using this Markov process. We choose the distribution for the noise to be $x = \pm\sigma$, where the two signs are chosen equiprobably, and the parameter σ determines the magnitude of the noise. This distribution gives maximal noise within the specified bounds. It is easily seen that $\langle E \rangle$ is correctly determined when $\sigma < 0.508$, and for larger noise the system approaches the limit in which the noise dominates the transition probabilities and the states are visited with equal frequency. Also shown in the figure are the corresponding measurements made by use of the Metropolis algorithm; these deviate from the correct result even for small values of σ.

It is most important to realize that it is not only the mean and its variance which are unchanged by the addition of noise with use of our method; *all* averages measured over the Markov chain will be correct. This is clearly illustrated in Fig. 2, in which we show the measured relative frequencies of the five states from the same simulation as gave Fig. 1.

We have established that a noisy estimate for Δ leads to just as good a sequence of bosonic configurations as an exact evaluation, provided only that the estimate is unbiased and that the value of Δ lies in a bounded range (say, $0 < \Delta < 2$). We shall now give several examples of how such noisy estimates may be computed efficiently in terms of the inverse of the fermion matrix.

It has been observed[1] that a local change $U \rightarrow U + \delta U$ in the bosonic field leads to a fermion determinant ratio

$$\frac{\det[M(U+\delta U)]}{\det[M(U)]} = \det[1 + M^{-1}(U)\delta M(U)], \tag{11}$$

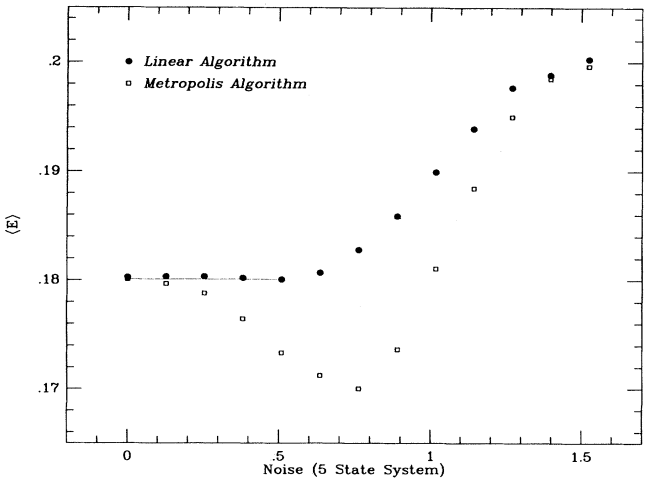

FIG. 1. The average energy of the five-state system as a function of noise in the updating process. The exact value is indicated by the line. The method (linear algorithm) shows bias only for violations of the bounds.

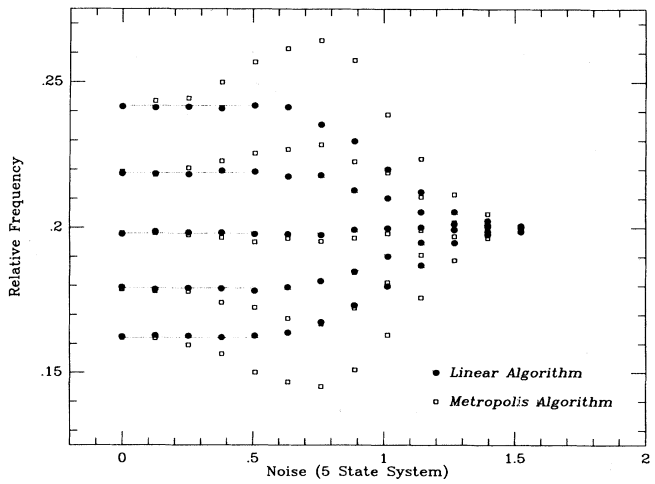

FIG. 2. The relative frequencies of the five states as a function of the noise. The lines correspond to the exact results.

where the nontrivial change δM in the fermion matrix is restricted to the neighborhood of the updated lattice site or link for models which have local boson-fermion couplings. Our Monte Carlo algorithm is immediately applicable to fermion methods which attempt to calculate the matrix elements of the inverse fermion matrix in a stochastic fashion.

von Neumann–Ulam random walks.—This technique which was described earlier[2] immediately satisfies the condition of generating an unbiased estimator. One calculates the inverse matrix element M_{ij}^{-1} from a properly chosen scoring procedure for random walks from site i to site j on the lattice.

According to the law of large numbers the sum of the scores for the random walks has a Gaussian probability distribution around the mean M_{ij}^{-1} with a width given by $\sigma/\sqrt{N}$, where σ^2 is the variance of the random walks and N is the number of walks in the procedure. By taking sufficiently many walks we can reduce the variance of the distribution $\rho(x)$ as much as we desire.

We tested our new Monte Carlo algorithm on a simple boson-fermion model, calculating the inverse matrix elements of the fermion matrix from von Neumann–Ulam random walks. The fermion matrix M was defined as

$$M_{ij} = -\Delta_{ij} + (m + gU_i^2)\delta_{ij}, \tag{12}$$

where U_i is a scalar boson field coupled to a spinless fermion field.[1] Δ_{ij} is the lattice Laplacian operator, m is the fermion mass, and g designates the boson-fermion coupling constant (this model calculation was discussed earlier[2] in the framework of the Metropolis procedure, whose results exhibit a systematic bias for noisy random walks).

The fermion-boson interaction dynamically generates a mass term, and the renormalized fermion mass is given by $m_R^2 = m^2 + \frac{1}{2}g$. We computed the renormalized fermion propagator to test our method. For the updating step we decreased the number of random walks and monitored various physical quantities, like the fermion propagator: The picture that emerged is similar to Fig. 1. With increasing noise from the random walks the variance of the measured physical quantities remained identical to that of the exact updating procedure. As the noise further increased the Metropolis method quickly became biased while the new algorithm remained unbiased much longer until the bound violations became significant.

Pruning.—This method which we tested on the Schwinger model[3] introduces some small bias into the calculation of the fermion determinant ratio because it is a stochastic iteration of the Jacobi method for inverting large sparse matrices, and it can therefore only be carried out for some fixed number of Jacobi iteration steps. This bias can be made negligibly small, provided that the fermions are sufficiently heavy. On the other hand, this method deals with sign cancellations more effectively than the von Neumann–Ulam method, and thus is expected to give a smaller variance.

Pseudofermions.—This technique[4] does not directly lend itself to our approach. As it is normally used, the fermion Green's functions $\langle\bar{\psi}(x)\psi(y)\rangle$ are computed once per sweep, and then used to update all the links in the lattice successively. Even though the changes in the link variables are constrained to be small (thereby increasing the correlations between sweeps as a side effect), this introduces a bias into the values of Δ used. An alternative would be to perform a brief pseudofermion Monte Carlo calculation for each bosonic link update, as there is no need to reduce the variance in the measurement of $\langle\bar{\psi}(x)\psi(y)\rangle$ by a large amount. Unfortunately, although one only needs to sample few equilibrium pseudofermion configurations, it is still necessary to perform enough pseudofermion sweeps for that Markov process to converge to the correct distribution, and it is unclear how long this convergence time is. Variants of the pseudofermion technique which utilize the new algorithm are under study.

In conclusion, we have developed a new Monte Carlo algorithm which has great potential for fermionic applications in lattice gauge theories and condensed-matter physics. We have presented the theoretical justification and illustrated the method on some simple examples. However promising the algorithm is, its ultimate success will depend on the efficient calculation of an unbiased stochastic estimate of the fermion determinant ratio.

We should also stress that the method has a general applicability beyond the obvious fermion problems which we have discussed here.

We would like to thank Brian Pendelton for his help during the early stages of this work. We thank Jorge Hirsch for discussions and advice. This research was supported in part by the National Science Foundation under Grant No. PHY82-17853, supplemented by funds from the National Aeronautics and Space Administration, and by the U. S. Department of Energy under Grant No. DEAT0381ER40029.

---

[a]On leave from Central Research Institute for Physics, Budapest, Hungary.

[1]D. J. Scalapino and R. L. Sugar, Phys. Rev. Lett. **46**, 519 (1981).

[2]J. Kuti, Phys. Rev. Lett. **49**, 183 (1982).

[3]A. D. Kennedy, J. Kuti, and B. J. Pendleton, to be published.

[4]F. Fucito, E. Marinari, G. Parisi, and C. Rebbi, Nucl. Phys. B **46**[FS2], 369 (1981).