Journal of Statistical Physics, Vol. 63, Nos. 5/6, 1991

Exact Monte Carlo for Few-Fermion Systems

M. H. Kalos[1]

We have reconsidered the fundamental difficulties of fermion Monte Carlo as applied to few-body systems. We conclude that necessary ingredients of successful algorithms include the following: There must be equal populations of random walkers that carry positive and negative weights. The positions of positive walkers should be selected from a distribution that uses Green's functions to couple all walkers. The positions of negative walkers should be generated from those of positive walkers by means of odd permutations. The correct importance functions that take into account the global interactions of the populations are different for positive and negative walkers. Use of such importance functions breaks the symmetry that otherwise would exist between configurations (of the entire population) and configurations derived by interchanging positive and negative walkers. Based upon these observations, we have constructed a stable and accurate algorithm that solves a fully-polarized, three-dimensional, three-body model problem.

KEY WORDS: Quantum Monte Carlo; fermions.

1. INTRODUCTION

By now, the "sign problem" that arises in the Monte Carlo integration of the Schrödinger equation for fermion systems is well known to those who know it well.[1] Briefly, the requirement of antisymmetry of particles with like spins requires that the wavefunction be negative as often as positive. It is straightforward to represent the wavefunction by a population of random walkers that carry signs, but the usual algorithm—each walker moves independent of the others—causes each population to converge to the same limit, a symmetrical one. Having a population of positive walkers and a

[1] Laboratory of Atomic and Solid State Physics and Center for Theory and Simulation in Science and Engineering, Cornell University, Ithaca, New York 14853.

822/63/5-6-30

separate population of negative walkers with the same asymptotic distribution implies zero overlap with any test function, symmetric or antisymmetric. This difficulty is not merely technical. Clearly, the requirement that wavefunctions be antisymmetric is a global property, one that cannot be satisfied by an ensemble of independent walkers. There is another way to put the difficulty: The simple algorithms have the property that positive and negative walkers can be exchanged with impunity. This (higher) symmetry again leads us to the conclusion that the asymptotic overlap with any test function is zero.

This paper provides a summary of the technical issues, followed by a proposal for a solution. There are two key ingredients for this new approach. The first is the use of an ensemble of interacting walkers. By itself, however, such an ensemble and the algorithms that sample new states of the ensemble do not break the symmetry between positive and negative walkers. The second key idea follows from a critical reevaluation of the meaning of the importance function in such ensembles. It turns out that the importance function is no longer the wavefunction itself (except in the limit of infinite population). Instead, positive and negative walkers have different (but related) importance functions. It is this difference that breaks the symmetry and permits, in principle, an exact, asymptotically stable method.

We sketch a model problem based on Gaussian kernels. A numerical solution of a fully polarized, three-dimensional, three-body system in this model is described.

2. TECHNICAL BACKGROUND

Consider the Schrödinger equation for a system of N particles $\mathbf{r}_1$, $\mathbf{r}_2,\ldots,\mathbf{r}_N$ which we denote as R. They interact with a potential $V(R)$ which is usually (but not necessarily) a sum of one-body and two-body terms. In appropriate units, the Schrödinger equation is

$$\left[-\sum_i \nabla^2 + V(R)\right]\psi_k(R) = [-\nabla_R^2 + V(R)]\,\psi_k(R) = E_k\psi_k(R)$$

Shift the energy scale by V_0 such that

$$V_0 + E_0 \geqslant 0$$

Then the Green's function for $[-\nabla^2 + V(R) + V_0]$, namely the solution of

$$[-\nabla_R^2 + V(R) + V_0]\,G(R, R_0) = \delta(R - R_0)$$

is nonnegative and integrable. G may be used to rewrite the Schrödinger equation as an integral equation

$$\psi_k(R) = E_k \int G(R, R')\, \psi_k(R')\, dR'$$

The iteration of this equation may be accomplished by sampling $G(R_1, R_0)$ as a stochastic kernel for R_1 conditional on R_0, then $G(R_2, R_1)$ for R_2 conditional on R_1, and so on. The limiting distribution for R_n is $\psi_0(R_n)$. When $V(R)$ is symmetric in the coordinates of all particles, ψ_0 is also symmetric.

Sampling $G(R_1, R')$ can be carried out in many ways. There is an exact sampling method[2] that relies on first passage across surfaces within configuration space plus upper bounds to $V(R)$ within the subdomains. A variety of other methods[3] use numerical approximations to Brownian motion to approximate $G(R, R')$. From the point of view of finding antisymmetric solutions, there is not much to choose among them.

2.1. Importance Sampling

A change that has substantial value in reducing the fluctuations in the random walks arises from the use of an importance sampling transformation. Consider the integral equation

$$\psi_0(R) = E_0 \int G(R, R')\, \psi_0(R')\, dR'$$

and multiply through by a function $\psi_T(R)$. Inside the integral multiply and divide by $\psi_T(R')$. The result is an integral equation of the same structure,

$$\psi_T(R)\, \psi_0(R) = E_0 \int [\psi_T(R)\, G(R, R')/\psi_T(R')]\, \psi_T(R')\, \psi_0(R')\, dR'$$

With

$$\tilde{\psi}(R) = \psi_T(R)\, \psi_0(R)$$

$$\tilde{G}(R) = [\psi_T(R)\, G(R, R')/\psi_T(R')]$$

we get

$$\tilde{\psi}(R) = E_0 \int \tilde{G}(R, R')\, \tilde{\psi}(R')\, dR'$$

The variance reducing properties follow from the fact that if

$$\psi_T(R) = \psi_0(R)$$

then

$$E_0 \int \tilde{G}(R, R')\, dR = 1 \qquad \text{for every} \quad R'$$

and that when ψ_T is close to ψ_0 the latter normalization is close to one. Sampling $\tilde{G}$ can be carried out in the domain or in the Brownian (i.e., diffusion) style.

For later purposes, it is useful for us to reconstruct the motivation for this transformation. It is one of a class of transformations in Monte Carlo methods that have the same structure: one seeks to alter the sampling so that the random points have a new density which is the original multiplied by the "value" or "importance" of the point. Since we are dealing with a homogeneous equation and seek asymptotic distributions, the importance of a walker at R' is the overlap with a test function $\psi_T(R)$ of the asymptotic distribution that arises from a walker at R'. That is, let

$$\eta_1(R; R') = E_0 \int G(R, R'')\, \delta(R'' - R')\, dR''$$

$$\vdots$$

$$\eta_{l+1}(R; R') = E_0 \int G(R, R'')\, \eta_l(R''; R')\, dR''$$

The importance $i(R')$ is calculated as

$$i(R') = \lim_{l \to \infty} \int \psi_T(R)\, \eta_l(R; R')\, dR$$

It is easily seen that

$$i(R') \propto \psi_0(R')$$

When we repeat this calculation for interacting ensembles of positive and negative walkers, we will find a different result.

2.2. Approaches to Fermion Calculations

2.2.1. Fixed Node Approximation. Anderson[4] was the first to treat interacting electrons by introducing an artificial boundary condition, namely that the solution must vanish at the nodes of an antisymmetric trial function $\psi_T(R)$. Since the correct nodes are in general unknown, and since the solution with this boundary condition on the opposite sides does not, in general, have continuous derivatives, the method is necessarily

approximate. It can, however, be very accurate, attaining chemical accuracy for few-electron systems. Unfortunately, it has proved difficult to improve systematically.

2.2.2. Transient Projection. If one iterates the Schrödinger equation in integral form (or alternatively simulates a diffusive process) from an initial guess that has an antisymmetric component, then the asymptotic antisymmetric component is exact. That is, let

$$\chi_1(R) = \psi_{AT}(R) + \psi_{ST}(R)$$

include antisymmetric and symmetric test functions, respectively.

Let

$$\chi_{l+1}(R) = E^* \int G(R, R')\, \chi_l(R')\, dR'$$

Then the fermion energy is exactly

$$E_F = \lim_{l \to \infty} \frac{\int \chi_l(R)\, H\psi_{AT}(R)\, dR}{\int \chi_l(R)\, \psi_{AT}(R)\, dR}$$

Unfortunately, for the reasons mentioned in the Introduction, the Monte Carlo estimates of the integrals that appear in numerator and denominator are asymptotically zero. More precisely, the signal-to-noise ratio decreases geometrically with l.

With a very good choice of ψ_{AT}, the estimates of E_F can be approximately constant yielding good results for the energy.

2.2.3. Correlated Walks. In 1982, David Arnow and his collaborators[5] introduced the idea of correlating pairs of walkers of opposite sign so as to achieve some degree of cancellation in their descendants and thereby avoid or delay the convergence to the ground state. They used populations of positive and negative walkers, $\{R_j^+\}$ and $\{R_j^-\}$, respectively. Then the Monte Carlo representation of the distribution at any stage is

$$\chi_l(R) = \sum_j [\delta(R - R_j^+) - \delta(R - R_j^-)]$$

and

$$\chi_{l+1}(R) = E^* \int G(R, R')\, \chi_l(R')\, dR'$$

$$= E^* \left[\sum G(R, R_j^+) - \sum G(R, R_j^-) \right]$$

When the R^+ and R^- are paired to enhance the overlap of $G(R, R_j^+)$ and $G(R, R_j^-)$ we may consider

$$G(R, R_j^+) - G(R, R_j^-)$$

as a composite propagator. Where this is positive, $\chi_{l+1}^+(R)$ is sampled; negative values sample $\chi_{l+1}^-(R)$. In the next section we will generalize the idea to correlate all walkers of both signs. Note, however, that the method so far does not break the plus/minus symmetry.

3. A MODEL PROBLEM

We will now introduce some integral identities that yield random walks like those that occur in Green's function Monte Carlo (GFMC).[2] These will permit simple analysis and intuitive understanding of that technique.

These integral equations have the form

$$\varphi(y) = E_0 \int G(y, z)\, W(z)\, \varphi(z)\, dz$$

where $G(y, z)$ is a stochastic kernel such that

$$G(y, z) = G(z, y)$$

$$G(y, z) \geqslant 0 \qquad \text{for all } y, z$$

$$\int G(y, z)\, dy = 1 \qquad \text{for every } z$$

and

$$W(z) \geqslant 0$$

A random walk that is described by this equation is as follows. A walker at z becomes N walkers ($N \geqslant 0$) with the expected value of N satisfying

$$\langle N \rangle = E_0 W(z)$$

A simple implementation is

$$N = \lfloor E_0 W(z) + \xi \rfloor$$

where ξ is a uniformly distributed random number on [0, 1].

We know introduce an importance function transformation. Let us use the eigenfunctions of our model equation:

$$\varphi_k(y) = E_k \int G(y, z)\, W(z)\, \varphi_k(z)\, dz$$

These are not orthogonal in the usual sense. It is easy to see, however, that the following generalization of the orthogonality relation holds:

$$\int \varphi_k(y)\, W(y)\, \varphi_l(y)\, dy = 0 \qquad \text{if} \quad E_l \neq E_k$$

Therefore, the functions $[W(y)]^{1/2}\, \varphi_k(y)$ are orthogonal. An arbitrary function $v(y)$ may be expanded as

$$v(y) = \sum_l c_l [W(y)]^{1/2}\, \varphi_l(y)$$

Writing $G(y, z)$ in terms of the functions of this complete set of eigenfunctions, we have

$$E_0 \int G(y, z)\, v(z)\, dz = \sum_l c_l \left(\frac{E_0}{E_l}\right) [W(y)]^{1/2}\, \varphi_l(y)$$

If one iterates many times,

$$\begin{aligned} E_0^n \int G(y, z_n) \int G(z_n, z_{n+1}) \cdots \int G(z_2, z_1)\, v(z_1)\, dz_1 \cdots dz_n \\ = \sum_l c_l \left(\frac{E_0}{E_l}\right)^n [W(y)]^{1/2}\, \varphi_l(y) \\ \to c_0 [W(y)]^{1/2}\, \varphi_0(y) \end{aligned}$$

As c_0 is given by

$$c_0 = \int [W(z)]^{1/2}\, \varphi_0(z)\, v(z)\, dz$$

if

$$v(z) = \delta(z - y_0)$$

we have

$$c_0 = [W(y_0)]^{1/2}\, \varphi_0(y_0)$$

Thus, the asymptotic population at y resulting from a walker that starts at y_0 is proportional to $W(y_0)\,\varphi_0(y_0)$. This is then the importance function. To use it, we multiply our integral equation through by $W(y)\,\varphi_0(y)$ and multiply and divide the integrand on the right by $\varphi_0(z)$ to obtain

$$\varphi_0(y)\,W(y)\,\varphi(y) = E_0 \int [W(y)\,\varphi_0(y)\,G(y,z)/\varphi_0(z)]\,\varphi_0(z)\,W(z)\,\varphi(z)\,dz$$

$$\tilde{\varphi}(y) = E_0 \int W(y)\,\tilde{G}(y,z)\,\tilde{\varphi}(z)$$

with

$$\tilde{\varphi}(y) = \varphi_0(y)\,W(y)\,\varphi(y)$$

and

$$\tilde{G}(y,z) = \varphi_0(y)\,G(y,z)/\varphi_0(z)$$

If we put our model equation in this form, we get

$$\exp[-(1+\beta)\,y^2] = \int \left[\frac{(1+2\beta)^2}{4\pi\beta}\right]^{1/2} \exp\left[\frac{-(1+2\beta)^2}{4\beta}\left(y - \frac{z}{1+2\beta}\right)^2\right] \times \exp[-(1+\beta)\,z^2]\,dz$$

with

$$\left[\frac{(1+2\beta)^2}{4\pi\beta}\right]^{1/2} \exp\left[\frac{-(1+2\beta)^2}{4\beta}\left(y - \frac{z}{1+2\beta}\right)^2\right] \equiv \tilde{G}(y,z)$$

and

$$\exp[-(\alpha+\beta)\,y^2] \equiv \tilde{\varphi}(y)$$

The transition kernel here satisfies

$$\int \tilde{G}(y,z)\,dz = 1 \qquad \text{for all } y,\ z$$

so that $N = 1$ identically and there is no multiplication (or termination) of the walkers. This circumstance, in which the population size has no variance, is clearly a more efficient random walk realization of the problem.

Also note that $\tilde{G}$ is not centered at z:

$$\tilde{G} \propto e^{-2(y-z/2)^2} \qquad \text{for} \quad \beta = 1/2$$

The walker is shifted to $z/2$ and the distribution is more narrow.

This $\tilde{G}$ was built using a knowledge of the exact eigenfunction. In practice we use an approximate trial function for φ_0, which reduces the fluctuations of the random walkers.

We can summarize our results for the model integral equation:

$$u_0(y) = e^{-y^2/2}, \qquad E_0 = 1$$

$$u_0(y) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(y-z)^2} \sqrt{2}\, e^{-z^2/2} u_0(z)\, dz$$

$$\sqrt{2}\, e^{-z^2/2} \equiv W(z), \qquad \text{plays role of potential}$$

$$u_k(y) = E_k \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(y-z)^2} \sqrt{2}\, e^{-z^2/2} u_k(z)\, dz$$

$$\tilde{u}_k(y) \equiv W(y)\, u_0(y)\, u_k(y)$$

$$\tilde{u}_k(y) = E_k \int_{-\infty}^{\infty} \left(\frac{2}{\pi}\right)^{1/2} e^{-2(y-z/2)^2} \tilde{u}_k(z)\, dz$$

$$u_1(y) = y e^{-y^2/2}, \qquad E_1 = 2$$

3.1. Ensembles of Interacting Walkers

The generalization of the method of Arnow *et al.* to this model problem and to the fully interacting ensembles is straightforward. We will seek to sample new points $y^{\pm}$ from the composite density

$$\sum_l [\tilde{G}(y, z_l^+) - \tilde{G}(y; z_l^-)]$$

We assume here that there are equal numbers of positive and negative walkers, a correct relation for ground states of fermion systems. Technically, we use a rejection method; that is, new points are sampled from a probability density proportional to

$$\sum_l [\tilde{G}(y, z_l^+) + \tilde{G}(y, z_l^-)]$$

When the ratio

$$\frac{\sum [\tilde{G}(y, z_l^+) - \tilde{G}(y, z_l^-)]}{\sum [\tilde{G}(y, z_l^+) + \tilde{G}(y, z_l^-)]} \leqslant 1$$

is positive, then y is accepted as a positive point y^+. Negative points y^-

are obtained by cycling through odd permutations of the particles and applying one to each y^+.

It is fruitful to think of the ensemble of walkers as constituting a pair of droplets of two kinds of atoms living in $3N$-dimensional space. For long-term stability these droplets, of positive and negative walkers, respectively, should be immiscible. The use of the full sum of Green's functions provides the analog of a pair potential, attractive between like atoms, repulsive between unlike atoms, that provides for the necessary stability. But the symmetry of positive and negative walkers remains. Thus, for any configuration of the droplets, another, with positive and negative walkers interchanged, is equally likely in the long-term solution of the system. We need a one-body potential that acts differently on positive and negative walkers to break the symmetry and to fully stabilize the system. As indicated above, the importance function provides the analog of a one-body force.

3.2. Importance Sampling Revisited

The developments of Section 2.1 apply here with two modifications. The first is that the test function $\psi(R)$ used there must be some antisymmetric function $\psi_{AT}(y)$. Second, the evolution of all walkers is coupled: a high density of negative walkers inhibits a positive walker from being selected in some neighborhood. Thus, any negative walker has an asymptotic distribution like the distribution of the other negative walkers, one which avoids the neighborhoods where positive walkers are dense. Taking into account the sign of the walkers, this asymptotic distribution has a positive overlap with $\psi_{AT}(y)$; the asymptotic distribution of positive walkers is mapped by the use of any odd permutation and has exactly the same overlap with $\psi_{AT}(y)$. Thus, the importance functions are both positive. Their behavior as a function of y cannot be written down in closed form; it depends on the details of the algorithm, including the size of the populations. However, their qualitative behavior is clear: $i^+(y)$, the importance for positive walkers, resembles the solution $\psi(y)$, or, one hopes, $\psi_{AT}(y)$, where they are positive and not too close to their nodal surfaces.

Where $\psi_{AT}(y)$ is negative, $i^+(y)$ falls off rapidly, reflecting the increasing density of negative walkers. A form such as

$$i^+(y) = \int G(y, z) \max[\psi_{AT}(z), 0]\, dz$$

might plausibly be used with a similar *ansatz* for $i^-(y)$. The form used in our test problem will be described below.

3.3. A Numerical Experiment

Several programs were written and run in the course of developing these ideas and in exploring their consequences. We describe here some of the characteristics of the last, the one that most completely embodies the characteristics described above.

We attempted to sample populations of positive and negative walkers each having a stable overlap with the solution of the model equation for a system of three particles, each in three dimensions, and all polarized in the same direction. For the direction chosen, that is,

$$\psi(R) = \begin{vmatrix} 1 & 1 & 1 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix} \exp\left(-\frac{1}{2}R^2\right)$$

A fairly poor trial function was chosen as follows:

$$\varphi_0(x) = \exp[-(x+x_s)^2] + \exp[-(x-x_s)^2]$$
$$\varphi_1(x) = \exp[-(x+x_s)^2] - \exp[-(x-x_s)^2]$$

We build up orbitals as products

$$\Phi_{ijk}(r) = \varphi_i(x)\,\varphi_j(y)\,\varphi_k(z)$$

and use as trial or test function

$$\psi_{AT}(R) = \begin{vmatrix} \Phi_{000}(r_1) & \Phi_{000}(r_2) & \Phi_{000}(r_3) \\ \Phi_{010}(r_1) & \Phi_{010}(r_2) & \Phi_{010}(r_3) \\ \Phi_{001}(r_1) & \Phi_{001}(r_2) & \Phi_{001}(r_3) \end{vmatrix}$$

The only parameter is x_s, which can be varied to minimize the eigenvalue. The value $x_s = 0.625$ yields $\lambda_v = 4.28$. The use of a crude trial function was deliberate: we did not want a spurious convergence or stability to derive from a special choice of ψ_{AT}. The importance function is derived from the trial function by means of an intermediate "mapping" function $\mu(s)$, which was chosen as

$$\mu(s) = \tfrac{1}{2}[s - (s^2 + \alpha^2)^{1/2}]$$

so that

$$i^{\pm}(R) = \exp\{\pm\beta\mu[\psi_{AT}(R)]\}$$

The parameters α and β can be adjusted to minimize the population fluctuations, but after a quick verification that the computation was insensitive to them, intermediate plausible values of $\alpha = 0.4$, $\beta = 0.4$ were used.

Separate calculations were run with populations of each kind of walker sets at 50, 100, and 200 in turn. There was no evidence of "instability" or decay of signal-to-noise ratio. In fact, the smaller population system was iterated 2×10^4 times. If one notes that the ratio of eigenvalues is 4, and that the state singled out for solution is degenerate and a second excited state, it is clear that without stabilization by the importance functions, noise would have dominated within a few iterations.

The computed eigenvalue was observed to be very nearly a linear function of L^{-1}, where L is the fixed population size. Extrapolated to $L^{-1} = 0$, the eigenvalue is

$$\lambda = 3.9975 \pm 0.0024$$

in satisfactory agreement with the correct value of 4.

4. CONCLUSION

The techniques outlined above for stabilization and projection of an antisymmetric excited state appear, on the basis of a simple but exigent model calculation, to work. The methods have been adapted to calculate properties of few-electron systems, and tests of this new algorithm are now underway.

ACKNOWLEDGMENTS

Much of the theoretical development and most of the programming of the model problem took place during a visit by the author to the Laboratoire de Physique Théorique des Liquides of the Université Pierre et Marie Curie (Jussieu). The author is happy to express his gratitude to Michel Moreau and Claire Lhuillier for their hospitality and support during that visit. It is a pleasure to acknowledge the help of Liane Strantz in the preparation of the manuscript and Prof. Lhuillier, Silvio Vitiello, and Shi-wei Zhang for their helpful discussions regarding this work. It was supported in part by the Université Pierre et Marie Curie. The Cornell Theory Center is funded by the U.S. National Science Foundation, by New York State, by IBM, and by Cornell University.

REFERENCES

1. K. E. Schmidt and M. H. Kalos, in *Applications of the Monte Carlo Method in Statistical Physics*, K. Binder, ed. (Springer-Verlag, 1984).
2. D. M. Ceperley and M. H. Kalos, in *Monte Carlo Methods in Statistical Physics*, K. Binder, ed. (Springer-Verlag, 1979); M. H. Kalos, D. Levesque, and L. Verlet, *Phys. Rev. A* **9**:2178 (1974).
3. D. M. Ceperley and B. J. Alder, *J. Chem. Phys.* **81**:5833 (1984); C. J. Umrigar, K. J. Runge, and M. P. Nightingale, in *Monte Carlo Methods in Theoretical Physics*, S. Caracciolo and A. Fabrocini, eds. (ETS Editrice, Pisa, 1991).
4. J. B. Anderson, *J. Chem. Phys.* **63**:1499 (1975); **65**:4121 (1976); **73**:3897 (1980).
5. D. M. Arnow, M. H. Kalos, M. A. Lee, and K. E. Schmidt, *J. Chem. Phys.* **77**:5562 (1982).