Generalization of Metropolis and heat-bath sampling for Monte Carlo simulations

Bruce M. Boghosian*

Center for Computational Science, Boston University, 3 Cummington Street, Boston, Massachusetts 02215
(Received 23 March 1998)

For a wide class of applications of the Monte Carlo method, we describe a general sampling methodology that is guaranteed to converge to a specified equilibrium distribution function. The method is distinct from that of Metropolis in that it is sometimes possible to arrange for unconditional acceptance of trial moves. It involves sampling states in a local region of phase space with probability equal to, in the first approximation, the square root of the desired global probability density function. The validity of this choice is derived from the Chapman-Kolmogorov equation, and the utility of the method is illustrated by a prototypical numerical experiment. [S1063-651X(98)04408-0]

In Monte Carlo calculations, it is often necessary to sample from a probability density function (PDF) that is known only to within a multiplicative constant. That is, we often know an analytic form for

$$F(\mathbf{x}) = Z f(\mathbf{x}),$$

where $f(\mathbf{x})$ is the PDF from which we would like to sample, and where the constant Z is unknown. In computational statistical physics, for example, it is often necessary to sample from the Boltzmann-Gibbs distribution

$$f(\mathbf{x}) = \frac{1}{Z} \exp[-\beta H(\mathbf{x})], \tag{1}$$

where $\mathbf{x}$ coordinatizes the phase space, $H(\mathbf{x})$ is the system Hamiltonian, and β is the inverse temperature. A simple analytic form is generally available for $H(\mathbf{x})$ but not for Z. Indeed, if such a form were known for Z, it would not be necessary to resort to Monte Carlo methods.

One way to break this apparent impasse was provided by an ingenious algorithm due to Metropolis *et al.* [1] in 1953. This algorithm provides a Markov process on the state space whose time-asymptotic state is the desired PDF, $f(\mathbf{x})$, but whose transition kernel $K(\mathbf{y} \rightarrow \mathbf{x})$ involves only *ratios* of this PDF. It does this by noting that a sufficient condition for a stationary solution of the Chapman-Kolmogorov equation for the Markov process,

$$f(\mathbf{x}) = \int d\mathbf{y}\, f(\mathbf{y}) K(\mathbf{y} \rightarrow \mathbf{x}), \tag{2}$$

is that of *detailed balance*,

$$f(\mathbf{x}) K(\mathbf{x} \rightarrow \mathbf{y}) = f(\mathbf{y}) K(\mathbf{y} \rightarrow \mathbf{x}). \tag{3}$$

To see this, we need only integrate both sides of Eq. (3) with respect to $\mathbf{y}$, using the normalization constraint on the transition kernel,

$$\int d\mathbf{y}\, K(\mathbf{x} \rightarrow \mathbf{y}) = 1; \tag{4}$$

the result is Eq. (2). We then note that the choice

$$K(\mathbf{y} \rightarrow \mathbf{x}) = T(\mathbf{y} \rightarrow \mathbf{x}) \min\left[1, \frac{f(\mathbf{x}) T(\mathbf{x} \rightarrow \mathbf{y})}{f(\mathbf{y}) T(\mathbf{y} \rightarrow \mathbf{x})}\right] \tag{5}$$

satisfies Eq. (3) and involves only ratios of the PDF.

A simple application of the Metropolis method to a statistical physics problem might then begin with state $\mathbf{y}$ and select a new state $\mathbf{x}$ according to some *trial PDF*, $T(\mathbf{y} \rightarrow \mathbf{x})$. A convenient (but not necessary) condition on the trial PDF is

$$T(\mathbf{y} \rightarrow \mathbf{x}) = T(\mathbf{x} \rightarrow \mathbf{y}). \tag{6}$$

This new trial state is then *accepted* with probability

$$\min\{1, \exp\{-\beta[(H(\mathbf{x}) - H(\mathbf{y})]\}\},$$

so that the net transition probability is given by Eq. (5). Thus, the trial state is accepted with probability unity if it results in a decrease of energy, and with probability exp$(-\beta \Delta H)$ if it results in an energy increase of ΔH. If the trial is not accepted, then the current state is retained. More sophisticated implementations of the Metropolis algorithm bias the selection of a trial state so that Eq. (6) is violated [2], but the detailed-balance condition, Eq. (3), is maintained.

In this paper, we describe a fundamentally different kind of Markov process for the sampling of a PDF $f(\mathbf{x})$ that is known only to within a multiplicative constant. The method is distinct from that of Metropolis in that moves may, in some circumstances, be accepted unconditionally. Nevertheless, it admits $f(\mathbf{x})$ as a time-asymptotic solution of the Chapman-Kolmogorov equation, Eq. (2). The proposed Markov process uses a transition rate of the form

*Electronic address: bruceb@bu.edu

$$K(\mathbf{y}\rightarrow\mathbf{x})=\frac{T(\mathbf{y}\rightarrow\mathbf{x})g(\mathbf{x})}{\displaystyle\int d\mathbf{z}\,T(\mathbf{y}\rightarrow\mathbf{z})g(\mathbf{z})}, \tag{7}$$

where $T(\mathbf{y}\rightarrow\mathbf{x})$ is the trial PDF, and $g(\mathbf{x})$ is to be determined in terms of the desired time-asymptotic state $f(\mathbf{x})$. Note that this form obeys Eq. (4) manifestly. Essentially, it tells us to sample from $g(\mathbf{x})$, normalized over only those states accessible by the trial PDF $T(\mathbf{y}\rightarrow\mathbf{x})$. We note that the dependence of $g(\mathbf{x})$ on $f(\mathbf{x})$ may well be complicated (nonlinear and/or nonlocal), but it must be such that the transition kernel of Eq. (7) is unchanged when $f(\mathbf{x})$ is scaled by a multiplicative constant.

If we insist that the trial process be symmetric in the sense of Eq. (6), then it is straightforward to verify that the Chapman-Kolmogorov equation, Eq. (2), with transition kernel given by Eq. (7), is satisfied by

$$f(\mathbf{x})=g(\mathbf{x})\int d\mathbf{y}\,T(\mathbf{y}\rightarrow\mathbf{x})g(\mathbf{y}). \tag{8}$$

Thus we must invert Eq. (8) to get $g(\mathbf{x})$ in terms of $f(\mathbf{x})$. Equation (8) is a nonlinear integral equation for $g(\mathbf{x})$, but the reward for inverting it is a Markov process with specified trial PDF, $T(\mathbf{y}\rightarrow\mathbf{x})$, and stationary state $f(\mathbf{x})$. The utility of Eq. (8) is the central observation of this paper.

One limit in which Eq. (8) has a trivial solution is that for which all states are equally accessible. If $T(\mathbf{y}\rightarrow\mathbf{x})$ is independent of $\mathbf{x}$, it is easy to see that Eq. (8) is solved by $g(\mathbf{x})\propto f(\mathbf{x})$. This solution is the *heat-bath* Monte Carlo algorithm. The proposed method can thus be regarded as a generalization of the heat-bath algorithm for nonuniform trial PDF's.

We can better understand the character of the solutions to Eq. (8) if we rewrite it as follows:

$$g(\mathbf{x})=\sqrt{\frac{f(\mathbf{x})}{\displaystyle\int d\mathbf{y}\,T(\mathbf{x}\rightarrow\mathbf{y})[g(\mathbf{y})/g(\mathbf{x})]}}. \tag{9}$$

Let us suppose that $g(\mathbf{y})$ is reasonably constant over the set of $\mathbf{y}$ for which $T(\mathbf{x}\rightarrow\mathbf{y})$ is appreciable. For simplicity, let us also suppose that the trial PDF is normalized,

$$\int d\mathbf{y}\,T(\mathbf{y}\rightarrow\mathbf{x})=1.$$

Then the integral in the denominator is nearly unity, and $g(\mathbf{x})\approx\sqrt{f(\mathbf{x})}$ to first approximation. In this case, it makes sense to solve Eq. (9) by the following successive approximation scheme:

$$g(\mathbf{x})=\lim_{\ell\rightarrow\infty}g^{(\ell)}(\mathbf{x}),$$

$$g^{(0)}(\mathbf{x})=\text{const},$$

$$g^{(\ell)}(\mathbf{x})=\sqrt{\frac{f(\mathbf{x})}{\displaystyle\int d\mathbf{y}\,T(\mathbf{x}\rightarrow\mathbf{y})[g^{(\ell-1)}(\mathbf{y})/g^{(\ell-1)}(\mathbf{x})]}}.$$

The first few approximations to $g(\mathbf{x})$ are then

$$g^{(1)}(\mathbf{x})=\sqrt{f(\mathbf{x})}$$

$$g^{(2)}(\mathbf{x})=\sqrt{\frac{f(\mathbf{x})}{\displaystyle\int d\mathbf{y}\,T(\mathbf{x}\rightarrow\mathbf{y})[\sqrt{f(\mathbf{y})/f(\mathbf{x})}]}}$$

$$\text{(10)}$$

$$g^{(3)}(\mathbf{x})=\left(\frac{f(\mathbf{x})}{\displaystyle\int d\mathbf{y}\,T(\mathbf{x}\rightarrow\mathbf{y})\left\{f(\mathbf{y})\int d\mathbf{z}\,T(\mathbf{x}\rightarrow\mathbf{z})\sqrt{[f(\mathbf{z})/f(\mathbf{x})]}\Big/ f(\mathbf{x})\int d\mathbf{z}\,T(\mathbf{y}\rightarrow\mathbf{z})\sqrt{[f(\mathbf{z})/f(\mathbf{y})]}\right\}}\right)^{1/2}$$

$$g^{(4)}(\mathbf{x})=\left(\frac{f(\mathbf{x})}{\displaystyle\int d\mathbf{y}\,T(\mathbf{x}\rightarrow\mathbf{y})\dfrac{f(\mathbf{y})\int d\mathbf{z}T(\mathbf{x}\rightarrow\mathbf{z})\left(f(\mathbf{z})\int d\mathbf{w}\,T(\mathbf{x}\rightarrow\mathbf{w})\sqrt{[f(\mathbf{w})/f(\mathbf{x})]}\Big/ f(\mathbf{x})\int d\mathbf{w}\,T(\mathbf{z}\rightarrow\mathbf{w})\sqrt{[f(\mathbf{w})/f(\mathbf{z})]}\right)}{f(\mathbf{x})\int d\mathbf{z}\,T(\mathbf{y}\rightarrow\mathbf{z})\left(f(\mathbf{z})\int d\mathbf{w}\,T(\mathbf{y}\rightarrow\mathbf{w})\sqrt{[f(\mathbf{w})/f(\mathbf{y})]}\Big/ f(\mathbf{y})\int d\mathbf{w}\,T(\mathbf{z}\rightarrow\mathbf{w})\sqrt{[f(\mathbf{w})/f(\mathbf{z})]}\right)}}\right)^{1/2}.$$

In all cases, note that if $f(\mathbf{x})$ is multiplied by a factor C, the $g^{(\varnothing)}(\mathbf{x})$ scale by an overall factor $\sqrt{C}$, but this factor cancels when inserted into Eq. (7) for the transition kernel. Thus the transition kernel depends only on ratios of the PDF, as promised.

At this point, we make two observations: First, all of the above considerations apply to a discrete state space, if we just replace the integrals by sums as appropriate. Second, all of the above considerations can be applied to situations for which $T(\mathbf{x}\rightarrow\mathbf{x})=0$, so that the walker steps to another state with probability unity, though there are no guarantees that Eq. (8) will admit a solution in that case.

As a first simple example of this methodology for which the integral equation, Eq. (8), may be solved *without any approximation at all*, consider a five-state system with $\mathbf{x},\mathbf{y}\in\{1,2,3,4,5\}$, so that $T(\mathbf{x}\rightarrow\mathbf{y})$ may be represented as a matrix which we take to be

$$T=\begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Note that the diagonal elements $T(\mathbf{x}\rightarrow\mathbf{x})=0$ so that moves that leave the state unchanged are expressly forbidden. The analog of the integral equation, Eq. (8), is the set of nonlinear algebraic equations

$$f_j=g_j(g_{j-1}+g_{j+1}),$$

where we have used subscripts modulo 5 to denote functional arguments for simplicity. We note that scaling the g_j by C and the f_j by C^2 leaves the system unchanged, and we search for solutions only to within an arbitrary multiplicative constant. It may be verified that the following is such a solution:

$$\begin{aligned} g_j=C(&+f_j+f_{j+1}-f_{j+2}+f_{j+3}-f_{j+4}) \\ \times(&+f_j-f_{j+1}+f_{j+2}+f_{j+3}-f_{j+4}) \\ \times(&+f_j-f_{j+1}+f_{j+2}-f_{j+3}+f_{j+4}), \end{aligned} \quad (11)$$

where again all subscripts are taken modulo 5. The constant C turns out to be equal to

$$C=\frac{1}{\sqrt{2(-f_1+f_2-f_3+f_4+f_5)(-f_1+f_2+f_3-f_4-f_5)(+f_1+f_2-f_3+f_4-f_5)(+f_1-f_2+f_3+f_4-f_5)(+f_1-f_2+f_3-f_4+f_5)}},$$

but it is unnecessary to know this in order to use this solution, since only the ratios of the g_j's matter.

To be concrete, suppose that we take $f_1=1, f_2=2, f_3=3, f_4=2$, and $f_5=1$; obviously, these should be normalized by the factor $1/9$, but let us pretend that we do not know this normalization factor. The above equations give $g_1=g_2=g_4=g_5=3C$, and $g_3=9C$. Thus, if we simulate a Markov process with transition matrix

$$K=\begin{pmatrix} 0 & \frac{g_2}{g_2+g_5} & 0 & 0 & \frac{g_5}{g_2+g_5} \\ \frac{g_1}{g_1+g_3} & 0 & \frac{g_3}{g_1+g_3} & 0 & 0 \\ 0 & \frac{g_2}{g_2+g_4} & 0 & \frac{g_4}{g_2+g_4} & 0 \\ 0 & 0 & \frac{g_3}{g_3+g_5} & 0 & \frac{g_5}{g_3+g_5} \\ \frac{g_1}{g_4+g_1} & 0 & 0 & \frac{g_4}{g_4+g_1} & 0 \end{pmatrix}=\begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix},$$

the result will be the desired equilibrium state. Note that this Markov process is nothing like the Metropolis algorithm, in that trial states are accepted unconditionally. Also note that it is distinct from the *heat-bath* algorithm, wherein *all* states must be sampled with probability proportional to $f(\mathbf{x})$.

It is of course not always possible to do what we have done in the simple example above. If we have one state that is much more probable than any other, our intuition tells us

that a Markov process cannot force us to leave that state with probability unity every time we are in it, since the result would be our spending half of our time in a much less probable state. To see how the solution for the five-state system breaks down in that case, let us suppose that state 3 has a high probability f_H, and the other four states have a low probability f_L. We find that $g_1=g_5=(2f_L-f_H)f_H^2$, $g_2=g_4=(2f_L-f_H)^2f_H$, and $g_3=f_H^3$. We note that $\text{sgn}(g_1)$

TABLE I. Iterative solution with $N=5$.

$g_j^{(\ell)}/g_1^{(\ell)}$	$\ell=1$	$\ell=25$	$\ell=50$	$\ell=75$	$\ell=100$
$j=2,4$	1.41421	1.20022	1.05966	1.01876	1.00600
$j=3$	1.73205	2.72733	2.90754	2.96972	2.99019

$=\mathrm{sgn}(g_5)$ and $\mathrm{sgn}(g_2)=\mathrm{sgn}(g_3)=\mathrm{sgn}(g_4)$, but that these signs will not be the same if $f_H>2f_L$. If they are not the same, our transition matrix will involve negative probabilities. Hence we arrive at another requirement on the solutions of Eq. (8): Any legitimate solution for $g(\mathbf{x})$ must be of definite sign.

It is one thing to treat a five-state system, but quite another to treat, say, a two-dimensional Ising model on a 10 $\times 10$ lattice, for which there are 2^{100} distinct states. In such a situation, it will be no more possible to store $g(\mathbf{x})$ on a computer than it is to store $f(\mathbf{x})$. Thus, we must search for ways to obtain partial solutions to Eq. (8) for large systems *on the fly*; more specifically, if we are in a state $\mathbf{x}$, we must be able to easily compute $g(\mathbf{y})$ for all the states for which $T(\mathbf{x}\rightarrow\mathbf{y})$ is nonzero. This is enough information to sample the next state and then we can repeat the process. So for the remainder of this paper we turn our attention to larger systems.

It is likely that analytic solutions in the spirit of Eqs. (11) exist for arbitrarily large systems, but they seem to become correspondingly more difficult to find. What is interesting is that the iterative process given in Eqs. (10) actually becomes *easier* to solve for larger systems, as long as the distribution $f(\mathbf{x})$ is smooth in the state space. To see this, let us consider an N-state system with periodic, nearest-neighbor state connectivity and

$$f_j=1+2\,\sin^2\left(\frac{\pi j}{N-1}\right)$$

for $j=0,\ldots,N-1$; note that when $N=5$ this reduces to the numerical example used above for which we know that the exact solution for the g_j obeys $g_2/g_1=g_4/g_1=g_5/g_1=1$ and $g_3/g_1=3$. (Recall that only *ratios* of the g_j's matter.) These results are not particularly close to the first approximation of the iterative process of Eq. (10), which would be $g_5^{(1)}/g_1^{(1)}=\sqrt{f_5/f_1}=1$, $g_{2,4}^{(1)}/g_1^{(1)}=\sqrt{f_{2,4}/f_1}=\sqrt{2}$, and $g_3^{(1)}/g_1^{(1)}=\sqrt{f_3/f_1}=\sqrt{3}$. This is because that iterative process was predicated on the smoothness of $f(\mathbf{x})$, and there is no obvious sense in which a function defined on five discrete points is smooth. Nevertheless, as Table I illustrates, the iteration does appear to converge. Table II shows that when

we let $N=50$ (without making the function f any more rapidly varying), the iteration converges much faster. (This is especially evident if we compare the values of the iteration number ℓ in the two tables.) Moreover, we see that the asymptotic result is not very different from the first approximation, which is proportional to the square root of the distribution function. This suggests that we develop a perturbation theory for $g(\mathbf{x})$ about $\sqrt{f(\mathbf{x})}$, and we turn our attention to this approach below.

Let us restrict our attention to continuous state spaces, and suppose that $T(\mathbf{x}\rightarrow\mathbf{y})=W(\|\mathbf{x}-\mathbf{y}\|)$, for some norm $\|\cdot\|$, where the region of support of $W(\|\mathbf{y}\|)$ is very small compared to the characteristic scale(s) of $f(\mathbf{x})$. Then we can Taylor expand $g(\mathbf{y})$ about the point $\mathbf{x}$ in Eq. (8). In particular, if we then suppose that the moments of the function W are

$$\int d\mathbf{y}\,W(\|\mathbf{y}\|)=1,$$

$$\int d\mathbf{y}\,W(\|\mathbf{y}\|)\mathbf{y}=\mathbf{t}, \qquad (12)$$

$$\int d\mathbf{y}\,W(\|\mathbf{y}\|)\mathbf{y}\mathbf{y}=\mathbf{t}\mathbf{t}+\boldsymbol{\sigma},$$

where $\mathbf{t}$ is the mean and $\boldsymbol{\sigma}$ is the variance tensor, then Eq. (8) becomes

$$f(\mathbf{x})=g(\mathbf{x})[g(\mathbf{x})+\mathbf{t}\cdot\boldsymbol{\nabla}g(\mathbf{x})+\tfrac{1}{2}(\mathbf{t}\mathbf{t}+\boldsymbol{\sigma}):\boldsymbol{\nabla}\boldsymbol{\nabla}g(\mathbf{x})+\cdots]$$

$$=[g(\mathbf{x})]^2\left[1+\frac{\mathbf{t}\cdot\boldsymbol{\nabla}g(\mathbf{x})}{g(\mathbf{x})}+\frac{(\mathbf{t}\cdot\boldsymbol{\nabla})^2g(\mathbf{x})}{2g(\mathbf{x})}\right.$$

$$\left.+\frac{\boldsymbol{\sigma}:\boldsymbol{\nabla}\boldsymbol{\nabla}g(\mathbf{x})}{2g(\mathbf{x})}+\cdots\right].$$

We can invert the above equation perturbatively to find

TABLE II. Iterative solution with $N=50$.

$g_j^{(\ell)}/g_1^{(\ell)}$	$\ell=1$	$\ell=2$	$\ell=3$	$\ell=4$	$\ell=100$
$j=10$	1.26302	1.26408	1.26378	1.26361	1.26261
$j=20$	1.66176	1.66450	1.66410	1.66388	1.66301
$j=30$	1.68466	1.68747	1.68707	1.68685	1.68598
$j=40$	1.30976	1.31107	1.31076	1.31059	1.31018

$$g = \sqrt{f}\left[1 - \frac{\mathbf{t}\cdot\boldsymbol{\nabla} f}{4f} + \frac{(\mathbf{t}\cdot\boldsymbol{\nabla} f)^2}{32f^2} - \frac{\boldsymbol{\sigma}:\boldsymbol{\nabla}\boldsymbol{\nabla} f}{8f} + \frac{\boldsymbol{\sigma}:(\boldsymbol{\nabla} f)(\boldsymbol{\nabla} f)}{16f^2} + \cdots\right],$$

where all functions and derivatives are understood to be evaluated at $\mathbf{x}$. Once again, it is manifest that the transition kernel, Eq. (7), computed from this $g(\mathbf{x})$ will depend only on ratios of $f(\mathbf{x})$. When the desired distribution function $f(\mathbf{x})$ is of Boltzmann-Gibbs form, as in Eq. (1), the above equation reduces to

$$g = e^{-\beta H/2}\left[1 + \frac{\beta}{4}\mathbf{t}\cdot\boldsymbol{\nabla} H + \frac{\beta^2}{32}(\mathbf{t}\cdot\boldsymbol{\nabla} H)^2 + \frac{\beta}{8}\boldsymbol{\sigma}:\boldsymbol{\nabla}\boldsymbol{\nabla} H - \frac{\beta^2}{16}\boldsymbol{\sigma}:(\boldsymbol{\nabla} H)(\boldsymbol{\nabla} H) + \cdots\right]. \tag{13}$$

This form is potentially useful for Monte Carlo simulations of molecular systems, for which the function $H(\mathbf{x})$ is known in analytic form and the state space is continuous, so that gradients of $H(\mathbf{x})$ can also be computed in analytic form. It is interesting that the first approximation to $g(\mathbf{x})$ in this case is $\sqrt{f(\mathbf{x})}$, which is also of Boltzmann-Gibbs form, but at *twice* the desired temperature.

To apply Eq. (13) to a molecular system with potential energy

$$V(\mathbf{q}) = \frac{1}{2}\sum_{i,j}^{N} v(|\mathbf{q}_i - \mathbf{q}_j|)$$

and PDF $\exp[-\beta V(\mathbf{q})]/Z$, we calculate g from Eq. (13) (with $H \to V$), and consider moves that take the walker at $\mathbf{q} = (\mathbf{q}_1, \ldots, \mathbf{q}_N)$ to one at

$$\mathbf{q} + \Delta\mathbf{q} = (\mathbf{q}_1 + \sigma\hat{\mathbf{e}}_1, \ldots, \mathbf{q}_N + \sigma\hat{\mathbf{e}}_N),$$

where the $\hat{\mathbf{e}}_i$ are spatial unit vectors, sampled with probability $g(\mathbf{q} + \Delta\mathbf{q})$. Note that all of the particles must move a distance σ in this scheme. The sampling of unit vectors $\hat{\mathbf{e}}_i$ may be carried out by the Metropolis or any other suitable method, but after that is completed the new walker position is $\mathbf{q} + \Delta\mathbf{q}$ with probability unity. For sufficiently small σ, this method is guaranteed to converge and unconditionally accepts moves. A potential pitfall is that this sufficiently small value of σ may be very small indeed.

As a final example, we turn our attention to the Ising model with N spins $\mathbf{s}$ and Hamiltonian

$$H(\mathbf{s}) = \frac{1}{2}\sum_{i}^{N}\sum_{j\in\mathcal{N}_i} Js_i s_j,$$

where $\mathcal{N}_i$ denotes the neighborhood of spin i. Let us introduce the operator $\hat{\theta}_i$ such that $\hat{\theta}_i\mathbf{s}$ is the state obtained from state $\mathbf{s}$ by flipping spin i. We also denote the resulting change in observable $F(\mathbf{s})$ by

$$\hat{\delta}_i F(\mathbf{s}) \equiv F(\hat{\theta}_i\mathbf{s}) - F(\mathbf{s}).$$

Let us suppose that our trial PDF can take us from state $\mathbf{s}$ to any one of the N states $\hat{\theta}_i\mathbf{s}$ for $i = 1, \ldots, N$. Once again, this condition expressly forbids the system from remaining in its present state. The analog of our integral equation, Eq. (8), is then

$$e^{-\beta H(\mathbf{s})} = g(\mathbf{s})\frac{1}{N}\sum_{i}^{N} g(\hat{\theta}_i\mathbf{s}) = g(\mathbf{s})\frac{1}{N}\sum_{i}^{N}[g(\mathbf{s}) + \hat{\delta}_i g(\mathbf{s})]$$

$$= [g(\mathbf{s})]^2\left[1 + \frac{1}{N}\sum_{i}^{N}\frac{\hat{\delta}_i g(\mathbf{s})}{g(\mathbf{s})}\right],$$

so that the analog of our approximation method, Eq. (10) is

$$g(\mathbf{s}) = \frac{e^{-\beta H(\mathbf{s})/2}}{\sqrt{1 + (1/N)\sum_{i}[\hat{\delta}_i g(\mathbf{s})/g(\mathbf{s})]}}.$$

Solving this involves $O(N)$ work at the very least and will yield N values of g, one for each proposed new state $\hat{\theta}_i\mathbf{s}$. We can sample from these to choose a site whose spin is then flipped with probability unity. As with our five-state problem, this is likely to become impossible when one state is much more probable than any of its neighbors, and this is likely to happen for sufficiently low temperature. A more fundamental problem is that the method as stated requires $O(N)$ work to flip one spin (albeit with unit probability). This is clearly prohibitive, but it may be possible to drastically improve this situation by more clever choices of the trial PDF. In particular, it would be interesting to study the relationship of this method to the various cluster algorithms that are known for the Ising model [3].

In conclusion, the integral equation, Eq. (8), is a very useful way to frame the problem of developing a Markov process with specified equilibrium and trial PDF. The method is not a panacea, in that it will not always be possible to find a solution for $g(\mathbf{x})$ with definite sign. The breakdown of the method when g has indefinite sign is reminiscent of the *sign problem* that is sometimes encountered in quantum Monte Carlo simulations [4], and it would be interesting to

study this relationship further. Moreover, unless a trial PDF is chosen very judiciously, the amount of work involved in finding g may mitigate the advantage of unconditional acceptance, as with the Ising model example above. In any case, we have shown that such a solution may be constructed perturbatively for sufficiently small step sizes in a continuous state space, and this observation ought to have immediate application to molecular Monte Carlo problems, as we have also shown. In addition, our application of this methodology to stochastic lattice Boltzmann methods is presented in one of the references [5].

[1] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[2] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, San Diego, 1996) and references therein.

[3] For a good review of this subject, see G. T. Barkema and M. E. J. Newman, in *Monte Carlo Methods in Chemical Physics*, edited by D. Ferguson, J. I. Siepmann, and D. G. Truhlar (Wiley, New York, 1997).

[4] See, for example, G. G. Batrouni and P. de Forcrand, Phys. Rev. B **48**, 589 (1993).

[5] B. M. Boghosian, J. Yepez, P. V. Coveney, and A. Wagner, Report No. BU-CCS-990701 (1999).