JOURNAL OF COMPUTATIONAL PHYSICS 24, 150–157 (1977)

Monte Carlo in Statistical Mechanics: Choosing between Alternative Transition Matrices

JOHN P. VALLEAU AND STUART G. WHITTINGTON

Department of Chemistry, University of Toronto, Toronto, Canada, M5S 1A1

Received September 30, 1976; revised November 17, 1976

The transition matrices for sampling along a realization of a Markov chain, suggested by Flinn and McManus (*Phys. Rev.* **124** (1961), 54) and by Metropolis *et al.* (*J. Chem. Phys.* **21** (1953), 1087) are compared by determining the convergence rates for several simple model systems. Although it is possible to construct model systems for which the Flinn and McManus matrix leads to more rapid convergence we argue that the Metropolis matrix will be superior, from a computational point of view, for many systems including all those (such as fluids) with many states accessible at each step.

INTRODUCTION

Since their inception in 1953, Metropolis-type Monte Carlo methods [1, 2] have played a useful role in the statistical mechanical studies of many-body systems. The idea of the method is to sample along a realization of a Markov chain, defined on the configuration space of the system, such that the unique limit distribution of the Markov chain is the distribution of physical interest (e.g., the Boltzmann distribution). Metropolis and coworkers [1] suggested an ingenious way of constructing a Markov chain, with Boltzmann limiting distribution, so that it is unnecessary to evaluate the partition function when computing mechanical quantities.

Of course it is possible to construct many different Markov chains which have the same unique limit distribution and one needs a criterion by which to determine whether one is better than another from a computational point of view.

Consider a transition matrix $\mathbf{P}$ defined on the set of states $i = 1, 2, 3,...$ such that the limit distribution of the Markov chain is $\{\pi_i\}$. Two popular forms for $\mathbf{P}$ are the original suggestion of Metropolis *et al.* [1]

$$p_{ij}^{M} = q_{ij} \min\{1, \pi_j/\pi_i\}, \qquad i \neq j, \tag{1}$$

or another suggested by Flinn and McManus [3] (FM)

$$p_{ij}^{F} = q_{ij}\pi_j/(\pi_i + \pi_j), \qquad i \neq j, \tag{2}$$

with p_{ii} in either case chosen to make $\mathbf{P}$ stochastic:

$$p_{ii} = 1 - \sum_{j \neq i} p_{ij}\,; \tag{3}$$

ISSN 0021-9991

q_{ij} are the elements of a symmetric but otherwise arbitrary stochastic matrix. We examine the usual case in which the q_{ij} are not themselves explicit functions of π_i, π_j. Both of these Markov chains have the required limit distribution so that both appear to be appropriate for a Monte Carlo calculation. However, in certain circumstances one of them may have computational advantages over the other.

Peskun [4] has discovered the very important and interesting result that the asymptotic variance of any mechanical quantity is never less for the Flinn and McManus form than the corresponding asymptotic variance for the Metropolis form. This demonstrates that the Metropolis form is always superior for very long realizations of the Markov chain. Unfortunately one is always limited to finite (and often to quite short) realizations in practice and a second concern is therefore the rate of convergence to the limit distribution.

Recently Cunningham and Meijer [5] compared the Metropolis and FM schemes for an Ising problem, and came to the conclusion that the FM method led to a smaller variance and more rapid convergence. On the other hand, most workers in the field have favored the Metropolis method, on the basis of somewhat intuitive arguments which suggest it may lead to more rapid convergence. This apparent disagreement has led us to reexamine the problem. In particular we have studied analytically a number of simple model systems in an attempt to understand the circumstances which will make one method or the other superior from the point of view of convergence rates.

The intuitive considerations supporting the Metropolis method are vague, but might be formulated in this way. Since the transition probability to a new state is always larger for the Metropolis case than for the FM case (for a given underlying symmetric matrix $\| q_{ij} \|$), the former chain tends to move around in the state space more than the latter. This greater mobility of the Metropolis chain should mean that the sampling will span the relevant space more rapidly and thus may lead to more rapid convergence.

An alternative hunch for the desirable characteristic of a transition matrix might be that at each step the chain should as closely as possible choose from the accessible states according to their relative probabilities in the limiting distribution $\{\pi_i\}$. By "accessible," in this context, we refer to the states which can possibly be reached in a single step; this accessibility of states is determined by the q_{ij} matrix, and its choice is determined by various computational considerations which do not concern us here.

These two intuitions turn out to be closely related for most simulations of physical interest, where at each step there are very many accessible states of different energies. This is because both are associated with a probability p_{ii} of repeating the state i which is normally very much larger than the other p_{ij}. That will be true for both the Metropolis and FM schemes, but we have $p_{ii}^M < p_{ii}^F$ (cf. Eqs. (1), (2), (3)). This is associated with a higher mobility in the Metropolis case but also with a relative distribution among the accessible states somewhat less distorted from that of the limiting distribution. These intuitive notions are later examined in the light of our exact results for simple model systems.

Convergence

In this section we discuss the property of the transition matrix which determines the rate of convergence to the limit distribution. Suppose that $\pi_i(t)$ is the probability that the system is in state i after t steps in a realization of the Markov chain. It is clear that

$$\pi_j(t + 1) = \sum_i p_{ij}\pi_i(t) \tag{4}$$

and it is easy to show that the solution of this set of difference equations is

$$\pi_j(t) = \sum_k \alpha_{jk}\lambda_k{}^t \tag{5}$$

where the λ_k are the eigenvalues of $\mathbf{P}$ and α_{jk} are constants related to the left eigenvectors of $\mathbf{P}$ and to the initial conditions. Since $\mathbf{P}$ is stochastic it will have a unit eigenvalue and all other eigenvalues will have moduli less than unity (provided that the chain is irreducible). It is clear that for a long realization the convergence rate is determined by the magnitude of the nonunit eigenvalue whose modulus is closest to unity. We shall call this the "subdominant eigenvalue." This gives a method for the examination of convergence rates for model systems.

Simple Model Cases

(a) *Two State Model*

It is easy to examine this simplest case in detail [6]. If, as is usual, q_{ii} is chosen to be zero, then the FM method leads to more rapid convergence than the Metropolis. Indeed, this is almost obvious, since the FM chain (2) then ensures that the limit distribution is attained in a single step!

More formally we examine [6] the eigenvalues of the corresponding transition matrices. Those for the Metropolis case (taking $\pi_1 \geqslant \pi_2$ and $q_{12} = q_{21} = q$) are $(1, 1 - q/\pi_1)$ and those for the FM case $(1, 1 - q)$. When $q = 1$, which would correspond to the usual choice of $q_{ii} = 0$ for systems of practical interest, the FM case has a subdominant eigenvalue of zero, associated with the instant convergence noted above. (Of course, one could choose values of q such that convergence would be more rapid for the Metropolis case.) It is interesting to note that $\pi_1 = \pi_2 = \frac{1}{2}$ and $q = 1$ leads to a subdominant eigenvalue of -1 for the Metropolis case. This is associated with an undamped oscillation (cf. Eq. (5)) corresponding to the strict alternation between the two states. In this case the chain does not converge, although averages of state functions will do so.

(b) *A Four State Problem: The Two Spin Ising Model*

The Ising problem studied by Cunningham and Meijer consisted of a sequence of two state problems associated with the "flipping" of one spin at a time (with $q_{ii} = 0$). One might conjecture that the above result on the two state system therefore explains

the faster convergence of the FM scheme noted by Cunningham and Meijer. We shall show that the situation is in fact more involved.

In the Cunningham and Meijer realization moves were attempted for the spins in an ordered sequence, rather than by choosing a spin at random. This "cyclic" procedure corresponds to a periodic inhomogeneous Markov chain. However it is not difficult to show (cf. [7]) that the states of the system at the end of each complete cycle form a realization of a Markov chain with a constant transition matrix, providing that the transition matrix for each step of the cycle satisfies (4). We examine the constant transition matrix representing a whole cycle of attempted moves.

In the two spin case there are four possible states which we can represent as

$$\begin{array}{cccc} \uparrow\uparrow & \uparrow\downarrow & \downarrow\uparrow & \downarrow\downarrow \\ 1 & 2 & 3 & 4 \end{array}$$

and there are two steps in the cycle. Now $\pi_1 = \pi_4$ and $\pi_2 = \pi_3$. We set the Boltzmann factor $\pi_2/\pi_1 \equiv \theta \leqslant 1$; this is the "ferromagnetic" case. Of course the conclusions would be the same for the antiferromagnetic choice.

With the Metropolis method the transition matrix describing one cycle of moves is

$$\mathbf{P}^M = \begin{pmatrix} (1-\theta)^2 & \theta(1-\theta) & 0 & \theta \\ 0 & 0 & \theta & 1-\theta \\ 1-\theta & \theta & 0 & 0 \\ \theta & 0 & \theta(1-\theta) & (1-\theta)^2 \end{pmatrix}. \tag{6}$$

It is interesting that although this has the correct limiting distribution, since it satisfies (4), it does not satisfy the condition of microscopic reversibility, which is

$$\pi_i p_{ij} = \pi_j p_{ji}. \tag{7}$$

The eigenvalues of the matrix are

$$(1, \theta^2, \tfrac{1}{2}(1 - 4\theta + \theta^2) \pm \tfrac{1}{2}(1 - 8\theta + 14\theta^2 - 8\theta^3 + \theta^4)^{1/2}).$$

For the case $\theta = 0$ these eigenvalues are (1, 1, 0, 0). The repeated unit eigenvalue corresponds to a reducible chain. The states break down into the two sets (1, 3) and (2, 4), neither of which can be reached from the other. The zero eigenvalues again indicate instant convergence within either set. For $\theta < (3 - 8^{1/2})$ all four eigenvalues are real and positive. For the remaining range of θ there are two conjugate complex eigenvalues, which we may as usual write in the form $\Lambda \exp(\pm i\phi)$, displaying their modulus Λ. These solutions will correspond to oscillatory behavior with a period depending on $\phi(\theta)$. The oscillation will be damped, provided that $\Lambda < 1$, by a factor Λ^t, so that (since Λ is closer to unity than θ^2) Λ is the quantity which should be compared with the subdominant eigenvalue of the FM method.

In the FM case the transition matrix becomes

$$\mathbf{P}^F = \begin{pmatrix} \alpha^2 & \alpha\beta & \beta^2 & \alpha\beta \\ \alpha\beta & \beta^2 & \alpha\beta & \alpha^2 \\ \alpha^2 & \alpha\beta & \beta^2 & \alpha\beta \\ \alpha\beta & \beta^2 & \alpha\beta & \alpha^2 \end{pmatrix} \tag{8}$$

where $\alpha = (1 + \theta)^{-1}$ and $\beta = \theta(1 + \theta)^{-1}$. Once again the matrix does not satisfy the microscopic reversibility condition. The eigenvalues are $(1, (1 - \theta)^2/(1 + \theta)^2, 0, 0)$ and are always real and positive. For $\theta = 0$ the chain is reducible, as for the Metropolis case, and the convergence is again instantaneous within each set of states. If $\theta = 1$, however, the eigenvalues are (1, 0, 0, 0) and the limiting distribution is attained in a single cycle.

It turns out that the relative convergence rates of the two methods now depend on the value of the Boltzmann factor θ. For $\theta > 0.293...$ the FM subdominant eigenvalue is smaller than Λ, while at lower values of θ it is larger than Λ (or, below $(3 - 8^{1/2})$, the Metropolis subdominant eigenvalue). The FM method is therefore preferable for high θ, and indeed becomes very much so as $\theta \to 1$. For low θ, however (where p_{ii} will tend to be high in each case) the Metropolis method will show faster convergence.

There is no *single* quantity corresponding to θ in Cunningham and Meijer's actual many-spin system, of course. An estimate of the corresponding magnitude comes from the Boltzmann factor associated with the interaction energy J between a pair of spins, $\exp(-|J|/kT)$. The authors studied systems in which this quantity varied from 0.37 to 0.91. Of course the value at which the FM method would become preferable is not known for this system, and indeed the situation may not be that simple.

(c) *Another Two Spin Ising Problem*

A simpler but related problem arises if at each step a trial "spin flip" is attempted for a spin chosen at random. The Metropolis and FM transition matrices are then

$$\mathbf{P}^M = \begin{pmatrix} 1 - \theta & \theta/2 & \theta/2 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \theta/2 & \theta/2 & 1 - \theta \end{pmatrix} \tag{9}$$

with eigenvalues $(1, 0, -\theta, 1 - \theta)$, and

$$\mathbf{P}^F = \begin{pmatrix} 1 - \beta & \beta/2 & \beta/2 & 0 \\ \alpha/2 & \beta & 0 & \alpha/2 \\ \alpha/2 & 0 & \beta & \alpha/2 \\ 0 & \beta/2 & \beta/2 & 1 - \beta \end{pmatrix} \tag{10}$$

with eigenvalues $(1, 0, (1 + \theta)^{-1}, \theta(1 + \theta)^{-1})$. It is clear that $\theta(1 + \theta)^{-1} \leqslant (1 + \theta)^{-1} \geqslant 1 - \theta$, so we must compare θ (from the Metropolis case) with $(1 + \theta)^{-1}$ (from the FM case). Once again convergence will be more rapid for the FM case if θ is large, in this case if $\theta > \frac{1}{2}(5^{1/2} - 1) = 0.618$. However, when there is a large energy difference associated with flipping a spin (so that θ is small) the Metropolis method will be superior. Of course the critical value may be expected to depend on the number of spins.

The above examples make it clear that the choice of method will depend on the form of the distribution $\{\pi_i\}$ being sought. In particular, small Boltzmann factors

(i.e., large energy differences) will tend to favor the Metropolis method, in line with intuitive considerations.

In addition, however, we may anticipate a dependence on the number of states in the problem and also on the proportion of these states which are accessible at each step of the chain. These effects can already be seen in the Ising problems (with $q_{ii} = 0$) where for one spin the FM method is always superior while for two spins the choice depends on the value of the Boltzmann factor. In order to examine these effects more clearly we turn to some other simple model problems.

(d) *An n-State Model with Uniform Probability*

A simple model to study is that in which all n states have the same probability and in which any new state can be reached at each step, according to

$$q_{ij} = (n-1)^{-1}, \qquad i \neq j,$$
$$q_{ii} = 0.$$

For the Metropolis case the eigenvalues are unity and $n-1$ repeated values of $-(n-1)^{-1}$ while for the FM case they are unity and $n-1$ repeated values of $\frac{1}{2}(n-2)(n-1)^{-1}$. Thus the FM method converges more rapidly when $n < 4$ and the Metropolis method for $n > 4$. When $n = 4$ all the eigenvalues are identical in the two cases. This confirms the notion that the existence of many states in the problem will tend to favor the Metropolis method.

In this example there was the greatest possible accessibility of states at each step. In typical physical applications only a certain proportion of the system states will be accessible in a single step of the chain. In order to study the effect of this we study systems with differing accessibility.

(e) *Six-State Models with Limited Accessibility*

In these models there is again uniform probability π_i for the states but the number of states accessible at each step has been varied. Thus, in addition to the above model, we have looked at the "ladder" and "next-neighbor" models, with q_{ij} given by the matrices

$$\mathbf{Q}_{\text{ladder}} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix},$$

$$\mathbf{Q}_{n-n} = \begin{pmatrix} 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 \end{pmatrix}.$$

The eigenvalues of the ladder model are $(1, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, -1)$ for the Metropolis case and $(1, \frac{3}{4}, \frac{3}{4}, \frac{1}{4}, \frac{1}{4}, 0)$ for the FM case, so with this very limited accessibility the FM method is indeed superior for $n = 6$. This therefore contrasts with the case of complete accessibility discussed before. The "next-neighbor" model is intermediate with regard to accessibility. In this case the subdominant eigenvalues are equal in magnitude ($-\frac{1}{2}$ for the Metropolis case and $\frac{1}{2}$ for the FM case) so that we would expect similar convergence rates in this case.

It is evident that the number of accessible states at each step is quite critical. This is not unexpected in view of the intuitive arguments mentioned earlier. Of course, as discussed before, a nonuniform limit distribution would also be expected to favor the Metropolis method in most instances.

Discussion

The above simple examples show that no blanket answer is possible as to which method will lead to faster convergence. Certain effects are, however, quite clear. The Metropolis method evidently becomes more favorable (i) as the number of states accessible at each step increases, and (ii) as the probability differences between the states increase, i.e., as the spread of reduced energies increases.

With this in mind it is possible to understand the success noted by Cunningham and Meijer [5] for the FM method, since in that application only two states are accessible at each step of the chain.

In more typical applications the number of accessible states is enormous at each step. For fluids a trial move usually occurs for one particle, into a local volume determined by a "maximum step size": the number of states thus depends on the number of significant figures kept by the computer, and will be very large. The calculations we have carried out therefore clearly support the intuitive arguments suggesting that the Metropolis method will be superior in that case.

It would be satisfying to use our exact results to evaluate those intuitive ideas. This is not too difficult in the case of the suggestion that convergence will be enhanced if at each step the chain assigns the system to new states with relative probabilities according as well as possible with their relative probabilities in the limit distribution. Suppose that m states, $k = 1, 2, \ldots, m$ are accessible from state i in a single step, and define the probability p_j of one of them as

$$p_j = \pi_j \Big/ \sum \pi_k .$$

Then the standard deviation

$$\sigma = \sum_i \pi_i m^{-1} \sum_{j=1}^{m} \{(p_{ij} - p_j)^2\}^{1/2}$$

provides a measure of the closeness of the one-step distribution to the limiting distribution. Some σ-values for both methods for some of our model systems are shown in

TABLE I

Model	$\lvert \lambda_M'/\lambda_F' \rvert$	(σ_M/σ_F)
Two-state, $q_{12} = q_{21} = q$	$\left\lvert \frac{q - \pi_1}{\pi_1(1-q)} \right\rvert$	$\left\lvert \frac{q - \pi_1}{\pi_1(1-q)} \right\rvert$
n-State, uniform limit distribution, $q_{ij} = (n-1)^{-1}$, $i \neq j$	$2/(n-2)$	$2/(n-2)$
Six-state, ladder	4/3	2
Six-state, next neighbor	1	$\frac{2}{3}$
Six-state, all states accessible at each step	$\frac{1}{2}$	$\frac{1}{2}$

Table I and compared with the subdominant eigenvalues λ' and hence the convergence rates: Small values of $\lvert \lambda' \rvert$ and σ should go hand in hand if the suggestion is correct.

This intuition seems rather successful. It can be seen that for the two state case, in fact, σ_M/σ_F exactly equals $\lvert \lambda_M'/\lambda_F' \rvert$, so the criterion even correctly predicts the value of q at which the two methods are equally effective. For the n-state cases where all states are of equal probability and are all accessible at each step, once again the criterion correctly predicts the changeover from FM to Metropolis at $n = 4$. It would also have led one to expect the FM method to be superior for the $n = 6$ "ladder" model, as is the case.

When applied to simulations where many states are accessible at each step, this criterion will favor the Metropolis method. This is because $p_{ii}^F \gg p_{ij}^F$ and the term $(p_{ii} - p_i)^2$ will therefore dominate the expression for the standard deviation. Typically this will be true of the Metropolis chain as well, but since $p_{ii}^F > p_{ii}^M$, then $\sigma_F > \sigma_M$ (for a given q_{ij} matrix) and the criterion favors the Metropolis method.

ACKNOWLEDGMENT

The financial support of the National Research Council of Canada is gratefully acknowledged.

REFERENCES

1. N. METROPOLIS, A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, AND E. TELLER, *J. Chem. Phys.* **21** (1953), 1087.
2. J. M. HAMMERSLEY AND D. C. HANDSCOMB, "Monte Carlo Methods," Methuen, London, 1964.
3. P. A. FLINN AND G. M. MCMANUS, *Phys. Rev.* **124** (1961), 54.
4. P. PESKUN, "The Choice of Transition Matrix in Monte Carlo Sampling Methods Using Markov Chains," thesis, University of Toronto, 1970.
5. G. W. CUNNINGHAM AND P. H. E. MEIJER, *J. Computational Phys.* **20** (1976), 50.
6. J. P. VALLEAU AND S. G. WHITTINGTON, *in* "Statistical Mechanics, Part A: Equilibrium Techniques" (Vol. 5 of Modern Theoretical Chemistry (B. Berne, Ed.)), Chap. 4, Plenum, New York, 1976.
7. W. W. WOOD, *in* "Physics of Simple Liquids" (A. N. W. Temperley, J. S. Rowlinson, and G. S. Rushbrooke, Eds.), Chap. 5, North–Holland, Amsterdam, 1968.