Rejection-free Monte Carlo technique

Rejection-free Monte Carlo technique

Koo-Chul Lee

Department of Physics and the Centre for Theoretical Physics, Seoul National University, Seoul 151-742, Korea

Abstract. An efficient rejection-free Monte Carlo algorithm for lattice systems is presented. In the microcanonical Monte Carlo technique, the sampling time grows prohibitively large due to a high rejection rate at low energies, as the size of the system increases. In this paper we report on an algorithm which improves the sampling efficiency enormously at low energies by selecting only sites which are acceptable for a trial move, thereby reducing the rejection rate to nil.

1. Introduction

Recently there has been an increasing number of studies which use the microcanonical Monte Carlo technique (MCMCT) [1–4]. The advantage of the MCMCT is that we can calculate whole (continuous) thermodynamic functions instead of thermodynamic points with much less effort than that needed to calculate a single thermodynamic point in the conventional Monte Carlo technique [3]. As we pointed out in [3], one of the difficulties in the MCMCT is low-temperature (microcanonical) slowing down and not critical slowing down, and we suggested a remedy to the problem.

Recently Care [4] used the MCMCT to calculate the density of states of the Blume–Capel model. In order to overcome low-temperature slowing down he simply reduced the width of the energy band of the microcanonical ensemble. This method is useful in some cases especially for continuous systems (in contrast to the lattice systems). However, it necessarily incurs repeated sampling due to rejected moves which would inevitably lower the precision of the Monte Carlo (MC) data.

In this paper we present a method which improves the sampling efficiency enormously at low microcanonical temperatures, where the density of states has a large slope, without incurring repeated sampling. In this method we bypass all the trials, which would be rejected if we selected random sites for trial moves, by maintaining a table from which acceptable sites for trial moves can be read off. A similar technique has been used in the conventional MC algorithm [5–7]. However, as we will show later, this technique becomes best suited for the MCMCT. We begin by reviewing the standard MCMCT in the next section and examine the origin of the low-energy slowing down. In section 3 we describe the rejection-free technique and present a test result, and in the final section we present a summary together with a few remarks concerning the new technique.

2. Standard procedure and origin of the low-energy slowing down

In this section we will briefly review the standard MCMCT and discuss the origin of the low-energy (or equivalently low-temperature) slowing down. We will take a spin-$\frac{1}{2}$ Ising

model of N spins in the absence of an external field as an example [3]. The energy of the system can be written as

$$E(\{S_i\}) = -J \sum_{\langle i,j \rangle} S_i S_j \tag{1}$$

where S_i is the spin variable assuming ± 1 values, J is the exchange energy and $\langle i, j \rangle$ runs over interacting nearest-neighbour pairs i, j. The canonical average $\langle A \rangle$ of any thermodynamic quantity $A(\{S_i\})$ is defined by

$$\langle A \rangle = \sum_{\{S_i\}}^{2^N} A(\{S_i\}) \exp(-\beta E(\{S_i\}))/Q \tag{2}$$

where β is the inverse temperature $1/k_B T$ with Boltzmann's constant k_B, and Q is the partition function defined by

$$Q = \sum_{\{S_i\}}^{2^N} \exp(-\beta E(\{S_i\})). \tag{3}$$

We can rewrite equations (2) and (3) in slightly different forms as

$$\langle A \rangle = \sum_E \Omega(E) \exp(-\beta E) \overline{A}(E)/Q \tag{4}$$

and

$$Q = \sum_E \Omega(E) \exp(-\beta E) \tag{5}$$

where $\overline{A}(E)$ is the microcanonical average of the variable A defined by

$$\overline{A}(E) = \sum_{\{S_i\}}^{\Omega(E)} {}' A(\{S_i\})/\Omega(E) \tag{6}$$

and $\Omega(E)$ is the density of state (DOS) at E. The prime in equation (6) indicates that the summation is over microscopic configurations with a fixed E. From this point of view, the calculation of $\langle A \rangle$ is reduced to the calculations of $\Omega(E)$ and the microcanonical average $\overline{A}(E)$.

In the MCMCT, we set up a random walk in the configuration space restricted to a narrow energy band which consists of few energy layers. We select a single spin out of N spins either randomly or sequentially and attempt to flip it. Whenever the attempted move takes the walker to a spin configuration $\{S_i\}$ which lies within the energy band the move is allowed; otherwise the move is rejected. $\overline{A}(E)$ are calculated simply by sampling As whenever the walker visits the layer at a given fixed E. That is $\overline{A}(E) = \sum_{k=1}^{N_d} A_k/N_d$, where A_ks are samples of the dynamical variable A and the N_d is the number of data points. $\Omega(E)$s are estimated consecutively by the ratio N_d^+/N_d to $\Omega(E + \Delta E) = \Omega(E) N_d^+/N_d$. In the above N_d^+ and N_d are the number of visits at energy layer E and $E + \Delta E$, two consecutive energies in the band.

After a sufficient number of data points are collected we move to the next energy band whose lowest energy layer overlaps the highest energy layer of the band we just left. By this method we can calculate $\overline{A}(E)$s and $\Omega(E)$s for all energies E. For $\Omega(E)$s one needs to know $\Omega(E_0)$ of the lowest energy layer of the lowest band. However, the unknown factor is irrelevant in the calculation of $\langle A \rangle$ since the factor cancels out in the average given by [4]. It is only relevant to the calculation of the absolute entropy or the free energy.

In the early version of the MCMCT, the energy band consisted of a few energy layers, for example, four in [2]. The samples are taken at a large interval in order to avoid repeated sampling of the same configuration. The problem of repeated sampling is especially severe where the rejection rate is high. In the refined MCMCT reported in [3] there is a new device to avoid this very problem.

In order to illustrate the refined MCMCT, which we will call the standard MCMCT hereafter, we will rewrite the energy of system (1) as

$$E = -4J(N^{++} - \tfrac{1}{2}qN^+) \tag{7}$$

where the constant term $-\tfrac{1}{2}JqN$ has been dropped. In the above, q, N^+ and N^{++} are the coordination number, the total number of up-spins, and the total number of interacting up-spin pairs respectively. In this description it is easy to recognize that the energy of the system is separated by $\Delta E = 4J$. The standard MCMCT is based on the fact that there is a number of classes of sites, classifying sites by the change of energy $\delta E(= -\Delta E(\delta N^{++} - \tfrac{1}{2}q\delta N^+))$. In a lattice with periodic boundary conditions the number of classes is $q+1$. For the square lattice there are five classes of sites, and we can number them as shown in table 1.

Table 1. Classifications of spins. The second column is the state of spin to be flipped. The third column is the number of nearest-neighbour spins that are up-state. The last three columns represent the change in the total number of up-spins, the change in the total number of up-spin pairs and the change in the total energy, $E_{new} - E_{old}$, respectively.

Class	Spin	Number of spin-up nearest neighbours	δN^+	δN^{++}	$\delta E = -\Delta E(\delta N^{++} - \tfrac{1}{2}q\delta N^+)$
1	Up	4	−1	−4	2
2	Up	3	−1	−3	1
3	Up	2	−1	−2	0
4	Up	1	−1	−1	−1
5	Up	0	−1	0	−2
5	Down	4	1	4	−2
4	Down	3	1	3	−1
3	Down	2	1	2	0
2	Down	1	1	1	1
1	Down	0	1	0	2

From table 1 we can easily see that the spin-flip trial is always successful at the middle layer if we make the energy band consist of at least $q+1$ (which equals five for the square lattice) layers. This is because any site belongs to one of the $q+1$ classes and the spin flip only causes the total energy change bounded by the two units above and below. If the current position of the random walker is in the middle layer the spin-flip move always leaves the walker within the energy band. If we take samples at the middle layer, the samples are always from fresh spin configurations. In order to calculate $\Omega(E)$s, we have to take samples of two consecutive energy layers. For this reason we construct an energy band consisting of $q+2$ energy layers and take samples at the middle two layers. By this elaboration only can one attain the high efficiency and precision reported in [3].

In summary, in the standard MCMCT we set up a random walk in the configuration space restricted to a narrow energy band given by

$$E_j - \tfrac{1}{2}q\Delta E \leqslant E(\{S_i\}) \leqslant E_j + (\tfrac{1}{2}q + 1)\Delta E. \tag{8}$$

For data collection we use configurations which satisfy

$$E(\{S_i\}) = E_j \qquad \text{and} \qquad E(\{S_i\}) = E_j + \Delta E. \tag{9}$$

In a sense, the standard MCMCT is already a rejection-free technique, at least at the energy taking middle two layers. This is accomplished by introducing padded layers above and below the data collecting layers. The sole purpose of creating the rejection-free layers in the standard MCMCT is to guarantee that the samples taken are from fresh configurations. However, in order to apply the rejection-free (RF) MCMCT we are proposing in this paper that the above prescription is imperative, as we will see in the next section.

At low energies, the density of states (DOS) grows rapidly as the energy increases or the microcanonical temperature (the inverse of the rate of the increment of the logarithm of DOS) is very low. Therefore, at low microcanonical temperatures, any unbiased random walk would be likely to put the system at the highest energy layer and any attempted spin flip would most likely make the system move out of the upper boundary of the energy band. We can understand this situation more vividly if we look at the configurational space.

At low energies, the majority of spins are aligned in one direction, say the up-direction and only a few spins are excited, that is in the down-direction. If we choose a spin site randomly, we would most likely select one of the aligned spins which belong to class one of the first row of table 1. The spin-flip move will most likely move the walker to a configuration belonging to the layer two units above. The move in turn rapidly puts the walker at the top-most energy layer in the band. At this point another randomly selected spin site would again be likely to be one of the aligned spins belonging to class one and the trial move on this spin would be likely to be rejected, since it put the walker out of the band. There would only be a small fraction of the number of spin sites near or at excited spins for which the trial move would be successful at any time. This fraction gets smaller as the size of the system grows larger. In general, this is the source of slowing down of the MC sampling at low energies.

3. Rejection-free technique and a test result

In order to overcome the difficulty described in the previous section, we introduce an RF technique. We only set up a random walk in the configuration space exactly in the same way as described in the previous section. The crux of the new technique is to maintain a pair of tables: one to look up which sites are successful for a trial move and the other being a cross reference table to look where the specific sites are located in the look-up table and to update the tables with minimum effort. We only select for a spin flip out of acceptable sites thereby eliminating any rejectable trial move.

Initially we can scan the whole system by going through every lattice site examining the spin states of the site and its interacting neighbours. We then classify the sites into the $q + 1$ classes as in table 1 and put them into one of the sites of the look-up tables. From the table we can easily tell which sites are acceptable or rejectable. For example, for the square lattice if the walker is in the configuration space belonging to the top-most layer in the band, the sites belonging to classes three, four and five are acceptable since the move will keep the walker within the energy band.

As long as we take data from the middle two layers, the path of the random walk and the MC data are exactly the same as those of the standard MCMCT of the previous section. Suppose at a certain point of the random walk there are N_a acceptable sites. In the standard MCMCT we select a spin out of N with a uniform probability, and whenever we select a rejectable site we pass and select another spin until we hit one of the acceptable sites. Only then we do move on to a new spin configuration. In the RF MCMCT we select a spin out of the N_a acceptable sites with a uniform probability, eliminating all the wasted trial moves of selecting rejectable sites. In either procedure, since the probability of moving to a new

specific spin configuration is the same, the path of the random walk is exactly the same.

Now if we were taking data from every layer of the band, then we need to know how many times the trial move would have been wasted in the standard procedure. This is because this number must be multiplied as a weight in the calculation of the average $\overline{A}$ in the RF procedure. This information is not available in the RF procedure. However, if we take data from the middle two layers only, we know this weight to be unity since there is no rejectable site in these two energy layers! This is the reason why we should keep the energy band given by (8) and take data from configurations which satisfy equation (9).

Once a new configuration is generated, we have to update the look-up table. Since the new configuration is modified by a single spin flip, the change of the configuration is limited to a change in a small number of lattice sites. Therefore, it is possible to update the look-up table by a procedure consisting of a finite number of steps. In the appendix we present one such algorithm where the updating procedure consists of a maximum number of $2q$ steps.

Since the computing time for the updating procedure in the RF algorithm depends only on the coordination number and not the system size, while the rejection rate in the standard procedure at a low energy of fixed $N_e (\equiv E/\Delta E)$ grows as the system size grows, we will know at a certain point where the efficiency of the RF technique wins over as the size of the system grows.

In general, the standard procedure is sufficient for a system of moderate size with low coordination number. On the other hand, when the coordination number or the number of spin states increases, the RF technique becomes far more efficient for systems of even smaller size at low microcanonical temperatures. However, when the microcanonical temperature is reasonably high, the efficiency of the standard procedure wins over due to the time consuming updating procedure of the double table in the RF procedure.

The RF MC method has been successfully used to calculate the universal scaling functions for the three-state Potts model on a square lattice of the size up to 33×33 [8], where the width of the energy band (8) necessary to avoid repeated sampling at the data collecting layers (9) is the same as that of the Blume–Capel model. We have also used the RF technique to calculate the complex scaling partition function in the complex temperature plane for the $d = 3$ Ising model on a simple cubic lattice of sizes up to $24 \times 24 \times 24$ [9].

In the following tables we present a test result which compares the efficiency of two algorithms at various energies (microcanonical temperatures) for two different system sizes using the spin-$\frac{1}{2}$ Ising model on a simple cubic lattices with periodic boundary conditions. For test purposes, we took only a small number of data points to save the computing time. In real simulations, 120 to 1200 times more data are collected [9].

The first column in tables 2 and 3 is the number of data points, N_{data}, the number of visits at the lower layer of (9) for each N_e. In the second column we placed N_e values together with the microcanonical temperatures defined by $T_{\text{mic}} = ((\ln \Omega(N_e + 1) - \ln \Omega(N_e - 1))/8)^{-1}$ in the bracket. The exact $\ln(\Omega(N_e))$ are calculable from the low-temperature series expansion coefficient [10] up to $N_e = 18$, and are placed in the third column. Up to this value, we calculated percentage relative deviations of the MC calculations from the known exact values for $\ln(\Omega(N_e))$ in the bracket in the fourth and sixth columns while for the N_e values larger than 18, the percentage relative deviation in the results obtained by two procedures are calculated and placed in the third column. The $\Omega(N_e)$s are not normalized for $N_e > 18$. In the fifth and seventh column we put the computing time for the two procedures, and the last column is the ratio of these two numbers which is a relative measure of the efficiency of the two algorithms. They are real computing times (in s) on a PC(i80486).

It is noteworthy that when the microcanonical temperature is about 3.25 the two

Table 2. The comparison table of efficiency of two algorithms for $N = 10^3$.

$N_{data}10^{-3}$	$N_e(T_{mic})$	$\ln(\Omega(N_e)_{exact})$	$\ln(\Omega(N_e)_{rf})$	t_{rf}	$\ln(\Omega(N_e)_{std})$	t_{std}	t_{std}/t_{rf}
20	15(3.55)	30.3852	30.4076(0.074%)	13	30.3802(−0.016%)	67	5.2
20	16(2.92)	29.8183	29.7911(−0.091%)	243	29.8346(0.055%)	1146	4.7
20	17(1.42)	33.0714	33.0066(−0.196%)	82	32.9876(−0.253%)	354	4.3
20	18(3.38)	35.4845	35.5118(0.077%)	17	35.5541(0.196%)	72	4.2
20	101(3.25)	(0.028%)	31.2900	47	31.2989	46	1.0
20	102(3.26)	(−0.030%)	32.5340	47	32.5243	47	1.0
20	103(3.23)	(−0.024%)	33.7633	47	33.7553	47	1.0
20	748(410.26)	(−0.175%)	27.9001	19	27.8514	5	0.3
20	749(352.33)	(−0.141%)	27.8983	19	27.8589	6	0.3
20	750(690.99)	(−0.092%)	27.8998	19	27.8741	5	0.3

Table 3. The comparison table of efficiency of two algorithms for $N = 20^3$.

$N_{data}10^{-3}$	$N_e(T_{mic})$	$\ln(\Omega(N_e)_{exact})$	$\ln(\Omega(N_e)_{rf})$	t_{rf}	$\ln(\Omega(N_e)_{std})$	t_{std}	t_{std}/t_{rf}
2	15(3.41)	40.8331	40.8389(0.014%)	1	40.8240(−0.022%)	30	30.0
2	16(2.85)	38.1920	38.2653(0.192%)	197	38.2856(0.245%)	6820	34.6
2	17(0.82)	43.5390	43.2567(−0.648%)	54	43.3550(−0.423%)	1931	35.8
2	18(4.08)	48.0244	48.0450(0.043%)	0	47.7215(−0.631%)	24	30.0
20	842(3.27)	(−0.043%)	38.4099	49	38.3936	47	1
20	843(3.24)	(0.041%)	39.6373	49	39.6534	48	1
20	844(3.25)	(0.044%)	40.8744	48	40.8924	47	1
20	5998(247.79)	(0.180%)	35.9485	20	36.0133	6	0.3
20	5999(416.29)	(0.216%)	35.9591	19	36.0366	6	0.3
20	6000(−561.02)	(0.246%)	35.9441	20	36.0325	6	0.3

computing times become equal to each other in both system sizes. In fact it is true for all other system sizes. We remark in passing that the microcanonical temperatures at low energies are irregular and sometimes become higher than 3.25, as shown in tables 2 and 3. The low-energy slowing down in this case is due to the high rejection rate of the padded layers above the middle layers where the microcanonical temperatures are still low. When the system reaches the highest microcanonical temperature at the maximum of N_e, $qN/8$, the standard algorithm becomes about three times faster than the RF algorithm for both sizes.

4. Conclusion and discussion

There are several advantages of the double look-up table technique in the MCMC method over the conventional MC method [5–7]. In the MCMCT, acceptable sites and rejectable sites are clearly divided while in the conventional MC technique it is only probabilistically classified and has probability that needs to be calculated. The best advantage of all is that in the MCMCT there is a sharp division by energy where the efficiency of one procedure wins over the other. Therefore, we can switch from one procedure to the other to maximize the efficiency in updating the ensemble. On the other hand, in the canonical MC setting even at moderate temperatures, the ensemble encompasses wide energy layers (see e.g. [3]) and

there is no way to apply separate updating procedures to maximize the efficiency.

We should remark again that the RF technique in the MCMC method does not alter the dynamical nature of the random walk. In the RF technique we merely bypass all the wasteful trials of the standard procedure of selecting rejectable sites, leaving no effect at all on the spin configuration. Therefore, all the properties of the dynamics of the random walk in the configurational space in the MCMCT remain intact.

We should also remark that it does not matter in what sequence the acceptable sites are placed in the look-up table if the random number we use is truly random (uncorrelated to any of the previously generated random numbers). In the MCMC method the random number generator is least demanding [1,3] to the point that thermodynamic functions for the $d = 2$ Ising model have been calculated without using a random number generator at all [11]. It is this property that makes the simple RF updating procedure described in this paper work very well, as we saw in the test result as well as in [8,9], even if we use random numbers of moderately good quality.

Finally, we emphasize again that in the RF algorithm the width of the microcanonical ensemble should not be made less than (8) although it may be made larger than (8), since there is no way of knowing how often the trial move rejected at the data collecting layers (9), had we used the standard algorithm with a narrower band.

In summary, the RF algorithm in the MCMCT, which bypasses all the wasted trial moves, can be enormously efficient at low microcanonical temperatures for systems which have large coordination number or large number of spin states. In fact, in a calculation where the absolute (normalized) DOS needs to be calculated, such as in [8,9], the RF algorithm is an indispensable technique even for a moderate size system.

Acknowledgments

Appendix. A method of updating the double look-up tables

We first label N spins by numbers running $1, \ldots, N$. Next we construct a double table in the form of an array, *order*[ ] and *site*[ ] of N elements. We initialize this table by putting numbers in the array. Any spin site at any moment belongs to one of $q + 1$ classes. We arrange them in sequence and put them in *order*[ ]. The spins belonging to the same class may be placed initially in any order. There are movable partitions which separate the classes. The movable partitions may be constructed using an array *boundary*[ ] of size $q + 1$ indexed from $-q/2$ to $q/2$. For convenience we use δE for the index k of the array *boundary* representing the class. The array *boundary*[k] contains the smallest of the indices of spins belonging to the kth class in *order*[ ].

Now we construct another table, *site*[ ], which is an array that contains indices of the array, *order*[ ], in positions indexed by a number which is the label of the lattice site. They are cross reference tables so that $i = site[order[i]]$ and $i = order[[site[i]]$.

Let us suppose that we are taking data in a microcanonical ensemble labelled by the integer N_e measured in the unit of $\Delta E \equiv 4J$. The energy layers in the band are labelled by

an integer p which takes values from $[-q/2, q/2 + 1]$ so that the microcanonical ensemble consists of $q + 2$ energy layers specified by $(N_e + p)$s. Suppose that our system is at the pth layer. Then we should select a spin site from the classes in $order[\ \ ]$ whose k value lies in

$$[k_{min}, k_{max}] \tag{10}$$

where $k_{min} = \max\{-q/2, p - q/2\}$ and $k_{max} = \min\{q/2, q/2 + 1 - p\}$. If we generate a random number i in the interval $[boundary[k_{min}], boundary[k_{max} + 1] - 1]$ (or $[boundary[k_{min}], N]$ if $k_{max} \leqslant q/2$), then we have a spin site, $j = order[i]$, which is one of the acceptable sites.

Now consider the updating part of the double array. If we flip the jth spin, the jth site changes the class it belongs to and rearrangement of the $order[\ \ ]$ is necessary. If the site belonged to the kth class before flip, it will move to the $-k$th class after the spin flip. This can be achieved by $2k$ numbers of exchange moves in the $order[\ \ ]$. Suppose k was a negative number initially, we can put the element in the $-k$th class by the following steps. We first exchange this element with the element in the highest position in the same class. Next, we shift the boundary one downward by decreasing the number in the $boundary[k]$ by one. Next, we exchange this element with the element of highest position in the new class and shift the boundary again by subtracting one from $boundary[k + 1]$. This exchange move repeats until the element is placed in the lowest position of the class of the final destination. This move must accompany the similar exchange move in the array $site[\ \ ]$. The maximum number of exchange moves is q, which is finite.

In addition to this updating of the very spin that flipped, we have to update the q interacting neighbours of the central spin that flipped. The position of neighbouring spins in the $order[\ \ ]$ can be found by reading values $site[j \pm 1]$, etc. To q neighbouring spins, since only one of the surrounding spins has changed its state, they move to one of the adjacent classes in $order[\ \ ]$. Therefore, we only need q number of exchange moves of the double array for the neighbouring spins. Altogether, we need to perform a maximum of $2q$ numbers of exchange moves of the double array accompanying the same number of updating steps for the partition.

References

[1] Creutz M 1983 *Phys. Rev. Lett.* **50** 1411
[2] Bhanot G, Salvador R, Black S, Carter P and Toral R 1987 *Phys. Rev. Lett.* **59** 803
[3] Lee K-C 1990 *J. Phys. A: Math. Gen.* **23** 2087
[4] Care C M 1993 *J. Phys. A: Math. Gen.* **26** 1481
[5] Boritz A B, Kalos M H, and Lebowitz J L 1975 *J. Comput. Phys.* **17** 10
[6] Novotny M A 1995 *Phys. Rev. Lett.* **74** 1
[7] Novotny M A 1995 *Comput. Phys.* **9** 46
[8] Lee K-C 1992 *Phys. Rev. Lett.* **69** 9
[9] Lee K-C 1993 *Phys. Rev.* E **48** 3459
[10] Domb C and Green M S 1974 *Phase Transitions and Critical Phenomena* vol 3 (New York: Academic) p 416
[11] Lee K-C *J. Kor. Phys. Soc.* **24** S32