Available online at www.sciencedirect.com
SCIENCE DIRECT
Computer Physics Communications 175 (2006) 36–40
Computer Physics Communications
www.elsevier.com/locate/cpc

Convergence and refinement of the Wang–Landau algorithm

Hwee Kuan Lee [a,b,*], Yutaka Okabe [a], D.P. Landau [b]

[a] *Department of Physics, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan*
[b] *Center for Simulational Physics, University of Georgia, Athens, GA 30602, USA*

Received 20 January 2005; received in revised form 13 May 2005; accepted 14 February 2006

Available online 2 May 2006

Abstract

Recently, Wang and Landau proposed a new random walk algorithm that can be very efficiently applied to many problems. Subsequently, there has been numerous studies on the algorithm itself and many proposals for improvements were put forward. However, fundamental questions such as what determines the rate of convergence has not been answered. To understand the mechanism behind the Wang–Landau method, we did an error analysis and found that a steady state is reached where the fluctuations in the accumulated energy histogram saturate at values proportional to $[\log(f)]^{-1/2}$. This value is closely related to the error corrections to the Wang–Landau method. We also study the rate of convergence using different "tuning" parameters in the algorithm.

PACS: 05.10.Ln; 65.40.Gr; 02.50.-r

Keywords: Monte Carlo; Density of states; Random walk

1. Introduction

Computational methods have been used extensively for solving complex problems in the past few decades. In particular, in statistical physics equilibrium quantities of a system with many degrees of freedom are measured. The framework of statistical physics is formalized such that all equilibrium quantities can be derived from the partition function,

$$\mathcal{Z}(T) = \sum_{\{\sigma\}} \mathrm{e}^{-E(\sigma)/k_{\mathrm{B}}T}, \tag{1}$$

σ is the state of the system, E is the energy corresponding to σ, k_{B} is the Boltzmann constant and T is the temperature. The summation is over all possible states and the number of possible states is a colossal number which cannot be enumerated. Nevertheless, computational methods such as Monte Carlo techniques [1] are used to sample the partition function; in particular, Metropolis importance sampling [2] has achieved considerable success. However, new techniques are emerging and are replacing the Metropolis importance sampling, especially near phase transition boundaries where the Metropolis importance sampling becomes inefficient. A class of new techniques, called generalized ensemble methods, such as the multicanonical method [3,4], the broad histogram method [5] and the flat histogram method [6,7], were developed based on re-writing the partition function as a sum over energies

$$\mathcal{Z}(T) = \sum_{\{\sigma\}} \mathrm{e}^{-E(\sigma)/k_{\mathrm{B}}T} = \sum_{E} g(E)[\mathrm{e}^{-E(\sigma)/k_{\mathrm{B}}T}], \tag{2}$$

where the partition function is reduced from a sum over all states to a sum over $\sim N$ energy levels. The partition function would be tractable if the energy density of states $g(E)$ could be calculated.

Recently, a systematic, iterative, random walk method [8–10] has been proposed as one of the generalized ensemble methods. Now generally known as the Wang–Landau algorithm, it has received much attention in literature and has been applied to a wide range of problems [11–14]. There have also been numerous proposed improvements and studies of the efficiency of this method [15–23]; however, there are still many unanswered questions, e.g., what determines the rate of convergence to the true density of states and is there any universality behavior related to this algorithm? In this paper, we attempt to quantify

* Corresponding author.
E-mail address: hweekuan@yahoo.com (H.K. Lee).

0010-4655/$ – see front matter
doi:10.1016/j.cpc.2006.02.009

the mechanism behind the Wang–Landau method and study the effects of using different "tuning" parameters.

The Wang–Landau algorithm [8–10] is an iterative process in which the density of states $g(E)$ is modified by a factor $f_k > 1$, and the refinement of the density of states is assured with monotonically decreasing modification factors, e.g., $f_{k+1} = \sqrt{f_k}$. For each time the energy level E is visited, $g(E)$ is multiplied by $f_k > 1$, and a histogram of energy is accumulated concurrently. It was proposed that the modification factor be decreased when the accumulated histogram satisfies a certain flatness condition which we call the stopping condition. In this paper we study the effects of different modification factors and stopping conditions, and derive an expression for the error term in the Wang–Landau method based on generalizations of the modification factors and stopping conditions. We shall consider arbitrary sequences of modification factors, $f_1 > \cdots f_k > f_{k+1} \cdots > 1$ and arbitrary sequences of corresponding stopping conditions $\lambda_1, \lambda_2, \ldots$. The stopping conditions $\lambda_1, \lambda_2, \ldots$ may or may not be the histogram flatness condition; other stopping conditions could be used, for example, stopping after some predetermined maximum number of Monte Carlo steps, or stopping after some number of times the random walker reaches the minimum energy state. This generalization is needed for theoretical error analysis in the next section.

2. Theoretical error analysis

In the Wang–Landau algorithm, for each visit to an energy level E, the density of states at that energy level increases by a factor $f_k > 1$, and the corresponding histogram increases by one. Assume that initially the unknown density of states were set to 1, i.e. $g_0(E) = 1\ \forall E$. The density of states at the end of the nth iteration is given by

$$\log g_n(E) = \sum_{k=1}^{n} H_k(E) \log(f_k), \tag{3}$$

where $H_k(E)$ is the accumulated histogram and f_k is the modification factor for the kth iteration. At this point, it is important to realize that the relative values of $g(E)$ are sufficient for calculating thermodynamics quantities. Hence, a constant factor can be extracted from $g_n(E)$, without losing any information, by a change of variable on the histograms,

$$H_k(E) \to H_k(E) - \min_E\{H_k(E)\} = \tilde{H}_k(E) \quad \text{for } k = 1, 2, \ldots, n, \tag{4}$$

where $\min_E\{H_k(E)\}$ is the minimum value of the accumulated histogram for the kth iteration. Then Eq. (3) becomes

$$\log g_n(E) = \sum_{k=1}^{n} \tilde{H}_k(E) \log(f_k) + \text{constant of energy}. \tag{5}$$

The second term in Eq. (5) is independent of energy, and from here onwards we shall refer to $\tilde{g}_n(E)$ as the density of states without the second term in Eq. (5), i.e.

$$\log \tilde{g}_n(E) = \sum_{k=1}^{n} \tilde{H}_k(E) \log(f_k). \tag{6}$$

To lay the foundation for deriving an expression for the error of the Wang–Landau method, we use the conjecture that the method converges to the true density of states with proper choice of parameters.

Conjecture 1. *Let the Wang–Landau algorithm be carried out with a sequence of modification factors, $\cdots f_k > f_{k+1} > \cdots f_\infty = 1$. There exists a sequence of stopping conditions $\lambda_1, \lambda_2, \ldots, \lambda_\infty$ such that,*

$$\lim_{n\to\infty} \tilde{g}_n(E) = \tilde{g}_\infty(E) = g^*(E) \times \text{constant}, \tag{7}$$

where $\tilde{g}_n(E)$ is the density of states calculated up to the nth iteration and $g^*(E)$ is the true density of states.

This conjecture does not give any error bounds on the density of states; it only says that, in the limit of an infinite number of iterations the Wang–Landau estimate converges to the true density of states. In addition, no constraint is imposed on the stopping conditions in the conjecture. The error term up to the nth iteration can be defined as

$$\sum_E \big[\log \tilde{g}_\infty(E) - \log \tilde{g}_n(E)\big] = \sum_E \sum_{k=n+1}^{\infty} \tilde{H}_k(E) \log(f_k). \tag{8}$$

An intuitive view of Eq. (8) is that, if an infinite number of iterations were performed, the exact answer would be obtained. When n iterations were done instead, the error of $\tilde{g}_n(E)$ will be the sum of all the rest of the iterations that were not carried out explicitly. Define

$$\Delta H_k = \sum_E \tilde{H}_k(E), \tag{9}$$

where ΔH_k is a measure of fluctuations in $\tilde{H}_k(E)$. By the assumption of convergence series (implied by Conjecture 1), the order of summations in Eq. (8) can be rearranged. Then, Eq. (8) becomes

$$\eta_n = \sum_E \big[\log \tilde{g}_\infty(E) - \log \tilde{g}_n(E)\big] = \sum_{k=n+1}^{\infty} \Delta H_k \log(f_k). \tag{10}$$

Eq. (10) shows that, assuming appropriate stopping conditions, η_n depends only on the fluctuation in the histogram and the sequence of modification factors f_k. When the values of f_k are predetermined (e.g., $f_{k+1} = \sqrt{f_k}$), ΔH_k becomes the only determining factor for η_n.

3. Results

We investigate the Monte Carlo time dependence of ΔH_k for each iteration with the Wang–Landau method, where the subscript k denotes the kth iteration. Simulations were performed on the ferromagnetic Ising model and on the fully frustrated Ising model with various system sizes. The Hamiltonian is

$$\mathcal{H} = -\sum_{\langle ij \rangle} J_{ij} \sigma_i \sigma_j, \tag{11}$$

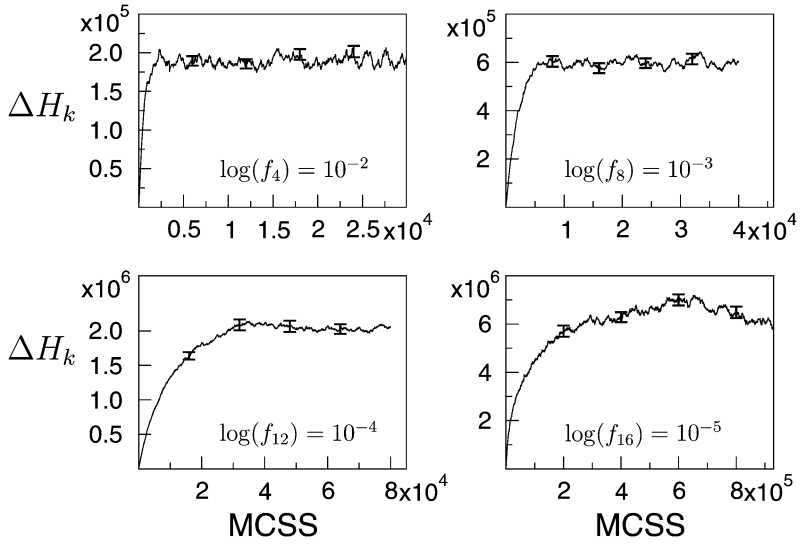

Fig. 1. ΔH_k versus Monte Carlo steps per site for 16×16 ferromagnetic Ising model for various $\log(f)$ values. From left to right, top to bottom, $\log(f)$ values are 10^{-2}, 10^{-3}, 10^{-4} and 10^{-5}.

where the sum is over nearest neighbors on a two-dimensional square grid and σ_i takes the values ± 1. $J_{ij} = 1$ for the ferromagnetic Ising model, and for the fully frustrated Ising model, J_{ij} takes the value -1 for every alternate horizontal nearest neighbors bonds and $+1$ otherwise.

Fig. 1 shows the time dependence of ΔH_k for several values of $\log(f)$; $\log(f) = 10^{-2}, 10^{-3}, 10^{-4}$ and 10^{-5} from left to right, top to bottom, respectively. We used the sequence of correction factors $\log(f_{k+1}) = \log(f_k)/1.78$ with $\log(f_1) = 0.1$, this sequence is chosen so that $\log(f_{k+4}) = \log(f_k)/10$. These graphs were generated by performing the Wang–Landau algorithm on a 16×16 ferromagnetic Ising model with numerical values averaged over 128 independent simulations. The Monte Carlo steps per spin, the horizontal axis of Fig. 1, are measured from the time when we decrease $\log(f)$. ΔH_k increases initially and eventually saturates, and for smaller $\log(f)$ values, saturation values are greater and number of Monte Carlo steps required to reach saturation are larger. Because the error term given by Eq. (10) depends only on ΔH_k, any computational effort after ΔH_k become saturated does not improve the accuracy of the final density of states $g_n(E)$. On the other hand, stopping the random walk before ΔH_k becomes saturated would make the simulation less efficient because insufficient statistics are accumulated in the kth iteration and much more statistics would have to be accumulated with smaller $\log(f)$ values for subsequent iterations. An optimal algorithm is to stop the simulation as soon as ΔH_k becomes saturated. The Wang–Landau algorithm in the original paper [8] suggested using the histogram flatness condition as a stopping condition, but this does not guarantee optimal efficiency.

It is difficult to predict the saturation value of ΔH_k for $k = 1, \ldots$. As shown in Fig. 1, we performed a set of simulations with more Monte Carlo steps than required for ΔH_k to reach saturation. In this way, we could measure the saturation values accurately. Fig. 2 shows a plot of saturation values versus $\log(f)$ for ferromagnetic Ising model (FMIM) and fully frustrated Ising model (FFIM). In double log scale, the data points

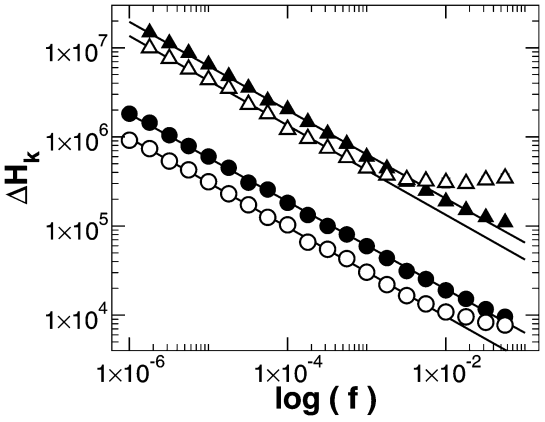

Fig. 2. Plots of saturation ΔH_k versus $\log(f)$ for 8×8 ferromagnetic Ising model (filled circles), 8×8 fully frustrated Ising model (empty circles), 16×16 ferromagnetic Ising model (filled triangles) and 16×16 fully frustrated Ising model (empty triangles). 128 independent simulations were performed for each data point and error bars were smaller than the size of the symbols.

fall on a straight line with the values of the slopes equal to -0.491 ± 0.004 for 8×8 FMIM, -0.501 ± 0.004 for 8×8 FFIM, -0.496 ± 0.006 for 16×16 FMIM and -0.502 ± 0.008 for 16×16 FFIM. To within error bars, the slopes seem to have an universal behavior

$$\max\{\Delta H_k\} \propto \log(f_k)^{-1/2} \tag{12}$$

as predicted by Zhou and Bhatt [19]. Our results suggest that the values of the slope is generic to the Wang–Landau algorithm and does not depend on system sizes and models. Certainly many more simulations on different models are needed to confirm the universality of the slope.

4. Effects of modification factors

We also looked at how the Wang–Landau algorithm performs with different sequences of modification factors. In the extreme case, we studied the effects of taking the limit of $f = 1$

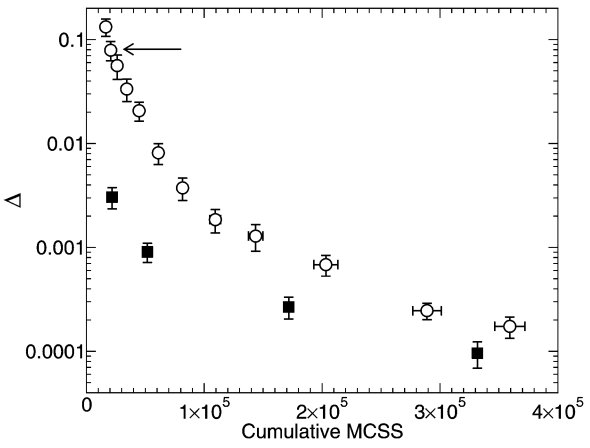

Fig. 3. Comparison of accuracy of the Wang–Landau method with two modification sequences. Data points were generated with the sequence $f_{k+1} = \sqrt{f_k}$ (empty circles) and with the sequence where the limiting value of $f = 1$ is used after 14 iterations (filled squares). Lattice size is 32×32 with energy range $E/N_E \in [-1.55, -1.35]$.

only after a few iterations. Assuming n iterations were performed with large modification factors, and on the $(n+1)$th iteration, the modification factor is set to 1. The background for implementation is as follows: Eq. (2) uses the Boltzmann weights ($B(E,T) = \exp(-E/k_B T)$) and the resulting energy distribution is,

$$P(E) = g^*(E)B(E,T)/\mathcal{Z} = g^*(E)\exp(-E/k_B T)/\mathcal{Z}, \quad (13)$$

where $g^*(E)$ is the true density of states. In general other weights can be used in summing the partition function. If one chooses $B(E,T) = 1/g_n(E)$, then the probability distribution of E for the $(n+1)$th iteration $P_{n+1}(E)$ will be given by,

$$P_{n+1}(E) = g^*(E)/g_n(E)\mathcal{Z}, \quad (14)$$

where $g_n(E)$ is the density of states calculated by the nth iteration. $\mathcal{Z}$ is an undetermined normalization constant. The true density of states can then be estimated by the accumulated histogram of the $(n+1)$th iteration.

$$g_{n+1}(E) = H_{n+1}(E)g_n(E) \times \text{constant}. \quad (15)$$

This is analogous to the iteration process employed in Lee's entropic sampling [4], but we used the fast diffusion of the Wang–Landau algorithm in the early stage. Fig. 3 compares the accuracy of the Wang–Landau method with two different modification sequences for the 32×32 ferromagnetic Ising model. The energy range was $E/N_E \in [-1.55, -1.35]$ where $N_E = 1024$ is the total number of lattice sites. The vertical axis is the error of density of states defined by,

$$\Delta = \frac{1}{m}\sum_{E}^{m}\left[1 - \frac{g_{n+1}(E)}{g^*(E)}\right]^2, \quad (16)$$

where $g^*(E)$ is the exact density of states calculated from a MATHEMATICA program provided by Beale [24]. $g_{n+1}(E)$ is the calculated density of states and m is the total number of energy levels in the summation over this energy range. We plot Δ for different sequences of modification factors. Empty circles were generated with modification factors $f_{k+1} = \sqrt{f_k}$ (with $f_0 = \exp(1)$) and stopping when the condition $(H_{\max} - H_{\min})/(H_{\max} + H_{\min}) \leqslant 0.1$ is satisfied. Where $H_{\max}$ and $H_{\min}$ are the maximum and minimum histogram counts respectively. Filled squares were obtained with a sequence of modification factors where the limiting value of $f = 1$ was used after 14 iterations. The arrow indicates the location which the modification factor was set to 1. We measure the errors (filled squares) at several Monte Carlo steps per site after we set $f = 1$. Error bars were obtained by averaging over several independent simulations. Accuracy increases rapidly immediately after setting $f = 1$, but in the long run, the limiting case becomes only about twice as accurate as the other sequence.

5. Conclusion

We derived an expression for the error term of the Wang–Landau algorithm. With this, we showed that the fluctuation of the accumulated histogram ΔH_k plays a central role in the accuracy of the Wang–Landau method. We have also proposed that stopping each iteration as soon as ΔH_k becomes saturated would be optimal. The dependence of the saturation values on the modification factor was also investigated and it was found that for the ferromagnetic Ising model and fully frustrated Ising model, $\max\{\Delta H_k\} \propto \log(f_k)^{-1/2}$. With this equation, the saturation values of ΔH_k for small modification factors can be predicted from values obtained with larger modification factors. Perhaps, a more efficient algorithm can be developed. We also studied the effects of using different sequences of modification factors (refinement), in which we presented the limiting case where the modification factor is set to 1 after 14 iterations. There are significant improvements of efficiency for short simulations and improvements become less for longer runs.

Acknowledgements

We wish to thank J.S. Wang for fruitful discussions. This work is supported by the Japan Society for Promotion of Science, the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, DOE-OS through BES-DMSE and OASCR-MICS under Contract No. DE-AC05-00OR22725 with UT-Battelle LLC, and by NSF Grant No. DMR-0341874. The computation of this work has been done using computer facilities of the Supercomputer Center, Institute of Solid State Physic, University of Tokyo (Japan) and the computer facilities of the Center for Computational Sciences, Oak Ridge National Laboratory (USA).

References

[1] D.P. Landau, K. Binder, A Guide to Monte Carlo Simulations in Statistical Physics, Cambridge University Press, 2000.
[2] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, M.N. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087.
[3] B.A. Berg, T. Neuhaus, Phys. Lett. B 267 (1991) 249.
[4] J. Lee, Phys. Rev. Lett. 71 (1993) 211.
[5] P.M.C. de Oliveira, T.J.P. Penna, H.J. Herrmann, Eur. Phys. J. B 1 (1998) 205.

[6] J.S. Wang, L.W. Lee, Comput. Phys. Comm. 127 (2000) 131.
[7] C. Yamaguchi, N. Kawashima, Y. Okabe, Phys. Rev. E 66 (2002) 36704.
[8] F. Wang, D.P. Landau, Phys. Rev. Lett. 86 (2001) 2050.
[9] F. Wang, D.P. Landau, Phys. Rev. B 64 (2001) 56101.
[10] D.P. Landau, S. Tsai, M. Exler, Am. J. Phys. 72 (2004) 1294.
[11] C. Yamaguchi, Y. Okabe, J. Phys. A 34 (2001) 8781.
[12] T.S. Jain, J.J. de Pablo, J. Chem. Phys. 116 (2002) 7238.
[13] N. Rathore, J.J. de Pablo, J. Chem. Phys. 116 (2002) 7225.
[14] Q. Yan, R. Faller, J.J. de Pablo, J. Chem. Phys. 116 (2002) 8745.
[15] B.J. Schulz, K. Binder, M. Müller, Internat. J. Modern Phys. C 13 (2002) 477.
[16] M.S. Shell, P.G. Debenedetti, A.Z. Panagiotopoulos, Phys. Rev. E 66 (2002) 56703.
[17] C. Yamaguchi, N. Kawashima, Phys. Rev. E 65 (2002) 56710.
[18] B.J. Schulz, K. Binder, M. Müller, D.P. Landau, Phys. Rev. E 67 (2003) 67102.
[19] C. Zhou, R.N. Bhatt, Phys. Rev. E 72 (2005) 025701(R).
[20] Q. Yan, J.J. de Pablo, Phys. Rev. Lett. 90 (2003) 035701.
[21] P. Dayal, S. Trebst, S. Wessel, D. Wütz, M. Troyer, S. Sabhapandit, S.N. Coppersmith, Phys. Rev. Lett. 92 (2004) 97201.
[22] S. Trebst, D.A. Huse, M. Troyer, Phys. Rev. E 70 (2004) 046701.
[23] P. Virnau, M. Müller, L.G. MacDowell, K. Binder, J. Chem. Phys. 121 (2004) 2169.
[24] P.D. Beale, Phys. Rev. Lett. 76 (1996) 78.