ELSEVIER

1 March 1999

PHYSICS LETTERS A

Physics Letters A 252 (1999) 257–262

Dynamic weighting in simulations of spin systems

Faming Liang [a], Wing Hung Wong [b]

[a] *Department of Statistics and Applied Probability, National University of Singapore, Singapore 119260*
[b] *Department of Statistics, University of California, Los Angeles, CA 90095-1554, USA*

Received 26 October 1998; revised manuscript received 21 December 1998; accepted for publication 23 December 1998
Communicated by L.J. Sham

Abstract

We establish the invariance property of a dynamically weighted Monte Carlo process and apply the method to the simulation of spin glasses and Ising models. For the two-dimensional Edwards–Anderson model, we obtain an ergodicity time of $O(L^{2.44})$, where L is the linear dimension of the lattice. The results suggest that dynamic weighting is a highly promising new tool for Monte Carlo simulation. © 1999 Published by Elsevier Science B.V.

PACS: 02.70.Lq; 05.50.+q; 75.10.Nr
Keywords: Dynamic weighting; Ising model; Invariance; Spin glass

At low temperature spin glasses and other complex systems have many local energy minima. Conventional Monte Carlo simulations are ineffective for such systems due to the difficulty in crossing the steep energy barriers separating the minima. A major goal of current Monte Carlo research is to find ways to overcome this difficulty. Although the cluster algorithms [1,2] have greatly reduced the autocorrelation time at criticality for the second order phase transition, they are not as effective in the case of the first order phase transition. In this respect, simulated tempering [3], exchange Monte Carlo [14,15,7], multicanonical algorithm [4,5] and a related algorithm [16] have presented some substantial progress. In simulated tempering, temperature is treated as a dynamic variable taking values in a ladder of suitably chosen levels near the critical temperature. This method has been very successfully in the simulation of spin glasses [6,7] and random field Ising models [3]. However, it performed less well if many levels are needed to cover a large energy variation such as in simulating an Ising system at a temperature below the critical point.

We have recently proposed dynamic weighting [8] as a new algorithm for general simulation and optimization tasks. This method relies on an additional dynamic variable, the importance weight, to help the system overcome steep energy barriers. A notable aspect of the proposed transition rule is that they do not necessarily satisfy detailed balance. Instead, it was proposed that the property of *invariance with respect to importance weighting* (IWIW) is used to guide the design of proper transition rules. In this Letter, we discuss the application of dynamic weighting to the simulation of Ising models and spin glasses. We outline the proof of the IWIW property for the transition rules, and present numerical evidence that the method has succeeded in achieving a great reduction in ergodicity times.

Suppose our task is to simulate a system x according to a density $f(x)$. We augment the system to be

PII S0375-9601(99)00006-7

(x, w), where w is a dynamic variable representing the importance weight for x. The following transition rule, called a weighted transition, is an important ingredient of our algorithm: draw y from a proposal transition function $T(x \to y)$ ($T(x \to y)$ take values from a domain where $f(y) > 0$) and compute the Metropolis–Hastings [9] ratio

$$r = r(x, y) = \frac{f(y)}{f(x)} \frac{T(y \to x)}{T(x \to y)} . \tag{1}$$

The proposed transition is accepted with probability $a = a(x, w; y) = wr/(wr+\theta)$. If it is accepted, we set the new state to (y, w') with weight $w' = wr/a$; otherwise, the proposal is rejected and the state is set to (x, w') with weight $w' = w/(1-a)$. Here $\theta = \theta(x, w)$ is a threshold function which can be used to control the ease of acceptance. A simulation run then produces a sequence of samples (x_i, w_i), $i = 1, 2, \ldots$. The expectation $E_f \rho(x)$ for a state function $\rho(x)$ can be estimated by the weighted average $[\sum \rho(x_i) w_i] / [\sum w_i]$ over samples obtained after the process has reached equilibrium.

We now verify that such a weighted transition satisfies IWIW, i.e., it would transform correctly weighted densities to correctly weighted densities. We assume that the sample space $\mathcal{X}$ is finite, and that the initial weight is $w_0 = 1$. In this case, the weight space $\mathcal{W}$ is countable. Suppose $g_0(x, w)$ is a correctly weighted density, i.e.,

$$\sum_w g_0(x, w) w = c f(x) , \tag{2}$$

for some c, and $g_1(x', w')$ is the density after one step transition from the density $g_0(x, w)$, then our task is to show that $g_1(x', w')$ is also correctly weighted in the sense that

$$\sum_{w'} g_1(x', w') w' = c' f(x') . \tag{3}$$

From the definition of the transition rule, we can write

$$\begin{aligned} g_1(x', w') = & \sum_x \sum_w g_0(x, w) T(x \to x') \\ & \times a(x, w; x') \delta\left(w', \frac{wr}{a(x, w; x')}\right) \\ & + \sum_x \sum_w \sum_y g_0(x, w) T(x \to y)(1 - a(x, w; y)) \\ & \times \delta(x', x) \delta\left(w', \frac{w}{1 - a(x, w; y)}\right) . \end{aligned}$$

The two terms in the above expression reflect the acceptance and the rejection scenario, respectively. Hence

$$\sum_{w'} g_1(x', w') w' = \mathrm{I} + \mathrm{II} ,$$

$$\begin{aligned} \mathrm{I} = & \sum_{w'} \sum_x \sum_w g_0(x, w) T(x \to x') a(x, w; x') \\ & \times \delta\left(w', \frac{wr}{a}\right) w' , \end{aligned}$$

$$\begin{aligned} \mathrm{II} = & \sum_{w'} \sum_x \sum_w \sum_y g_0(x, w) \\ & \times T(x \to y)(1 - a(x, w; y)) \\ & \times \delta(x', x) \delta\left(w', \frac{w}{1 - a}\right) w' . \end{aligned}$$

It suffices to show that each of (I) and (II) is proportional to $f(x')$.

Summing out w' in part (I) and then substituting (1) and (2) into it, we can see that it is proportional to $f(x')$. Similarly, we can prove that (II) is also proportional to $f(x')$. Hence the density is still correctly weighted after the transition.

The usual Metropolis rule can be regarded as a special type of IWIW transition: if we apply a Metropolis transition to x and leave w unchanged, the result satisfies IWIW. Thus correctly weighted densities will remain so when weighted transitions and Metropolis transitions are alternated in the same run of the Markov chain.

With dynamic weighting, the system can make a move against a steep energy barrier without waiting for a long time as in the standard Metropolis algorithm. This is achieved by adjusting the weight by carrying the ratio $r(x, y)$ when the proposal is accepted and dividing by the rejection probability when the proposal is rejected. When $\theta > 0$, the weights have a tendency

to increase if there are frequent rejections, and this eventually enables the sampler to escape from a deep local minimum. Thus the importance weights facilitate faster approach to equilibrium. We note that it is useful to alternate the weighted transitions and Metropolis or heat bath transitions in the same Markov chain. This will be illustrated in the examples below.

We start with the 2D Edwards–Anderson Ising model with the Hamiltonian

$$H = -\sum_{\langle i,j\rangle} J_{ij}\sigma_i\sigma_j\,, \tag{4}$$

where the spins $\sigma_i = \pm 1$, $\langle i,j\rangle$ denotes the nearest neighbors on the lattice, and J_{ij}'s are independent quenched random variables. (J_{ij} takes values ± 1 with probability 0.5.) The corresponding Boltzmann density at configuration σ is

$$f(\sigma) = \frac{1}{Z(K)}\exp\left\{K\sum_{\langle i,j\rangle} J_{ij}\sigma_i\sigma_j\right\}, \tag{5}$$

where $Z(K)$ is the partition function, and K is the coupling constant or inverse temperature.

We impose the constraint $\sum J_{ij} = 0$ on the realization of J and adopt periodic boundary condition for the configuration σ. Our simulations are performed on lattices of linear dimensions $L = 16, 24, 32, 48, 64$. As in simulated tempering, the simulations are taken on a temperature ladder, $K_1 < K_2 < \ldots < K_n$, where K_i is the inverse temperature at the ith level, and n is the number of levels used. We take $n = 0.625L$ and $K_1, K_2, \ldots, K_n$ to be roughly equally spaced in $0.4 \leqslant K \leqslant 2.5$. The number of levels is smaller than $n = 1.25L$ used by simulated tempering [6]. Within the same level, single-spin-flip heat bath update is used to generate new configurations, meanwhile the weights were left unchanged. The weighted transition rule is only used to govern jumps between levels. After each heat bath sweep, we randomly propose to move to another adjacent level with equal probability. In this experiment, the threshold function is set to be $\theta(x,w) = 1$ if $w < B$ and $\theta = 0$ if $w > B$, where B is a weight control parameter. (We have also performed computation with $\theta(x,w) \equiv 1$, and obtained similar results, although the distribution of the weight tends to have a heavier tail in that case.) We set $B = 10^6$ in all our computations. Each run of the Markov chain starts with one realization of J and one initial random configuration, and it proceeds until 10 000 configurations are obtained at the final level. Fig. 1 is the energy e per spin for this model against the temperature $1/K$. The smooth curve is a high temperature expansion result [10]

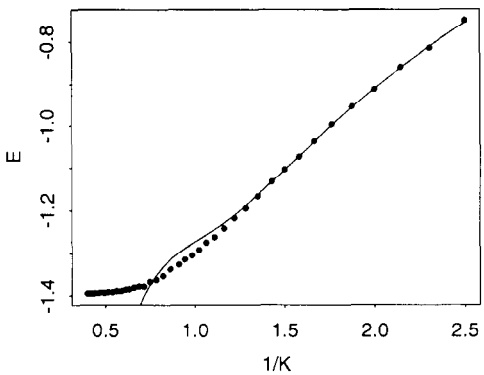

Fig. 1. The energy versus the coupling constant K of the two-dimensional $\pm J$ Ising spin-glass model. The smooth curve is a high-temperature expansion result. Points are averages of 10 runs on a 64^2 lattice.

$$e = -2x + 4x^7 - 4x^9 + \ldots\,, \tag{6}$$

where $x = \tanh(K)$. The expansion gives a good approximation up to $K \approx 0.6$. The points are averages over 10 independent runs. For each run, we discard the initial 100 to 500 samples (for warm-up) and sort the remaining according to the value of the state function of interest (in this case, the energy per spin). We stratify the samples into 10 strata with the $(j \times 10)$ percentiles as strata boundaries, $j = 1, 2, \ldots, 9$. Within each stratum, we truncate the highest 1% weights to the value of the 99% percentile of the weights within that stratum. The weighted estimate is then computed with these weights. The purpose of truncation is to avoid the possibility of having a few outlying weights dominate the weighted averaging process. Without stratification, truncation can produce biased estimates. We found that, with stratification according to the function of interest, this bias can be reduced to a minimal or non-existent level.

At the higher temperatures, our estimated values agree with the theoretical result. Our estimate of ground-state energy per spin, $e_0 = -1.426 \pm 0.01$, is consistent with previous estimates $e_0 = -1.394 \pm 0.0007$ [5], $e_0 = -1.407 \pm 0.008$ [10], and transfer-matrix result $e_0 = -1.4024 \pm 0.0012$ [11]. Our en-

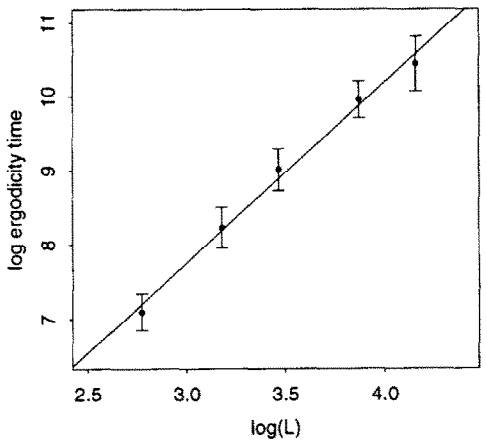

Fig. 2. Plot of the logarithm of ergodicity times versus the logarithm of linear lattice sizes. The straight line is the least square fit.

tropy estimate is $s_0 = 0.0769 \pm 0.0067$, which is also consistent with the estimate $s_0 = 0.081 \pm 0.004$ [5], $s_0 = 0.071 \pm 0.007$ [10] and $s_0 = 0.0701 \pm 0.005$ [11]. From these comparisons, we know that dynamic weighting has produced well-mixed samples and accurate weighted estimates.

Similar to [5,6], we define the ergodicity time τ_e as the mean time (in Monte Carlo sweeps) for the system to travel from the first level to the final level and back. Fig. 2 is a log–log plot (base e) of the ergodicity time versus the linear size of the lattice. The data are consistent with a straight line fit. We get the dependence

$$\tau_e \sim L^{2.44(12)}, \tag{7}$$

in the least square fitting. This is a substantial improvement over simulated tempering [6] and multicanonical method [5], in which the values of 4.27(8) and 4.4(3) are obtained respectively for the exponent. In our computation the normalizing constants of each level are estimated by a fully automatic procedure based on a short pilot run. We believe that the exponent of 2.44 can be further reduced to about 2.0 with more careful tuning of these constants. We also noticed that the exchange Monte Carlo [7] yields $\tau_e \sim V^{1.5}(L^{4.5})$ in the 3D case on a simple cubic lattice. But the comparison should not be made directly. The ergodicity time defined in Refs. [5,6] and this paper actually measures the "tunneling" time from the disordered phase to a ground state and back, whereas the highest β-value in Ref. [7] seems too low to sample the ground states.

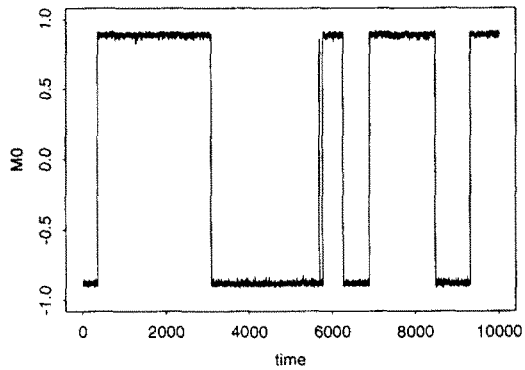

Fig. 3. Plot of spontaneous magnetization M_0 at the final temperature level ($K = 0.5$) versus time. The lattice size is 128^2.

Next we apply dynamic weighting to the 2D Ising model, which has a Hamiltonian given by (5) with $J_{ij} \equiv 1$. When the temperature is at or below the critical point ($K = 0.4407$) there are two oppositely magnetized states separated by a very steep energy barrier. Because of its symmetry, this model is more amenable to theoretical analysis. However, for a sampling algorithm that does not rely on the symmetry of the model (as in simulated tempering and our method), this is a harder problem than spin glasses.

We performed dynamic weighting simulation on lattices of size 32^2, 64^2 and 128^2. The levels of K are uniformly spaced in the range $[0.4, 0.5]$, with number of levels equal to 6, 11, 21, respectively. As in the last example, single-spin-flip heat bath update is used to generate new configurations within the same level, and the weighted transition is used to govern jumping between levels. The choice of θ remains the same as above. In the initial configuration all spins are set to be $+1$. We performed five independent runs. In each run, the simulation continues until 10 000 configurations are obtained at the final temperature level. In the 128^2 case, the average number of sweeps in each run is 776547.

Fig. 3 plots the spontaneous magnetization M at the lowest temperature level for one simulation on the 128^2 lattice. It is clear that the algorithm has succeeded in crossing the very steep barrier separating the two ground states and the system is able to traverse freely between the two energy wells.

Fig. 4 plots the expectation of the absolute value

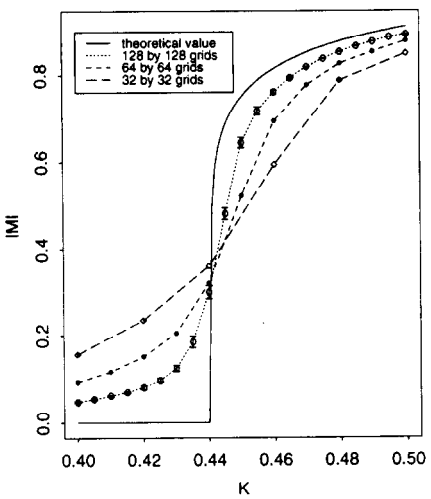

Fig. 4. The expectation of the absolute value of spontaneous magnetization versus the coupling constant K in Ising models with lattice size 32^2, 64^2 and 128^2. The smooth curve is the Onsager result. The points are averages of 5 independent runs. For clarity of picture, error bars are given only in the case of the 128^2 lattice.

of the spontaneous magnetization at various values of K for lattices of size 32^2, 64^2 and 128^2. The smooth curve is the celebrated infinite lattice result found by Onsager [12], and proved by Yang [13]. The expectations are estimated by weighted averaging as discussed before. The only difference is that now the stratification is according to spontaneous magnetization. Clearly, the phase transition phenomenon has been captured with increasing accuracy as the lattice size grows. The critical point can be estimated quite well from the crossing of the curves for the 64^2 and 128^2 cases. We note that a major strength of our method is that a single run of the process is sufficient to yield accurate estimates over the entire temperature range extending well below the critical point.

For comparison, we applied simulated tempering [3] to the Ising model in the same range 0.4 to 0.5 for K. With 51 equally spaced levels of K, tempering performs well for the 32^2 lattice. For the case of 64^2 and 128^2 lattices, we simulated until 50000 configurations are obtained at the final level, but the system was not able to sample both energy wells in the same run. This is so regardless of the number of levels used in the temperature ladders. (We have tried runs with 11, 21, 51, 101 levels for the 64^2 lattice, and 21, 51, 101, 201 levels for the 128^2 lattice.)

In another comparison, we applied the exchange Monte Carlo method [14,15,7] to the same Ising model. With the same 51 equally space temperature level as used in the last experiment, the exchange method works well for the 32^2 and 64^2 lattices. So we can conclude that the exchange method works better than tempering. But for the 128^2 lattice, the system was not able to sample both energy wells in the same run even after 1 000 000 Monte Carlo steps have been run. (Various temperature schemes have been tried, including 21, 51, 101, 201 levels.) What we observed is that a configuration which has stayed at a lower temperature for some time, typically has very small probability of being exchanged to the next higher temperature level, and when it does so the chance is extremely high that it will be moved back to the lower temperature during the next few exchanges.

We suspect that the reason for the poor behavior is that the Boltzmann distribution varies drastically with K. To satisfy the algorithms' requirement that successive distributions should have considerable overlap, many levels are needed. Since the time for a traversal of the temperature ladder grows at least as the square of the number of levels, this leads to a very large ergodicity time when the lattice size is large.

In other experiments, we have used dynamic weighting in the simulation of random field Ising models on a 20^3 lattice. Again, the algorithm performs effectively and does not suffer from any slowing down in the critical range $0.2 < K < 0.3$.

In conclusion, we have introduced a methodology for incorporating importance weighting dynamically into the Monte Carlo process. When used together with recently proposed techniques in temperature/energy laddering [3,4], dynamic weighting can greatly reduce the relaxation time in simulations of various spin systems. We have reported elsewhere [8] that dynamic weighting is also very effective in optimization problems such as neural network training and the traveling salesman problem. These suggest that the method is a promising tool for many similar or more challenging tasks in scientific computation, such as VLSI design, protein folding, and the simulation of materials.

Acknowledgement

This research is supported by NSF DMS-9703918 to W. H. Wong, and by internal research funds from

the division of physical sciences at UCLA.

References

[1] R.H. Swendsen, J.S. Wang, Phys. Rev. Lett. 58 (1987) 86.
[2] U. Wolff, Phys. Rev. Lett. 62 (1989) 361.
[3] E. Marinari, G. Parisi, Europhys. Lett. 19 (1992) 451.
[4] B.A. Berg, T. Neuhaus, Phys. Lett. B 267 (1991) 249.
[5] B.A. Berg, T. Celik, Phys. Rev. Lett. 69 (1992) 2292.
[6] W. Kerler, P. Rehberg, Phys. Rev. E 50 (1994) 4220.
[7] K. Hukushima, K. Nemoto, J. Phys. Soc. Japan 65 (1996) 1604.
[8] W.H. Wong, F. Liang, Proc. Natl. Acad. Sci. USA 94 (1997) 14220.
[9] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087; W.K. Hastings, Biometrika 57 (1970) 97.
[10] J.S. Wang, R.H. Swendsen, Phys. Rev. B 38 (1988) 4840.
[11] H.F. Cheung, W.L. McMillan, J. Phys. C 16 (1983) 7027.
[12] L. Onsager, Nuovo Cimento (Suppl.) 6 (1949) 261.
[13] C.N. Yang, Phys. Rev. 85 (1952) 808.
[14] C.J. Geyer, Computing Science and Statistics: Proc. 23rd Symposium on the Interface (1991) p. 156.
[15] M.C. Tesi, E.J.J. Van Rensburg, E. Orlandini, S.G. Whittington, J. Phys. A 29 (1996) 2451.
[16] A.P. Lyubartsev, A.A. Martsinovski, S.V. Shevkunov, P.N. Vorontsov-Velyaminov, J. Chem. Phys. 96 (1992) 1776.