Available online at www.sciencedirect.com

ScienceDirect

Physics Procedia 68 (2015) 120 – 124

Physics Procedia

28th Annual CSP Workshop on "Recent Developments in Computer Simulation Studies in Condensed Matter Physics", CSP 2015

0.234: The myth of a universal acceptance ratio for Monte Carlo simulations

Christopher C. J. Potter[a], Robert H. Swendsen[b]

[a]*Department of Mathematical Sciences, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213 USA*
[b]*Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

Abstract

Two well-known papers by Gelman, Roberts, and Gilks have proposed the application of the results of an interesting mathematical proof to practical optimizations of Markov Chain Monte Carlo computer simulations. In particular, they advocated tuning the simulation parameters to select an acceptance ratio of 0.234. In this paper, we point out that although the proof is valid, its significance is questionable, and its application to practical computations is not advisable. The simulation algorithm considered in the proof is very inefficient and produces poor results under all circumstances.

Peer-review under responsibility of The Organizing Committee of CSP 2015 Conference

Keywords: Monte Carlo simulation; efficiency; global updating

1. Introduction

The broad range of uses of Monte Carlo (MC) computer simulations make the optimization of such computations an important practical issue, as well as an intrinsically interesting subject of research. Two widely cited papers by Gelman et al. (1996) (GRG) and Roberts et al. (1997) have claimed that for MC simulations of models with a large number of degrees of freedom, the optimal acceptance ratio should have the numerical value of 0.234. The elegant simplicity of the GRG result, as well as the fact that it is based on a mathematical theorem, has brought it a great deal of attention, with each paper receiving more than 750 citations in the literature. Unfortunately, the acceptance ratio of 0.234 is a very poor choice for practical applications. The main reason is that the mathematical proof is only valid for an extremely inefficient version of MC, in which each MC step attempts to alter all variables simultaneously (global updating). A second reason is that their proof is limited to the special case of independent variables, which is not of practical interest. Finally, the criterion for efficiency used in the GRG proof is not appropriate for serious

* Robert H. Swendsen. Tel.: 412-268-5211 ; fax: 412-681-0648.
E-mail address: swendsen@cmu.edu

Peer-review under responsibility of The Organizing Committee of CSP 2015 Conference
doi:10.1016/j.phpro.2015.07.120

MC simulations. On the other hand, the unusual MC algorithm considered in Gelman et al. (1996) and Roberts et al. (1997) does have interesting properties that provide a cautionary tale for commonly used algorithms.

We'll first discuss what had been known about optimizing Monte Carlo simulations prior to the GRG work, and then go on to describe the GRG theorem and its consequences.

2. How to judge the efficiency of an MC algorithm

The standard criterion for determining the efficiency of MC simulations was derived by Müller-Krumbhaar and Binder (1973), who showed that the accuracy with which an observable X can be determined in an MC simulation was given by

$$\delta X = \sqrt{\left(\frac{1+2\tau_X}{N_{MC}}\right)\left(\langle X^2\rangle - \langle X\rangle^2\right)} \tag{1}$$

where $\tau_X = \sum_{t>0} f_X(t)$ and

$$f_X(t) = \frac{\langle X(t')X(t'+t)\rangle - \langle X\rangle^2}{\langle X^2\rangle - \langle X\rangle^2}. \tag{2}$$

This criterion obviously depends on the quantity being measured, and a given algorithm can estimate different quantities with different efficiencies. In particular, local variables, like the position of a single particle, and global variables, like the total energy, are not necessarily optimized at the same acceptance ratio. This is indeed the case for the algorithm used by Gelman and co-workers.

3. Previous work on optimizing Monte Carlo simulations

Prior to the work of GRG, Bouzida et al. (1991), Bouzida et al. (1992), and Bouzida et al. (1993) had analyzed the optimization of MC algorithms in the context of simulations of biological molecules. Bouzida and co-workers considered algorithms that updated multiple variables simultaneously because of the need to deal with very anisotropic (and time-dependent) potentials in biological models. They demonstrated that the optimal acceptance ratio depended on the number of variables being updated, and found optimal acceptance ratios for updating one, two, or three variables in a single MC step.

4. The GRG theorem

As mentioned above, Gelman et al. (1996) and Roberts et al. (1997) constructed a mathematical proof that the optimal acceptance ratio is 0.234, but only for a special simulation algorithm, a limited definition of "optimal," and a large number of updated variables. The model they considered consisted of N variables, subject to a Hamiltonian of the form

$$H = \sum_{j=1}^{N} \phi\left(x_j\right), \tag{3}$$

where $\phi\left(x_j\right)$ is a fairly general function of the variable x_j. The most important feature of Eq. (3) is that the random variables $\left\{x_j | j = 1,\ldots,N\right\}$ are mutually independent. This means that the partition function can be written as

$$\begin{aligned} Z_N &= \int dx_1 \cdots \int dx_N \exp\left[-\beta \sum_{j=1}^{N} \phi\left(x_j\right)\right] = \int dx_1 \cdots \int dx_N \prod_{j=1}^{N} \exp\left[-\beta\phi\left(x_j\right)\right] \\ &= \prod_{j=1}^{N} \int dx_j \exp\left[-\beta\phi\left(x_j\right)\right] = \left(Z_1\right)^N, \end{aligned} \tag{4}$$

where $\beta = 1/k_B T$, and

$$Z_1 = \int dx_1 \exp\left[-\beta\phi\left(x_1\right)\right]. \tag{5}$$

Since the evaluation of Z_1 in Eq. (5) requires only a one-dimensional integral, it is trivial to evaluate by direct numerical integration. The model defined in Eq. (3) is not a suitable candidate for a serious Monte Carlo simulation.

If an MC simulation is to be performed, the standard Metropolis method simulates each variable independently. A particular variable, x_k, would be chosen, a trial move, $\delta x_k \in [-\Delta, +\Delta]$, would be proposed, and it would be accepted with a probability $\exp\left[-\beta\left(\phi(x_k + \delta x_k) - \phi(x_k)\right)\right]$. As shown by Bouzida et al., if $\phi(\cdot)$ represents a simple harmonic oscillator, the optimal acceptance ratio would be close to 0.5 for arbitrary N.

GRG did not consider the standard Metropolis method. Instead, they used a multi-variable (global) updating scheme, in which a trial move perturbed every variable,

$$\left\{x_j | j = 1, \ldots, N\right\} \rightarrow \left\{x_j + \delta x_j | j = 1, \ldots, N\right\}, \tag{6}$$

and was accepted with a probability

$$\exp\left[-\beta \sum_{k=1}^{N} \left[\phi(x_k + \delta x_k) - \phi(x_k)\right]\right]. \tag{7}$$

GRG did not refer to Müller-Kumbhaar and Binder's work, but they used essentially the same criterion. However, the observable they used was $X = x_1$, rather than a more appropriate global variable like the total energy. They correctly noted that any function of x_1 would give the same correlation time, but they did not discuss the MC measurement of global variables. For global updating, using local ($X = x_1$) relaxation time as a criterion, GRG proved that the optimal acceptance ratio is close to 0.234. However, as shown by Potter and Swendsen (2013), global variables, such as the total energy, give different – and larger – correlation times, as well as shifting the optimal acceptance ratio to higher values. This is illustrated in Table 1, which gives the optimal acceptance ratios and the corresponding correlation times for systems of N simple harmonic oscillators using the GRG global updating algorithm. It can be seen that the values are similar for local and global energies, but the optimal acceptance ratios for the total energy are much higher.

N	acc ratio	τ_x (local)	acc ratio	τ_E (total)	acc ratio	τ_E (local)
1	0.431	1.7	0.495	1.8	0.495	1.8
2	0.338	3.5	0.425	3.6	0.387	2.7
4	0.295	5.9	0.407	6.1	0.336	4.6
8	0.261	12.5	0.395	13.8	0.275	7.5
16	0.249	24.6	0.373	25.5	0.258	13.7

Table 1. Optimal acceptance ratios and correlation times for displacement and energy in a system of simple harmonic oscillators. The single-variable update for this system is equivalent to global update with $N = 1$ for both the optimal acceptance ratios and the correlation times.

5. The inefficiency of global updating

The most important reason for the irrelevance of the GRG theorem for practical MC simulations is that global updating is extremely inefficient. To see the origin of this inefficiency, suppose that choosing $|\delta x_1| < \delta_1$ for single-variable updating gives an optimal acceptance ratio A. For an N-dimensional move $\delta\vec{x} = \{\delta x_j | j = 1 \cdots N\}$ achieving the same acceptance ratio would require reducing the single-variable step size to $|\delta_j| < \delta_N \approx \delta_1/\sqrt{N}$. This, in turn, would increase the number of steps needed by a factor of N. Potter and Swendsen (2013) have demonstrated exactly this behavior for both single-variable and global correlation times. The increase in the τ's for increasing N can also be seen clearly in Table 1.

6. Algorithm testing with a double-well potential

Another test of the efficiency of the GRG algorithm is provided by a simple double well potential

$$\phi(x) = x^4 - x^2 \tag{8}$$

When simulated with single-variable MC at an inverse temperature of $\beta = 1/k_B T = 1$, this system relaxes quite rapidly. However, for global updating, relaxation becomes so slow that we did not find it practical to compute correlation times. Instead, we measured the time-dependent relaxation of the average position

$$X = \frac{1}{N} \sum_{j=1}^{N} x_j. \tag{9}$$

It can be seen from Fig. 1 that relaxation is very slow for large N. With $N = 1024$, the value of X is still above 0.45 after 10^4 sweeps. Additional data for the double-well potential is shown in Table 2, which shows the increase with N of the number of MC sweeps needed to reduce the mean displacement by a factor of two.

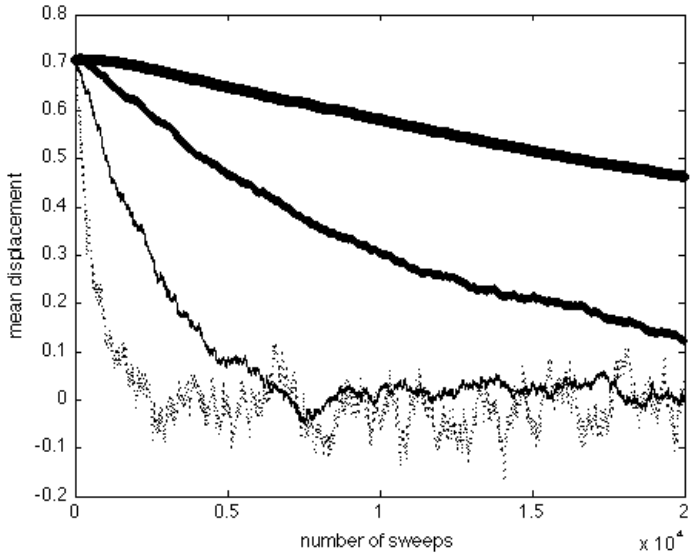

Fig. 1. Time evolution of mean displacement for globally updated independent symmetric double peak system with $N = 16$, 64, 256, and 1024, in order of increasing line thickness. The simulation times are in units of 10^4 MC sweeps. Initial displacement was $\sqrt{2}/2$ which is a local maximum for this distribution, and the expectation value for the mean displacement is 0. A single-variable update would have the same performance as the globally updated system with $N = 1$, which is not shown here because the mean displacement reaches 0 within at most a few steps and is statistical noise thereafter, but it is clear that for smaller N the equilibrium is reached in fewer time steps.

7. Conclusions

It should be emphasized that the GRG theorem is not generally valid for large systems, but only for Monte Carlo moves that involve many variables. If a large system is simulated with MC moves that each involve a small number of variables, the optimal acceptance is higher than 0.234, and the overall efficiency is much greater.

However, the extreme inefficiency of the global updating algorithm considered by Gelman et al. (1996) and Roberts et al. (1997) makes the optimization of its acceptance ratio relatively unimportant, since it should almost never be used for serious computer simulations.

N	time to halve mean displacement
4	32
16	248
64	2119
256	7920
1024	29412

Table 2. Time to halve mean displacement for a system of symmetric anharmonic oscillators with all oscillators initialized to displacement $1/\sqrt{2}$, using GRG global updating. This is the first time step during which mean displacement was $1/2\sqrt{2}$ or less. For each N, the acceptance ratio was the optimal value for measuring displacement. The plot for single-variable updating (global update with $N = 1$) is not shown because relaxation it too rapid. It is clear that the displacement-halving time in creases with increasing N, and single-variable updating relaxes more quickly than a global updating.

References

Bouzida, D., Kumar, S., Swendsen, R.H., 1991. Almost Markov processes, in: Landau, D.P., Mon, K.K., Schüttler, H.B. (Eds.), Computer Simulation Studies in Condensed Matter Physics III (Athens, Georgia, February 1990). Springer-Verlag, Berlin-Heidelberg p. 193, pp. 12–16.

Bouzida, D., Kumar, S., Swendsen, R.H., 1992. Efficient Monte Carlo methods for computer simulation of biological molecules. Phys. Rev. A 45, 8894–8901.

Bouzida, D., Kumar, S., Swendsen, R.H., 1993. A simulated annealing approach for probing biomolecular structures, in: Mudge, T., Milutinovic, V., Hunter, L. (Eds.), Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences 1993, IEEE Computer Society Press. pp. 736–742.

Gelman, A., Roberts, G.O., Gilks, W.R., 1996. Efficient Metropolis jumping rules. Bayesian Statistics 5, 599–608.

Müller-Krumbhaar, H., Binder, K., 1973. Dynamic properties of the Monte Carlo method in statistical mechanics. J. Stat. Phys. 8, 1–24.

Potter, C., Swendsen, R.H., 2013. Efficiency and time-dependent cross correlations in multivariable monte carlo updating. Phys. Rev. E. 88, 053301.

Roberts, G.O., Gelman, A., Gilks, W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. Ann. Appl. Probab. 7, 110–120.