1 2 3 4 5 6 7 8 9 10 11 12 13

Optimal Multigrid Algorithms for the Massive Gaussian Model and Path Integrals

A. Brandt & M. Galun

Dept. of Applied Mathematics & Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel

Abstract

Multigrid algorithms are presented which, in addition to eliminating the critical slowing down, can also eliminate the "volume factor". The elimination of the volume factor removes the need to produce many independent fine-grid configurations for averaging out their statistical deviations, by averaging over the many samples produced on coarse grids during the multigrid cycle. Thermodynamic limits of observables can be calculated to relative accuracy ε_r in just $O(\varepsilon_r^{-2})$ computer operations, where ε_r is the error relative to the standard deviation of the observable. In this paper, we describe in details the calculation of the susceptibility in the one dimensional massive Gaussian model, which is also a simple example of path integral. Numerical experiments show that the susceptibility can be calculated to relative accuracy ε_r in about $8\varepsilon_r^{-2}$ random number generations, independently of the mass size.

KEY WORDS: Multigrid; massive Gaussian model; Monte Carlo; critical slowing down; volume factor; thermodynamic limit; path integrals.

1 Introduction

One of the aims in statistical physics is to calculate various average properties of configurations governed by the Boltzmann distribution. This is usually done by measuring these averages over a sequence of Monte Carlo iterations. Unfortunately, such processes tend to suffer from *several* independent inefficiency factors that multiply each other and thus produce very expensive computations.

The best known of these inefficiency factors is the *critical slowing down* (CSD). This is the phenomenon, typical to simulations of critical systems, that with the increase in lattice size there also comes an increase in the number of full Monte Carlo passes over the lattice needed to produce a new configuration which is statistically "useful", i.e., substantially independent of, or only weakly correlated to, a former configuration. Considerable efforts have been devoted to reduce the critical slowing down. For simple cases with real variables, classical multigrid methods [5,7,10] can eliminate the CSD. For more complicated models, (e.g. ϕ^4 or discrete models) more recent publications report on simulation techniques that partially [7,12] or completely [1,8,9,13] eliminate the CSD. This means that the time to produce an independent configuration is proportional to the number of gridpoints.

The elimination of the CSD is very important but there is another no less important factor of slowness, this is the *volume factor*. To calculate a thermodynamic quantity to a certain *relative* accuracy ε_r, one needs to produce $O(\varepsilon_r^{-2})$ essentially independent configurations to average out the deviation exhibited by each of them, where the *relative* accuracy ε_r is the error relative to the standard deviation of the observable in question. Also, the size of the grid must increase as some positive power of ε_r^{-1}. Thus, even if the CSD has been completely eliminated, the overall work increases as $O(\varepsilon_r^{-2} N^d)$, where N is the linear lattice size and d is the dimension. An important advantage of the multigrid approach is that it can drastically reduce the volume factor N^d as well, by averaging over many samples produced in prolonged Monte Carlo passes on coarse grids. Indeed, we will exhibit cases in which the volume factor is completely eliminated, together with the CSD.

The elimination of both the volume factor and the CSD means that a thermodynamic limit can be calculated to an accuracy of $\pm\varepsilon$ in *optimal time*, i.e. in only $O(\varepsilon^{-2})$ *computer operations*. This is just the same order of complexity as needed to calculate, by statistical trials, any simple "pointwise" average, such as the frequency of "heads" in coin tossing. By contrast, both the volume and the CSD factors *multiply* the statistical factor (ε^{-2}) in the operation count of conventional algorithms.

The elimination of the volume factor has first been demonstrated [3,4,6] for the Gaussian model with constant coefficients. It has been shown there, for the one-dimensional Gaussian model, that the susceptibility can be calculated to accuracy ε_r in about $4\varepsilon_r^{-2}$ random number generations, while the average energy per degree of freedom requires $3\varepsilon_r^{-2}$ such generations for a similar accuracy. In the two-dimensional Gaussian model, the susceptibility can be measured to accuracy ε_r in

about $20\varepsilon_r^{-2}$ random number generations.

In this paper we treat the one dimensional *massive* Gaussian model and we show that using an appropriate multigrid algorithm the susceptibility can be calculated in an optimal time. Stated differently, we show that the multigrid algorithm effectively produce an independent sample in just $O(1)$ computer operations. More precisely, the result is that the susceptibility is calculated to relative accuracy ε_r in less than $8\varepsilon_r^{-2}$ random generations, essentially independently of the mass size, although the algorithm flow does change with that size (see Sec 2.4). The computational time of this "optimal multigrid" algorithm is thus smaller by a factor N^d compared to that of a *conventional* multigrid algorithm, e.g. such as that of Refs 5, 7 and 10, which measure the observable only once per multigrid cycle (cf. Table III).

From quantum mechanics point of view, the one dimensional massive Gaussian model describes an action of a discretized path integral with quadratic potential (see Sec. 2.2). Therefore, the elimination of the volume factor is important also for path integral calculations.

2 One dimensional massive Gaussian model: fast claculation of the susceptibility

A multigrid algorithm for simple continuous-state models has been described by Brandt, Ron and Amit [5], and independently, by Goodman and Sokal [7]. The later also tested it for the Gaussian model and reported that it indeed eliminates the CSD, for dimension $d \geq 2$. The two approaches are not the same: while Brandt et al. use linear interpolation, Goodman and Sokal employ constant interpolation. We have recently shown [3,4,6] that our multigrid Monte Carlo approach (unlike Goodman and Sokal) can be used not only for eliminating the CSD, but also for eliminating the volume factor. Here we extend this work to the one dimensional massive Gaussian model.

2.1 Continuous case

The Hamiltonian associated with the continuous case is:

$$\mathcal{H}(u) = \int_0^L u_x^2 dx + m^2 \int_0^L u^2 dx, \tag{1}$$

where $u = u(x)$ is a real continuous function (configuration) defined for $0 \leq x \leq L$ and m is a real number denoting the mass size. Homogeneous Dirichlet boundary conditions, $u(0) = u(L) = 0$, are used. Consequently, a general configuration $u(x)$ can be expanded by:

$$u(x) = \sum_{j=1}^{\infty} c_j \sin(j\pi x/L), \tag{2}$$

where the Fourier coefficient c_j are real. The magnetization is given by:

$$M(u) = \frac{1}{L}\int_0^L u(x)dx = \frac{2}{\pi}\sum_{j=1}^{\infty *}\frac{c_j}{j}, \tag{3}$$

where $\sum^*$, here and below, stands for a summation over odd integers. The probability density of each configuration u is given by the density function of the Boltzmann distribution

$$P(u) = \frac{e^{-\mathcal{H}(u)/T}}{Z(T)}, \tag{4}$$

where T is the temperature and $Z(T)$ is a normalization factor. It can be shown (see App. A) that the average magnetization ($\langle M\rangle$) and the susceptibility ($\langle M^2\rangle - \langle M\rangle^2$) are:

$$\langle M\rangle = 0 \tag{5a}$$

$$\langle M^2\rangle - \langle M\rangle^2 = \langle M^2\rangle = \frac{4LT}{\pi^2}\sum_{j=1}^{\infty *}\frac{1}{\pi^2 j^4 + m^2L^2j^2}. \tag{5b}$$

We define any statistics for the continuum as the limit of the statistics for systems truncated to a finite number of Fourier components.

2.2 Discrete case

In order to measure such statistical averages numerically, it is necessary to discretize the system. On a grid with meshsize $h = L/N$, the discretized Hamiltonian $\mathcal{H}_h(u)$, approximating (1), can be written as:

$$\mathcal{H}_h(u) = \frac{1}{h}\sum_{i=1}^{N}(u_i - u_{i-1})^2 + m^2 h\sum_{i=1}^{N-1}u_i^2, \tag{6}$$

where $u_i = u(x_i)$ are the variables at gridpoints $x_i = ih$, $0 \le i \le N$, respectively. For the simplicity of the multigrid algorithm we assume $N = 2^k$. Assuming again zero boundary conditions, $u_0 = u_N = 0$, a general grid configuration can be represented by

$$u_i = \sum_{j=1}^{N-1} c_j \sin(j\pi x_i/L). \tag{7}$$

The discrete magnetization is given by

$$M_h(u) = \frac{h}{L}\sum_{i=0}^{N}u(x_i) = \frac{h}{L}\sum_{j=1}^{N-1 *}c_j\frac{\cos(j\pi h/(2L))}{\sin(j\pi h/(2L))}. \tag{8}$$

Similarly to the continuous case, the probability distribution is given by (4) where $\mathcal{H}_h(u)$ replaces $\mathcal{H}(u)$. Therefore, one can derive (see App. A):

$$\langle M_h \rangle = 0 \tag{9a}$$

$$\langle M_h^2 \rangle = \frac{Th^4}{L^3} \sum_{j=1}^{N-1}{}^{*} \frac{\cos^2(j\pi h/(2L))}{4\sin^4(j\pi h/(2L)) + m^2 h^2 \sin^2(j\pi h/(2L))}. \tag{9b}$$

From quantum mechanics viewpoint , the discretized Hamiltonian (6) can be the action of a discretized path integral u, where u_i , $i = 0, \ldots, N$ are the positions of a single particle, which travels from u_i to u_{i+1} in timestep h. The left part of the Hamiltonian stands for the kinetic energy of the particle and the right part stands for the quadratic potential energy of the particle. The density function for the path u is just as the Boltzmann distribution we have described above.

2.3 Description of the multigrid cycle

Consider the following generalized Hamiltonian:

$$\begin{aligned} \mathcal{H}_h(u) = \frac{1}{h}\sum_{i=1}^{N}(u_i - u_{i-1})^2 + h\sum_{i=1}^{N-1}\phi_i u_i \\ + m^2 h\, b_h \sum_{i=1}^{N-1} u_i^2 + m^2 h\, a_h \sum_{i=1}^{N-1} u_i u_{i-1}. \end{aligned} \tag{10}$$

On the finest grid, $\phi_i = 0 \quad (i = 1, \ldots, N-1), a_h = 0$ and $b_h = 1$ are taken. The more general form of the Hamiltonian is needed for the algorithm recursion.

The coarse grid with meshsize $H = 2h$ is constructed by taking every other grid-point. The coarse-grid function $u^H = (u_0^H, \ldots, u_I^H, \ldots, u_{N/2}^H)$ describes a *displacement* of the fine-grid function $u^h = (u_0, \ldots, u_i, \ldots, u_N)$; i.e., it modifies the latter through *interpolation and addition:*

$$u^h = \tilde{u}^h + I_H^h u^H, \tag{11}$$

where $\tilde{u}^h$ is the fine-grid configuration at the stage of switching to the coarse-grid and I_H^h denotes the *linear* interpolation from grid H to grid h.

The fine-grid Hamiltonian $\mathcal{H}_h(u^h)$ resulting from that interpolation can be written as follows:

$$\mathcal{H}_h(\tilde{u}^h + I_H^h u^H) = \mathcal{H}_h(\tilde{u}^h) + \mathcal{H}_H(u^H), \tag{12}$$

where $\mathcal{H}_h(\tilde{u}^h)$ is given by (10) and $\mathcal{H}_H(u^H)$ is:

$$\mathcal{H}_H(u^H) = \frac{1}{H}\sum_{I=1}^{N/2}(u_I^H - u_{I-1}^H)^2 + H\sum_{I=1}^{N/2-1}\phi_I^H u_I^H + m^2 H b_H \sum_{I=1}^{N-1}(u_I^H)^2 + m^2 H a_H \sum_{I=1}^{N-1} u_I^H u_{I-1}^H, \tag{13}$$

with

$$a_H = \frac{a_h}{2} + \frac{b_h}{4}, \quad b_H = \frac{a_h}{2} + \frac{3b_h}{4} \tag{14a}$$

and

$$\begin{aligned}\phi_I^H = {} & \frac{-\tilde{u}_{i-2}^h + 2\tilde{u}_i^h - \tilde{u}_{i+2}^h}{2h^2} \\ & + \frac{\phi_{i-1}^h + 2\phi_i^h + \phi_{i+1}^h}{4} \\ & + \frac{m^2 b_h}{2}(\tilde{u}_{i-1}^h + 2\tilde{u}_i^h + \tilde{u}_{i+1}^h) \\ & + \frac{m^2 a_h}{4}(\tilde{u}_{i-2}^h + 2\tilde{u}_{i-1}^h + 2\tilde{u}_i^h + 2\tilde{u}_{i+1}^h + \tilde{u}_{i+2}^h) \\ & (I = i/2 = 1, \ldots, N/2 - 1),\end{aligned} \tag{14b}$$

representing fine-to-coarse induced field-like terms. The coefficients a^H and b^H depends only on a^h and b^h. The coarse field terms ϕ_I^H are calculated from the details of the fine-grid configuration at coarsening and are fixed throughout the processing on the coarser level. The variables of the coarse grid u_I^H are initially set to zero, corresponding to zero initial displacements.

Having calculated the field ϕ^H once for all, $\mathcal{H}_H$ is directly calculated in terms of the coarse grid configuration u^H; there is no need to explicitly perform (12) in order to relax the coarser level. One can therefore run a long Monte Carlo process with $\mathcal{H}_H$ (13) before explicitly updating u^h by (11).

The entire algorithm can be described by a sequence of multigrid cycles for the finest level. A cycle for any given ("current") level is *recursively* defined by the following five steps.

1. ν_1 Monte Carlo sweeps are first made on the current level. Then, if this level is the coarsest, go to 5.
2. The next coarser level is created from the current one by determining the coefficients (14a) and the coarse field-like terms (14b).
3. γ multigrid cycles for the coarse level are performed. (γ may change from the current level to another.)

4. Update the current level by performing (11).

5. Additional ν_2 Monte Carlo sweeps are finally made on the current level.

The Monte Carlo sweeps are performed by changing each variable in its turn randomly according to its associated distribution, regarding its neighbors as fixed.

The values of ν_1, ν_2 and γ are discussed below.

The massive Gaussian model displays criticality as $m \to 0$. The described cycle, even with $\gamma = 1$ (a V-cycle), would eliminate the critical slowing down, but the volume factor still remains intact. However, the main issue here is to eliminate the volume factor as well, for any mass size m; the way to do so is described next.

2.4 Fast sampling of susceptibilty

As in the simple Gaussian model the susceptibility ($5b$) is dominated by contributions from large-scale fluctuations (low-frequency Fourier components), regardless of the size m^2. Therefore, the purpose of the simulation is to sample quickly as many such fluctuations as possible. The way to do so is to use a cycle index γ larger than 1 and to calculate the susceptibility over the many measurements on the coarsest level. Furthermore, the optimal multigrid algorithm differs from the one that has been described for the simple Gaussian model [3,4,6]; the cycle index may change from one level to another depending on the parameter m.

The magnetization M_h can be evaluated on any level (plug (11) in (8)), without going back to finer levels. Thus, many measurements of M_h^2 can be made within a cycle, and their average $\overline{M_h^2}$ can be used as an estimate for the discrete susceptibility $\langle M_h^2 \rangle$. In practice, measurements are taken only on the coarsest level, after each relaxation sweep there, since only there substantial changes in M_h are introduced.

We next study the number s_i of relaxation sweeps the algorithm needs to perform on level i, i.e. on the grid with meshsize $h_i = 2^i h, \quad (i = 0, 1, \ldots, \ell = \log_2(N/2))$, in order to achieve accuracy ε in the estimation of the susceptibility. The *total* expected error ε in measuring $\langle M^2 \rangle$ is calculated by Fourier analysis in App. B. From ($B2$), the total error in measuring $\langle M^2 \rangle$ *relatively* to the standard deviation σ, where $\sigma = \sqrt{\langle M^4 \rangle - \langle M^2 \rangle^2} = \sqrt{2}\langle M^2 \rangle = O(\frac{LT}{\pi^2 + m^2 L^2})$, is

$$\varepsilon_r = \frac{\varepsilon}{\sigma} = O\left(\sum_{i=0}^{l} s_i^{-1/2} h_i^3 L^{-3} \frac{\pi^2 + m^2 L^2}{\pi^2 + m^2 h_i^2} \right) + r.d.e \qquad (15)$$

where the last term added here (r.d.e) is the relative discretization error estimator which is discussed in App. C. It is important to emphasize here that raising p, the order of the discretization error, beyond $p = 3$ has no benefit. (See ($C1$). This point, incidentally, was missed in Ref. 4.)

The total work (operations) on all the levels is clearly

$$W = \sum_{i=0}^{l} s_i O(L/h_i). \tag{16}$$

The optimal choice for s_i (yielding either minimal ε for a given W or minimal W for a given ε) is obtained when $\frac{\partial \varepsilon_r}{\partial s_i} + \lambda_1 \frac{\partial W}{\partial s_i} = 0$, which by (15) and (16) yields

$$s_i = \lambda_2 \left(L^{-4} h_i^4 \frac{\pi^2 + m^2 L^2}{\pi^2 + m^2 h_i^2} \right)^{2/3} = \lambda_3 2^{8i/3} \left(\frac{\pi^2 + m^2 L^2}{\pi^2 + m^2 h_i^2} \right)^{2/3},$$

where λ_1, λ_2 and λ_3 are independent of i. Hence, the optimal cycle index at level i is

$$\gamma_{i_{opt}} = \frac{s_{i+1}}{s_i} = 2^{8/3} \left(\frac{\pi^2 + m^2 h_i^2}{\pi^2 + 4m^2 h_i^2} \right)^{2/3}. \tag{17}$$

The actual values of (17) for constructing an optimal multigrid cycle are given in Table I as a function of mh_i. However, we will see in the experiments that the results are not much sensitive to changes of γ within quite large margins. In fact, analyzing the following three cases would show wide ranges of γ at which the optimal order $W = O(\varepsilon_r^{-2})$ is still obtained. For each of the cases we will use *fixed* γ, hence $s_i = s\gamma^i$ where $s = (\nu_1 + \nu_2) \cdot \#\text{cycles}$. Since $h_i L^{-1} = O(2^{i-l})$, we can perform the summations in (15) and (16). Using the discussion in App. C and the relation $\sigma = O(\frac{LT}{\pi^2 + m^2 L^2})$, the general relative discretization error can be calculated in each of the three cases:

1. For the case $h_i < \frac{\pi}{m}$ at all levels $(i = 0, \ldots, l)$

$$\varepsilon_r = O\left(s^{-1/2} \frac{\gamma^{-l/2} - 2^{-3l}}{1 - 2^{-3}\gamma^{1/2}} \right) + O(2^{-lp_*}) \tag{18}$$

and

$$W = O\left(s \frac{\gamma^l - 2^l}{1 - 2\gamma^{-1}} \right) \tag{19}$$

for any $2 < \gamma < 2^6$, where $p_* = \min(3, p)$. Actually, by choosing γ and the approximation order p so that γ is significantly smaller than 2^{2p_*}, the second term in (18) can be ignored, yielding $W = O(\varepsilon_r^{-2})$. While $\gamma = \gamma_{opt}$ indeed minimizes $W\varepsilon_r^2$, the other cycle indices $(2 < \gamma < 2^{2p_*})$ give practically the same efficiency. This case is very similar to the simple Gaussian case [4].

2. For the case $h_i > \frac{\pi}{m}$ at all levels $(i = 0, \ldots, l)$

$$\varepsilon_r = O\left(s^{-1/2} \frac{\gamma^{-l/2} - 2^{-l}}{1 - 2^{-1}\gamma^{1/2}} \right) + O(2^{-l}) \tag{20}$$

and W is as before (19), for any $2 < \gamma < 4$. As any γ in this range is already smaller than 2^2, the second term in (20) can be ignored, yielding again $W = O(\varepsilon_r^{-2})$.

3. As $h \to 0$ the last case will evolve eventually to the case $h_i < \frac{\pi}{m}$ for $i = 0, 1, \ldots, k-1$ and $h_i > \frac{\pi}{m}$ for $i = k, k+1, \ldots, l$. Generally, in this case

$$\begin{aligned} \varepsilon_r = O\Bigg(& s^{-1/2}\gamma^{-l/2} \quad \sum_{j=0}^{\hat{k}} (2^{-1}\gamma^{1/2})^j \Bigg) \\ & + O\Bigg(s^{-1/2}\gamma^{-l/2} \quad \frac{1-(2^{-3}\gamma^{1/2})^k}{1-2^{-3}\gamma^{1/2}} m^2 L^2 (2^{-3}\gamma^{1/2})^{\hat{k}} \Bigg) \\ & + O(2^{-lp_*}), \end{aligned} \tag{21}$$

and W is as before (19), for any $2 < \gamma < 2^6$, where $\hat{k} = l - k$ stays constant as $h \to 0$. As mentioned earlier, a multigrid cycle as described in Table I indeed minimizes $W\varepsilon_r^2$, but by choosing any fixed γ in the domain $(2 < \gamma < 2^{2p_*})$ the third term in (21) can be neglected, again yielding the optimal efficiency $W = O(\varepsilon_r^{-2})$.

2.5 Numerical results

We have tested the multigrid algorithm for different values of m with grid of sizes up to 512. Our main aim was to show that using appropriate values of γ the susceptibility can be calculated in an optimal time, while the use of unsuitable values of γ undermines optimality. The susceptibility has been measured over just *one* cycle. Within the cycle, many measurements are taken, in fact after each Monte Carlo step on the coarsest level, the level with just one internal point, i.e. $h_l = L/2$. The average of the measurements, $\overline{M_h^2}$, is an approximation for $\langle M_h^2 \rangle$ (9*b*), which is also an approximation for the thermodynamic limit $\langle M^2 \rangle$ (5*b*). The *relative* accuracy is defined as $\varepsilon_r = \frac{|\overline{M_h^2} - \langle M^2 \rangle|}{\sigma}$ and it is averaged over an ensemble of 10000 runs [1] . We define α to be the expected value of $\#\mathrm{RAN} \cdot \varepsilon_r^2$, where #RAN is the amount of work spent in the cycle, measured by the number of times a random number is generated. Thus, α should turn out to be a constant if and only if the algorithm solves to relative accuracy ε_r in $O(\varepsilon_r^{-2})$ operations. We measured α for three different cases. Results are presented in Table II for $L = 1$, $T = 1$, $h_0 = 1/N$ and $h_l = 1/2$, showing that the algorithm is not sensitive in a wide range of suitable γ. We see that any appropriate cycle index will lead to the optimal efficiency, i.e. α tends to a constant as N grows (see

[1] The experiments for $m = 64$ and $N = 512$ using Tab. I and $\gamma = 6$ are made over an ensemble of 4000 runs and 400 runs, respectively.

cases $m = 0.5$, $\sigma = 0.05749$ and $m = 64$, $\sigma = 1.672 \cdot 10^{-4}$). In the last case, $m = 400$, $\sigma = 4.397 \cdot 10^{-6}$, α turns out to be a constant when cycle index 3 is used, but cycle index 6, as explained above, is too big for this case. For any case, cycle index 2 (W-cycle) is below the optimal range, demonstrating logarithmic growth of α. The main conclusion is that an optimal algorithm, with practically constant α, can always be devised.

In Table III, we compare between our optimal multigrid Monte Carlo algorithm and a conventional multigrid algorithm, where the susceptibillity is measured *once* per V-cycle [5]. It is clear that better accuracy means using larger grids. Therefore, as the accuracy is improved the ratio between the complexity of the two algorithms increases. For example, in order to achieve a certain accuracy in the case $m = 64$ and $N = 512$, it would cost a conventional algorithm 330 times the work required by the optimal multigrid algorithm, as presented here. Practically, while the computational time of the conventional algorithm [5] is $4N\varepsilon_r^{-2}$ (the cost of conventional algorithms as described in Refs. 7 and 10 would be even somewhat bigger), the computational time of our algorithm is about $8\varepsilon_r^{-2}$, independently of the grid size N. (Note that, for maximal efficiency, a conventional algorithm should use the smallest possible N which still gives $r.d.e$ comparable to ε_r. According App. C, this would mean $N = O(\varepsilon^{-1/p_*})$.)

3 Summary

The calculation of a thermodynamic limit of any observable to a relative accuracy ε_r usually requires by a Monte Carlo process

$$O(N^{d+z}\varepsilon_r^{-2})$$

computer operations, where ε_r is the error relative to the standard deviation of the obervable, N is the linear dimension of the lattice needed to approximate the thermodynamic limit to accuracy ε_r, d is the dimension and z is the critical exponent.

Multigrid algorithms potentially can reduce and even eliminate not only the *critical slowing down* factor N^z but also the *volume factor* N^d.

The parameters of the multigrid algorithm, such as the cycle index γ and the coarse-to-fine interpolation order depends not only on the involved model and its discretiztion but also on the observable in question. For the optimal calculation of the susceptibility in the one dimensional massive Gaussian model it is essential to use linear interpolation and a cycle index which varies with the mass size. In this case the critical slowing down *and* the volume factor are completely eliminated leading to the optimal efficiency $O(\varepsilon_r^{-2})$.

Acknowledgments

The research has been supported in parts by grants No. G0289-065.07/93 from the German-Israeli Foundation for Research and Development (GIF), No. 379/93

from the Israeli Academy of Science and Humanities, F49620-92-J-0439 from the United States Air Force and by the Carl F. Gauss Minerva Center for Scientific Computation.

Appendix A: Fourier transform expressions

In the continuous case, by substituting (2) into (1) and into the left part of (3) one gets

$$\mathcal{H}(u) = \frac{\pi^2}{2L}\sum_{j=1}^{\infty} j^2 c_j^2 + \frac{m^2 L}{2}\sum_{j=1}^{\infty} c_j^2 \tag{A1}$$

and the right hand side of (3). From (4) and (A1), it can be shown by straightforward calculations that

$$\langle c_j \rangle = 0 \tag{A2}$$

$$\langle c_j^2 \rangle = \frac{LT}{\pi^2 j^2 + m^2 L^2} \tag{A3}$$

$$\langle c_j^4 \rangle = \frac{3L^2T^2}{(\pi^2 j^2 + m^2 L^2)^2}. \tag{A4}$$

Hence, the average magnetization ($\langle M \rangle$) and the susceptibility ($\langle M^2 \rangle - \langle M \rangle^2$) can be calculated using (3), (A2) and (A3), leading to results (5a) and (5b) in Sec. 2.1.

In the discrete case, by substituting (7) into (6) and into the left part of (8) one gets

$$\mathcal{H}_h(u) = \frac{2L}{h^2}\sum_{j=1}^{N-1} c_j^2 \sin^2(j\pi h/(2L)) + \frac{m^2 L}{2}\sum_{j=1}^{N-1} c_j^2, \tag{A5}$$

and the right hand side of (8). From (4) and (A5) it can be derived that

$$\langle c_j \rangle = 0 \tag{A6}$$

$$\langle c_j^2 \rangle = \frac{Th^2}{4L\sin^2(j\pi h/(2L)) + m^2 L h^2}. \tag{A7}$$

The average discrete magnetization (9a) and the discrete susceptibility (9b) in Sec. 2.2 are obtained by applying (A6) and (A7) to (8).

Appendix B: Fourier analysis of the expected error in the estimation of the susceptibility

The relaxation sweep on level i (with meshsize $h_i = 2^i h$: $i = 0, 1, \ldots, \ell = \log_2(N/2)$) strongly affects, hence effectively samples, only those Fourier coefficients c_j (cf. (2)) for which $j = O(L/h_i)$. Hence, the number s_i of relaxation sweeps needed to be performed on level i depends on the contribution of these components to the deviations in measuring $\langle M^2 \rangle$. By (3)

$$M^2 = \frac{4}{\pi^2}\sum_{j,k}^{*} c_j c_k/(jk). \tag{B1}$$

Consider first a term (j,k) in (15) for which both j and k are $O(L/h_i)$, hence the term is effectively sampled $O(s_i)$ times in a cycle. According to $(A2)$, $(A3)$ and $(A4)$ in App. A, the standard deviation of the term is

$$\frac{4}{jk\pi^2}\left(\langle (c_j c_k)^2\rangle - \langle c_j c_k\rangle^2\right)^{1/2} = O(h_i^4(\pi^2+m^2h_i^2)^{-1}L^{-3}T),$$

hence the standard deviation of its average over the $O(s_i)$ samples is $O(s_i^{-1/2}h_i^4(\pi^2+m^2h_i^2)^{-1}L^{-3}T)$. There are $O(h_i^{-2}L^2)$ such terms, where each pair of them is uncorrelated, hence their total contribution is

$$O(s_i^{-1/2}h_i^3(\pi^2+m^2h_i^2)^{-1}L^{-2}T).$$

In the case $j = O(L/h_{i-r})$ and $k = O(L/h_i)$ where $r \geq 1$ (i.e., $h_i > h_{i-r}$), the term (j,k) in (15) is effectively sampled as follows (see also ref. 4): in an inner loop, for a (nearly) fixed value of c_j, the values of c_k are averaged $O(s_i/s_{i-r})$ times, yielding an average whose deviation is of the order

$$O\left(\frac{c_j}{jk}\left(\frac{s_i}{s_{i-r}}\right)^{-1/2}\langle c_k^2\rangle^{1/2}\right) = O\left(\frac{c_j}{jk}\left(\frac{s_{i-r}}{s_i}\cdot\frac{LT}{\pi^2k^2+m^2L^2}\right)^{1/2}\right).$$

Then, in an outer loop, the c_j in this average is averaged over $O(s_{i-r})$ samples, giving results with deviations of order

$$\begin{aligned}
&O\left(\frac{s_{i-r}^{-1/2}}{jk}\left(\frac{s_{i-r}}{s_i}LT\right)^{1/2}(\pi^2k^2+m^2L^2)^{-1/2}\langle c_j^2\rangle^{1/2}\right)\\
&= O\left(\frac{s_i^{-1/2}}{jk}(\pi^2k^2+m^2L^2)^{-1/2}(\pi^2j^2+m^2L^2)^{-1/2}LT\right)\\
&= O\left(s_i^{-1/2}h_i^2h_{i-r}^2(\pi^2+m^2h_i^2)^{-1/2}(\pi^2+m^2h_{i-r}^2)^{-1/2}L^{-3}T\right).
\end{aligned}$$

There are $O(h_i^{-1}h_{i-r}^{-1}L^2)$ such terms, effectively uncorrelated, hence their total deviation is $O(s_i^{-1/2}h_i^{3/2}h_{i-r}^{3/2}(\pi^2+m^2h_i^2)^{-1/2}(\pi^2+m^2h_{i-r}^2)^{-1/2}L^{-2}T)$. Summing over integers $r \geq 0$ gives again

$$\begin{aligned}
&O\left(s_i^{-1/2}h_i^{3/2}(\pi^2+m^2h_i^2)^{-1/2}TL^{-2}\sum_{r\geq 0}h_{i-r}^{3/2}(\pi^2+m^2h_{i-r}^2)^{-1/2}\right)\\
&= O(s_i^{-1/2}h_i^3(\pi^2+m^2h_i^2)^{-1}L^{-2}T).
\end{aligned}$$

Therefore, the total error in measuring $\langle M^2\rangle$ is

$$\varepsilon = O\left(\sum_{i=0}^{l}s_i^{-1/2}h_i^3(\pi^2+m^2h_i^2)^{-1}L^{-2}T\right). \qquad (B2)$$

Appendix C: Calculation of the discretization error $\langle M^2\rangle - \langle M_h^2\rangle$.

To calculate the discretization error, observe first that for $\frac{N}{2} < j < N$ the term in $(9b)$ is smaller than

$$\frac{Th^4}{L^3(2+m^2h^2)} \le \min\left(\frac{Th^4}{2L^3}, \frac{Th^2}{m^2L^3}\right),$$

hence the sum of all these $\frac{N}{2}$ terms is less than

$$CLT\min\left(\frac{1}{N^3}, \frac{1}{m^2L^2N}\right),$$

where $C = .5$. A similar estimate, but with a different value of C, is obtained for the sum of all terms $\frac{N}{2} < j < \infty$ in $(5b)$.

For $j < \frac{N}{2}$, each term in $(9b)$ can be approximated by a Taylor expansion as follows:

$$\begin{aligned}&\frac{4LT}{\pi^2}\cdot\frac{1-(\frac{j\pi h}{2L})^2}{\pi^2j^4(1-\frac{2}{3}(\frac{j\pi h}{2L})^2)+m^2L^2j^2(1-\frac{1}{3}(\frac{j\pi h}{2L})^2)}\\ &=\frac{4LT}{\pi^2}\cdot\frac{1-\beta(\frac{j\pi h}{2L})^2}{\pi^2j^4+m^2L^2j^2},\end{aligned}$$

where $\frac{1}{3} < \beta < \frac{2}{3}$. Comparing this with the j^{th} term in $(5b)$, we conclude that the total discretization error for these terms is approximately

$$LT\beta N^{-2}\sum_{j=1}^{N/2-1}{}^{*}\frac{1}{\pi^2j^2+m^2L^2} \le CLT\min\left(\frac{1}{N^2}, \frac{1}{m^2L^2N}\right).$$

For general p-order discretization, a similar estimation would give

$$CLT\min\left(\frac{1}{N^p}, \frac{1}{m^2L^2N}\right).$$

Therefore, the total discretization error is

$$CLT\min\left(\frac{1}{N^{\min(3,p)}}, \frac{1}{m^2L^2N}\right) \le CLTN^{-\min(3,p)},$$

hence the *relative* (to σ) discretization error estimator is

$$r.d.e = \frac{C(\pi^2+m^2L^2)}{N^{p_*}+m^2L^2N}, \qquad (C1)$$

where $p_* = \min(3,p)$.

Clearly, there is no advantage in raising the order of the discretization error beyond $p = 3$.

Table I

Constructing an optimal multigrid cycle:
the optimal cycle index, $\gamma_{i_{opt}}$, at level i as a function of mh_i
(the mass size times the meshsize at level i)

mh_i	$\gamma_{i_{opt}}$	practical $\gamma_{i_{opt}}$
$<< 1$	6.35	6
.5	6.05	6
1	5.40	5
2	4.19	4
4	3.15	3
8	2.70	3
$>> 1$	2.52	3

Table II

Performane in measuring susceptibility:
showing α, number of random generations times the square of the obtained relative accuracy,
for the indicated values of the system size N and the cycle index γ

		N							
m	γ	4	8	16	32	64	128	256	512
0.5	2	2.4	5	7.3	9.8	12.3	14.6	17.4	
	3	2.2	3.5	4.5	5.1	5.3	5.9		
	6	2	3	3.6	3.8	3.9	4		
	7	2	2.8	3.5	3.7	3.8	3.8		
64	Tab. I	1.6	2.9	4.2	5.3	5.9	5.9	6.3	6.2
	3	1.6	2.9	4.2	5.3	6.1	6.5	6.7	6.8
	6	1.5	2.6	4.2	5.8	6.8	7.7	7.8	8.4
400	2	1.8	3.9	6.4	9.7	12.8	15.5	18.6	
	3 or Tab. I	1.6	2.9	4.3	5.6	6.8	7.5	8	
	6	1.5	2.8	4.9	7.7	11.2	16.2	20	

Table III

Computational time (in units of ε_r^{-2}) in measuring the susceptibility on a grid with N gridpoints to relative accuracy ε_r: conventional multigrid method (one measurement per cycle, as in Refs. 5, 7 or 10) vs. optimal multigrid method

		N							
m	multigrid algorithm	4	8	16	32	64	128	256	512
0.5	conventional	6.8	21.8	50.5	112.2	237	484.6		
	optimal	2	2.8	3.5	3.7	3.8	3.8		
64	conventional	4.9	17.4	52.6	115.6	236.7	485.2	992.9	2048.3
	optimal	1.6	2.9	4.2	5.3	5.9	5.9	6.3	6.2
400	conventional	4.7	16.8	50.6	112.2	235.1	501.9	960.9	
	optimal	1.6	2.9	4.3	5.6	6.8	7.5	8	

References

1. A. Brandt: *Multilevel Computations: Reviews and Recent Developments.* In Preliminary Proc. 3rd Copper Mountain conference on Multigrid Methods (April 1987). See also: *Multigrid Methods: theory applications and supercomputing*, S.F. McCormick, ed. (Marcel Dekker, 1988), pp. 35–62.

2. A. Brandt: *The Weizmann Institute Research in Multilevel Computation: 1988 Report.* In: Proc. 4th Copper Mountain Conf. on Multigrid Methods, J. Mandel et al, eds. *SIAM* (1989) pp. 13–53.

3. A. Brandt: Multigrid Methods in Lattice Field Computations. *Nuclear Physics B* (Proc. Suppl.) **26**:137-180 (1992).

4. A. Brandt, M. Galun and D. Ron: Optimal multigrid algorithms for calculating thermodynamic limits, *J. Stat. Phys.* **74**:313–348 (1994).

5. A. Brandt, D. Ron and D.J. Amit: Multi-level approaches to discrete-state and stochastic problems. *Multigrid Methods* II Proc. Cologne 1985. W. Hackbusch and U. Trottenberg, eds. (Springer Verlag, 1986), pp. 66–99.

6. M. Galun: Optimal Multigrid Algorithms for Model Problems in Statistical Mechanics. *M.Sc. Thesis,* Weizmann Institute of Science (1992).

7. J. Goodman and A.D. Sokal: Multigrid Monte Carlo methods for lattice field theories. *Phys. Rev. Lett.* **56**:1015–1018 (1986).

8. D. Kandel, E. Domany and A. Brandt: Simulations without critical slowing down—Ising and 3-State Potts models. *Phys. Rev.* **B40**:330 (1989).

9. D. Kandel, E. Domany, D. Ron, A. Brandt and E. Loh, Jr.: Simulations without critical slowing down. *Phys. Rev. Lett.* **60**:1591 (1988).

10. G. Mack and A. Pordt: Convergent perturbation expansions for euclidean quantum field theory. *Comm. Math. Phys.* **97**:267 (1985); G. Mack, in *Nonperturbative Quantum Field Theory,* G. t'Hooft et al. (Eds.), Plenum Press, NY: 309 (1988).

11. D. Ron: Development of Fast Numerical Solvers for Problems in Optimization and Statistical Mechanics. *Ph.D. Thesis,* Weizmann Institute of Science (1989).

12. R.H. Swendsen and J.S. Wang: Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* **58**:86–88 (1987).

13. U. Wolff: Collective Monte Carlo updating for spin systems. *Phys. Rev. Lett.* **62**:361–364 (1989).

Table Captions

I: Constructing an optimal multigrid cycle: the optimal cycle index, $\gamma_{i_{opt}}$, at level i as a function of mh_i (the mass size times the meshsize at level i).

II: Performane in measuring susceptibility: showing α, number of random generations times the square of the obtained relative accuracy, for the indicated values of the system size N and the cycle index γ.

III: Computational time (in units of ε_r^{-2}) in measuring the susceptibility on a grid with N gridpoints to relative accuracy ε_r: conventional multigrid method (one measurement per cycle, as in Refs. 5, 7 or 10) vs. optimal multigrid method.