CERN-TH.6172/91

July 1991

A GENERALIZED GUIDED MONTE CARLO ALGORITHM

Alan M. Horowitz

CERN Theory Division, Geneva, Switzerland

Abstract: A Hamiltonian-guided Monte Carlo algorithm for simulations of lattice field theories is presented. It is a generalization of the Hybrid Monte Carlo algorithm allowing the trajectory length to be shrunk to the step-size without losing on the speed of configuration decorrelation. Without the Monte Carlo step, this limit of the generalized algorithm reduces to the second-order Langevin equation. The latter is shown in a Gaussian model and in compact QED to be faster than the Hybrid algorithm. The performance of the two algorithms with the Monte Carlo step, HMC and L2MC, turned out to be comparable in compact QED. It is pointed out that L2MC is safer from rounding errors.

CERN-TH.6172/91

July 1991

1. Introduction

The Hybrid Monte Carlo (HMC) algorithm [1] is at present the most efficient exact method for numerical simulations of lattice field theories with dynamical fermions. It updates configurations with Hamilton's equations, occasionally randomizing the momenta (the Hybrid algorithm [2]), and tailors their distribution to the desired one $\exp(-H)$ by discarding configurations according to a global Monte Carlo decision. In principle the Hamiltonian $\bar{H}$ in Hamilton's equations is arbitrary but in practise one usually takes $\bar{H} = H$ and thus the Monte Carlo step serves to rid of the step-size errors in the Hybrid method. The HMC algorithm is a generalization of the Langevin guided global Monte Carlo (LMC) algorithm [3] which it becomes when the number of molecular dynamics steps, N_{MD} , between MC decisions is one. (The short-hand HMC will henceforth stand for "the HMC algorithm"). With $\bar{H} = H$, HMC is left with 2 parameters to be optimized, the step-size ϵ in the discretized Hamilton's equations and N_{MD} .

To date the parameter optimization is done empirically, but analytic estimates exist on how the parameters must scale with the system volume ($V \equiv L^4$) in order to keep the acceptance probability (P_{acc}) constant, at least for large volumes [4,5,6]. Specifically, for systems much larger than the correlation length, $P_{acc} \propto \text{erfc}(cN_{MD}\epsilon^3\sqrt{V})$ [6]. In HMC sweep-sweep correlations are minimized when $N_{MD} \propto 1/\epsilon$. Thus to keep P_{acc} constant requires $\epsilon \propto 1/L$. The Langevin limit $N_{MD} = 1$ yields a higher acceptance rate and better scaling behavior ($\epsilon \propto L^{-2/3}$), but this is overshadowed by the fact that the auto-correlation times, τ , are higher by a factor of order $1/\epsilon$: in HMC $\tau \propto 1/\epsilon$ while in LMC $\tau \propto 1/\epsilon^2$.

2. Generalized algorithm

The generalization proposed in this paper allows the $N_{MD} = 1$ limit to be taken while keeping the auto-correlations roughly the same. The trick is to mix the momenta with a random number with some mixing angle θ instead of replacing it entirely by the random number as in HMC. Thus the generalized algorithm with N_{MD} molecular dynamics steps between Monte Carlo decisions is, using the notation of a scalar theory with Hamiltonian $H = \frac{1}{2}\pi^2 + S(\phi)$,

$$\bar{\pi}_0 = \pi_0 \cos \theta + \xi \sin \theta \quad (1a)$$

$$\pi_{1/2} = \bar{\pi}_0 - \frac{1}{2}\epsilon S'_0 \quad (1b)$$

$$\phi_1 = \phi_0 + \epsilon\pi_{1/2}$$

$$N_{MD} - 1 \text{ full steps } \begin{cases} \pi_{n+1/2} = \pi_{n-1/2} - \epsilon S'_n \\ \phi_{n+1} = \phi_n + \pi_{n+1/2} \end{cases}$$

$$\pi_{N_{MD}} = \bar{\pi}_{N_{MD}-1/2} - \frac{1}{2}\epsilon S'_{N_{MD}}$$

where ξ is a Gaussian random number with $\langle \xi^2 \rangle = 1$. The process in eq. 1b is the leap-frog integration of Hamilton's equations. The half-steps at the beginning and end of the "trajectory" ensure the exact reversibility needed for the MC step. There is another ingredient in the generalized algorithm given below, concerning the sign of the momenta. But first I will discuss the algorithm in eq. 1 without the Monte Carlo decisions.

3. Limits H and L2

Actually only two limits of eq. 1 will be explored in this paper:

- i) $\theta = \pi/2$ and variable N_{MD} , which is the Hybrid algorithm [2], to be referred to as "H";
- ii) $N_{MD} = 1$ and variable θ , to be referred to as "L2".

Consider limit ii) for small θ . Set $\theta = \sqrt{2\epsilon\gamma}$. Then neglecting terms in θ^3 and higher, ϕ obeys

$$\Delta^2 \phi_{n+1} + \epsilon\gamma \Delta \phi_n + \epsilon^2 S'_n = \epsilon\sqrt{2\epsilon\gamma}\xi_n \quad (2)$$

where $\Delta \phi_n = \phi_n - \phi_{n-1}$. This is the second-order Langevin equation, (not to be confused with second-order integration of the first-order Langevin equation). Hence the denotation of limit ii) as L2.

It was pointed out in ref. 7, already before the Hybrid proposal [2], that L2 (more exactly the limit given in eq. 2) interpolates between the first order Langevin and molecular dynamics methods and should be optimal for some small value of γ . In a subsequent study, the optimal value of γ for the XY model on a 16^2 lattice was found to be .4, and analytic arguments and numerical support were given to the effect that the second-order equation is an order of magnitude more efficient than the Langevin equation [8]. The reason is simple. The step-size errors in L2 begin at order ϵ_{L2}^2 , while in L1 they are order ϵ_{L1} . Thus for fixed systematic error, $\epsilon_{L1} \propto \epsilon_{L2}^2$ and the ratio of auto-correlation times, $\tau_{L1}/\tau_{L2} \propto \epsilon_{L2}/\epsilon_{L1} \propto 1/\epsilon_{L2}$ (the dimensions are restored by γ).

Proofs that L2 and H qualify as legitimate algorithms and lead to $O(\epsilon^2)$ errors are given in [7,8,9].

In the literature, H is implemented with either fixed or variable numbers of MD steps between randomizations of the momenta. After each step in the variable case, denoted H_V , the momenta are replaced by random numbers with probability r or kept with probability $1-r$. Setting $r = 1/N_{MD}$ the average number of MD steps between randomization is then N_{MD} .

4. Comparison of H and L2

It turns out, as I now show, that H_V is dynamically similar to L2, given a certain relation between the parameters of each algorithm, γ (or θ) and r (or N_{MD}). In fact the relation comes immediately after considering in H_V the expected value of ϕ_2 given ϕ_1 and

ϕ_0 (assuming r of $O(\epsilon)$ and dropping a term of $O(r\epsilon^2)$):

$$\bar{\phi}_2 = 2\phi_1 - \phi_0 - r(\phi_1 - \phi_0) - \epsilon^2 S'_1. \quad (3)$$

The average here is over the noise (which is trivial) and the two possibilities, i.e., replacing by noise or keeping the momenta. Eq. 3 is identical to the corresponding average in L2 for small θ (eq. 2) if one sets

$$r = \epsilon\gamma \quad \text{or equivalently} \quad \frac{1}{2}\theta^2 = 1/N_{MD}. \quad (4)$$

Small θ in L2 means that $\sin \theta \approx \theta$ which is probably always true for optimal values of θ .

With the parameters related by eq. 4, the average dissipative force on the fields in each algorithm is the same. By the fluctuation-dissipation theorem, the average fluctuation in the fields is also the same. On explicit calculation, one sees however that they have different origins in each algorithm, precluding complete equivalence. For L2 $\langle(\phi_2 - \bar{\phi}_2)^2\rangle = 2\epsilon^3\gamma$ while in H_V one gets $r(\epsilon^2 + (\phi_1 - \phi_0)^2)$ which equals $2\epsilon^3\gamma$ only on average. Some numerical support for the relation $r = \epsilon\gamma$ already exists in the literature. In ref. 10 the optimal value of N_{MD} in H applied to the XY model on a 15^2 lattice with $\epsilon = .1$ was found to be around 25 [10] corresponding to $\gamma = .4$, which is exactly the optimal value for L2 as mentioned above.

From eq. 3 it follows that the auto-correlation function $G_H(\tau) = \langle\phi_\tau\phi_0\rangle$ for H_V obeys

$$\Delta^2 G_H + r\Delta G_H - \epsilon^2 \langle S'(\phi_\tau)\phi_0 \rangle = \delta_\tau. \quad (5)$$

The corresponding function $G_{L2}(\tau)$ in L2 satisfies the identical equation when $r = \epsilon\gamma$. When S is quadratic (e.g. a free field theory) eq. 5 becomes a closed equation for G_H and thus $G_H = G_{L2}$. The exponential auto-correlation time τ_1 for both algorithms for a mode with $S = \frac{1}{2}\omega^2\phi^2$ is

$$\left(\frac{1}{2}\gamma - \sqrt{(\frac{1}{2}\gamma)^2 - \omega^2}\right)^{-1}, \quad (6)$$

the real part of which is minimized when $\gamma = 2\omega$ giving $\tau_1 = 1/\omega$ [2,8].

Even in the Gaussian case, the auto-correlation functions of general operators are not the same for each algorithm. Consider e.g. $G_2 = \langle\phi_\tau^2\phi_0^2\rangle - \langle\phi^2\rangle^2$. In L2 $G_2 = 2G^2$ and thus the exponential auto-correlation time $\tau_2 = \tau_1/2$ and is minimized also at $\gamma = 2\omega$ where it has value $1/2\omega$. On the other hand, τ_2 in H_V , calculated using a method described in [11], is minimized at $\gamma = 2.12\omega$ where it has value $\sqrt{2}/\omega$, in agreement with the result reported in [9]. So the optimal auto-correlation time is about a factor of 3 shorter in L2, which also has the advantage of having the the same optimal γ for auto-correlations of ϕ and ϕ^2 .

5. Synthesis with Monte Carlo: L2MC

Now we come to the synthesis of L2 with a global Monte Carlo step, which I shall call L2MC. To make the process in eq. 1 reversible, the momenta must be negated at the end. The configuration is then accepted with probability

$$\min[1, \exp(H(\bar{\pi}_0, \phi_0) - H(\pi_1, \phi_1))]. \quad (7)$$

Since the Hamiltonian is invariant under $\pi \rightarrow -\pi$, we can unnegate the momenta after the Monte Carlo decision. This means that after an acceptance (with $N_{MD} = 1$) we keep (ϕ_1, π_1) while after a rejection we must take $(\phi_0, -\bar{\pi}_0)$. This completes the definition of L2MC. The proof of detailed balance is straightforward and the same as in HMC since the momentum $\bar{\pi}_0$ in eq. 1a has the Gaussian distribution, $\exp(-\frac{1}{2}\bar{\pi}_0^2)$.

If it weren't for this momenta reversal, it would be a near certain conclusion that L2MC is more efficient than HMC, since L2 is probably faster than H (based on the Gaussian example and the simulation results below), while L2MC has a higher acceptance rate than HMC. As it stands however, the question of efficiency becomes, as far as I see at present, a question for numerical experiment.

6. Numerical comparison

As a heuristic study where good statistics could be reached quickly, I compared algorithms on quenched compact QED on an 8^4 lattice. An algorithm is more efficient if in equal computer time it generates data with smaller statistical error. In a simulation with dynamical fermions equal computer time essentially means equal numbers of molecular dynamics steps since the cost of random number generation and energy evaluation is negligible compared to the conjugate gradient costs, which can be the same in each algorithm. So with an eye towards a comparison with dynamical fermions, I will measure efficiency in the quenched case also in terms of statistical errors per equal numbers of molecular dynamics steps.

I first present results of a numerical comparison of L2 and H, i.e. the algorithms without MC steps. Both were implemented with half-steps as given in eq. 1. The Hybrid algorithm was run with both fixed and random trajectory lengths. The value of β was held fixed at .94, in the disordered phase. Both the plaquette (P) and the 2×2 Wilson loop were measured. For each set of values for ϵ and γ or N_{MD} ten or more runs of 33000 molecular dynamics steps were made. Statistical errors were obtained by the binning method applied to each run, while errors in the errors were estimated from standard deviations over the runs. The integrated auto-correlation times τ_{int} were also measured. The algorithm comparisons given below are based only on the plaquette. The Wilson loop

showed on the average pretty much parallel behavior although its auto-correlation times were much *shorter*, by around a factor of 4.

For the plaquette the largest bin widths were always at least 2200 MD steps, which was over 10 times the largest measured values of τ_{int} . As carefully checked, this sufficed for reaching the asymptotic bin error, ΔP_{asympt} , defined as an average over the largest bins. The errors calculated with τ_{int} , ($\Delta P = \sqrt{2\tau_{int}\Delta P_{naive}}$), were always within 10%.

The algorithm comparisons given below are given in terms of the auto-correlation time τ_P for the plaquette defined by

$$\tau_P = \frac{1}{2}N_{MD}(\Delta P_{asympt}/\Delta P_{naive})^2$$

averaged over the runs for each set of parameters.

Before giving the comparisons I record here the static equilibrium results. With step-size .08 the value of the plaquette was .49893(5) from L2 and .49907(5) from H, while the value of the Wilson loop was .07245(7) from L2 and .07238(7) from H. With the MC step turned on, the plaquette values were .50390(5) from L2MC and .50397(5) from HMC, while the loop values were .07545(5) from L2MC and .07535(5) from HMC.

Figure 1 shows τ_P for L2 and for H_F and H_V (Hybrid algorithms with fixed and variable trajectory length respectively), as a function of the parameter r defined as $1/N_{MD}$ for H and $\frac{1}{2}\theta^2$ for L2. The step-size used was .08. According to the analysis leading to eq. 5, the algorithms are expected to behave similarly for equal values of r . Figure 1 indeed gives good support for this. The algorithms are optimal for $N_{MD} \approx 11$ for H_F and $\theta \approx .44$ for L2, which both fall in the range .09 to .1 for r . I made a less extensive investigation of H_V , but it appears to be less efficient than H_F , which in turn is less efficient than L2. (Actually when measured in computer time L2 is less efficient for this system [at least with my programs] because of the extra overhead in random numbers and energy evaluations.)

Based on considerations of ergodicity MacKenzie argued that H_V should be more efficient than H_F [12]. Since the system under study here is already at least nearly ergodic without the noise, these arguments do not seem to apply. Given ergodicity, then the superiority of H_F (at least for short distance correlation functions) might be understood simply as follows. In H_F there is an optimal N_{MD} or a narrow optimal range. Therefore in H_V one is averaging over optimal and non-optimal values of N_{MD} . This can, of course, be checked quantitatively.

Figure 2 shows τ_P for L2MC and HMC as functions of r for step-size .08. This value was around optimal, although .06 was nearly as good, while .1 was decisively worse. The performance of L2MC is remarkably similar to H_F MC (HMC with fixed N_{MD}). As was the case without the MC step H_V MC is slightly slower. Comparing with Figure 1 we see

the optimal value of r is within uncertainties the same with and without the MC step. The optimal trajectory length (ϵN_{MD}) in HMC is .8-.9, This is somewhat shorter than values found in other systems: e.g. around 2.5 in the XY model [9], and 2 in the Hubbard model [5].

The acceptance rate for L2MC was .911, independent of r (as expected), while the acceptance rates for H_FMC and H_VMC are shown in Figure 3. Such oscillatory behavior has been explained quantitatively in the context of free-field theory [11]. Because of the quick drop in acceptance rate after $N_{MD} = 1$ I did not see much advantage to be reaped from the general algorithm, eq. 1, with both N_{MD} and θ free to vary. Hence no numerical tests were made.

The ratio of τ_P in L2 to that in L2MC is .54 at the optimal θ (.44). Evidently, due to the momentum reversal one is paying a much higher price for exactness than suggested by the acceptance rate (.91). In H_FMC the corresponding ratio is .67 which is nearly the acceptance rate (.71).

7. Further discussion and summary

When dynamical fermions are included, another parameter enters the scene, namely the number of MD steps, NF_{MD} , between refreshment of the pseudo-fermions fields (see [1] for details of how fermions enter into HMC). In HMC this is usually taken to be equal to N_{MD} the number of steps between MC decisions, but it can be any integer multiple of N_{MD} . In L2MC where $N_{MD} = 1$ it would be unwise to take $NF_{MD} = 1$ for two reasons: i) for a given NF_{MD} there are $NF_{MD} + 1$ conjugate gradient inversions required, so the computer time per MD step with $NF_{MD} = 1$ is nearly double that of large NF_{MD} ; ii) with more noise, the exploration of phase space is less efficient. One possible advantage of L2MC is that NF_{MD} can be an arbitrary integer whereas in HMC it must be an integer multiple of N_{MD} , unless the pseudo-fermions are refreshed according to a probabilistic decision.

A potentially important advantage of L2MC concerns roundoff errors. As the volume (L^4) is increased or as the fermion mass (m) is decreased, the step-size must be decreased to keep the acceptance rate constant. As mentioned earlier, to do this in HMC requires $\epsilon \propto 1/L$, while it is estimated in ref. 6 that one needs $\epsilon \propto m^{\frac{3}{4}}$. If one sticks to the optimal trajectory length ϵN_{MD} of order one, then $N_{MD} \propto 1/\epsilon \propto L/m^{\frac{3}{4}}$ and at some point one may lose the struggle to protect reversibility against an avalanche of rounding errors. As the volume grows the problem is exacerbated by the need for greater precision in the fields and momenta due to the fact that the acceptance rate depends on the total energy which scales with the volume. In L2MC on the other hand where $N_{MD} = 1$ reversibility is much safer from rounding errors.

In summary I have proposed a generalized Hamiltonian-guided Monte Carlo algorithm and compared its complementary limits, H(MC) and L2(MC). The MC step aside, the limit L2 is, for small θ , the second-order Langevin equation. I showed that the average dissipative forces on the fields are the same in H and L2 provided their parameters were related by eq. 4. Nevertheless, for a Gaussian degree of freedom ϕ , the exponential autocorrelation time for ϕ^2 is around a factor of 3 lower in L2. The latter was also shown to be faster in the disordered phase of quenched compact QED, but only by about 15%. The relative speed of L2 and the fact that L2MC enjoys a higher acceptance rate than HMC must, however, be weighed against the necessary momentum reversals in L2MC.

In my exploratory study in quenched QED it turned out that HMC with fixed trajectory lengths (variable lengths were worse at least for short distance observables) required nearly the same number of MD steps as L2MC for independent data. This was probably a coincidence. Consider the dependence of the algorithms on the lattice size L . The acceptance rate stays constant in HMC when $\epsilon \propto 1/L$ and in L2MC when $\epsilon \propto 1/L^{2/3}$. It follows that if autocorrelation times are proportional to $1/\epsilon$ in each algorithm, as they are without the MC steps, then L2MC would increase in efficiency relative to HMC as L increases. So an important question is how the momentum reversals in L2MC modify the dependence of the autocorrelation times on ϵ . In any event it seems unlikely that HMC and L2MC will perform equally as the lattice grows, even if they do on small lattices. Even on moderate lattices QCD may distinguish between the algorithms more than quenched QED did. Whatever the case, the relative superiority of L2 concerning rounding errors should make it at least a useful standard of accuracy.

Acknowledgements

I thank Brian Pendleton, Tony Kennedy and Alan Sokal for valuable discussions. Financial support from CERN and the Deutsche Forschungsgemeinschaft (grant Me 567/3-2) is gratefully acknowledged. The simulations were done on the VP100 in Kaiserslautern.

References

1. S. Duane, A.D. Kennedy, B.J. Pendleton and D. Roweth, Phys. Lett. 195B (1987) 216.
2. S. Duane, Nucl. Phys. B257 [FS14] (1985) 652.
3. R.T. Scalettar, D.J. Scalapino and R.L. Sugar, Phys. Rev. B34 (1986) 7911; P.J. Rossky, J.D. Doll and H.L. Friedman, J. Chem. Phys. 69 (1978) 4628.
4. R. Gupta, G.W. Kilcup and S.R. Sharpe, Phys. Rev. D38 (1988) 1278.
5. M. Creutz, Phys. Rev. D38 (1988) 1228.
6. S. Gupta, A. Irbäck, F. Karsch and B. Petersson, Phys. Lett. 242B (1990) 437.
7. A.M. Horowitz, Phys. Lett. 156B (1985) 89.
8. A.M. Horowitz, Nucl. Phys. B280 [FS18] (1987) 510.
9. S. Duane and J.B. Kogut, Nucl. Phys. B275 (1986) 398.
10. E. Dagotto and J.B. Kogut, Phys. Rev. Lett. 58 (1987) 299.
11. A.D. Kennedy and B.J. Pendleton, Edinburgh preprint 90/498.
12. P. MacKenzie, Phys. Lett. 226B (1989) 369.

Figure captions

1. Auto-correlation times measured in MD steps for L2 (boxes), H_F (open circles) and H_V (filled circles, shifted slightly to the left to avoid overlap). The parameter r is $\frac{1}{2}\theta^2$ for L2 and $1/N_{MD}$ for H.
2. Same as Figure 1 but with MC step turned on.
3. Acceptance rate in H_F MC (open circles) and in H_V MC (filled circles). The errors are smaller than the points.

FIG 1

FIG 2

FIG 3