Adler's Overrelaxation Algorithm for Goldstone Bosons

Herbert Neuberger
Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08855
(Received 5 August 1987)

A very simple derivation of a closed-form solution to the stochastic evolution defined by Adler's overrelaxation algorithm is given for free massive and massless scalar fields on a finite lattice with periodic boundary conditions and checkerboard updating. It is argued that the results are directly relevant when critical slowing down reflects the existence of Goldstone bosons in the system.

PACS numbers: 11.15.Ha, 02.70.+d, 11.15.Ex

A well-known problem faced by Monte Carlo simulations[1] is that of critical slowing down. The essence of the problem is the following: The standard algorithms which work in practice do so in large part because only local updates are made. These updates can be shown to link into a Markov chain that converges to the desired distribution. A local change in the fields affects equally all Fourier modes. However, at equilibrium, these modes will often have very different typical sizes.

The large variation in typical sizes for Fourier modes is especially pronounced when the system has large correlations, be it because one tries to approach a continuum limit or be it because the system, even without going to infinite uv cutoff, is in a phase which contains Goldstone bosons accompanying the spontaneous breaking of a continuous symmetry. Both situations arise together in simulations of models describing strong interactions (QCD with light fermions) and systems relevant to weak interactions (the pure Higgs sector of the Weinberg-Salam model).

One approach to dealing with the problem of critical slowing down is to use Fourier-accelerated Langevin simulations.[2] The conceptual price paid is that one loses strict control over the form of the bare action. This might be a particularly heavy price to be paid when one deals with Higgs systems and the possible consequences of triviality, because there one wants to be able to claim that a complete sweep of a full physical range of a bare coupling has been carried out. Another way to fight against critical slowing down is to give up the strict local character of the updating procedure.[3]

Recently,[4] numerical experimentations with variants of a model proposed by Adler several years ago[5] have suggested that some progress can be made against critical slowing down without one's doing anything as drastically new as the multigrid method[3] or Langevin updating with Fourier acceleration.[2] Even more recently[6] Adler gave a quantitative analysis showing how his method achieves this goal.

Let me first present the basic idea of Adler's overrelaxation algorithm (AOA) and explain why, qualitatively, it may ameliorate critical slowing down. Suppose we have a system a subset of whose variables, $\Phi(x)$, enter only quadratically in the action, $\partial^n S/\partial\Phi^n(x)=0$ for $n=3,4,\ldots$. AOA prescribes that we perform sequentially local updates of these variables by the following transition probabilities at site x:

$$\xi(x)=\beta[\partial^2 S/\partial\Phi^2(x)][2\omega(2-\omega)]^{-1},\quad [\partial S/\partial\Phi(x)]\,|_{\Phi(x)=\bar{\Phi}(x;\Phi(y),y\neq x)}=0,$$
$$P(\Phi(x)\to\Phi'(x))=[\xi(x)/\pi]^{1/2}\exp\{-\xi(x)[\Phi'(x)-\omega\bar{\Phi}(x;\Phi(y),y\neq x)-(1-\omega)\Phi(x)]^2\}. \tag{1}$$

The rest of the variables are dealt with by use of other methods. Recently, Brown and Woch,[4] Creutz,[4] and Adler[6] have generalized AOA to actions which are not "multiquadratic." These developments will not concern us here. With Eqs. (1) detailed balance is satisfied and, as long as $0<\omega<2$, so are the other requirements necessary for the convergence of the Markov chain to the desired probability distribution $\exp\{-\beta S[\Phi]\}$. We see that AOA is a variant of the heat-bath algorithm, becoming identical to it for $\omega=1$. $\omega=2$ gives a deterministic, microcanonical, procedure.

For $1<\omega<2$ AOA may do better than the heat-bath algorithm for critical slowing down: The local updates are made to know something about the desire of the system to give disparate sizes to the various Fourier modes of $\Phi(x)$ by a feedback mechanism which uses the fact that at approximate equilibrium the instantaneous values of $\Phi(x)$ are more or less adequately correlated. For example, the exact conditional probability density $P(\Phi(x)=\Phi\,|\,\Phi(0)=\Phi_0)d\Phi$ of the Gibbs ensemble approximately predicts the distribution of $\Phi(x)$ even before an update; for $\omega\neq 1$ the new value is directly influenced, thus building more nonlocality into the distribution. This argument is too crude to lead to an assessment of by how much critical slowing down will be affected. Therefore, it is better to try to investigate analytically solvable cases and learn through them how and why AOA works. This was part of the objective of Ref. 6.

Adler's quantitative analysis[6] deals with a massless free field with Dirichlet or Neumann boundary conditions updated sequentially on a d-dimensional lattice by

first stepping in unit increments in the x_1 direction, then in the x_2 direction, and so forth. Because of the particular order of sweeping through the lattice, the type of boundary conditions, and the necessity of carrying out the analysis on a continuum version of the model, although one does learn how AOA deals with critical slowing down, one might still wish to see an exactly solvable case which is closer to what happens in practice.

The purpose of the present Letter is to fill in this gap; it will turn out that although we deal with a somewhat more realistic situation, the mathematical analysis is simpler than in the case studied by Adler. In Ref. 3 it was pointed out that the rate of convergence is governed by the spectral radius of the matrix defining the linear relationship between the old values of the fields and the new ones and that this radius can be easily estimated. Here we shall simply calculate all the eigenvalues of the matrix, explicitly diagonalizing it. We are still looking at a free-field theory but we use checkerboard updating, periodic boundary conditions, and work directly on the lattice. The free field can be massive without complicating the analysis. The results will be seen to confirm Adler's conclusions.[6]

The action is given by

$$S[\Phi]=\tfrac{1}{2}\sum_{\mathbf{x},\mu}[\Phi(\mathbf{x}+\hat{\boldsymbol{\mu}})-\Phi(\mathbf{x})]^2+m^2\sum_{\mathbf{x}}\Phi^2(\mathbf{x}). \tag{2}$$

Here $\mathbf{x}$ denotes a lattice site and $\boldsymbol{\mu}$ a direction; $x_\mu=0,1,2,\ldots,L-1$, $\mu=1,\ldots,d$. Periodic boundary conditions are imposed so that $\Phi(\mathbf{x})$ can be assumed to be defined for any $\mathbf{x}$ with $\Phi(\mathbf{x}+\hat{\boldsymbol{\mu}}L)=\Phi(\mathbf{x})$ $\forall\boldsymbol{\mu},\mathbf{x}$. L will be taken to be even, $L=2N$.

The Markov chain is defined by Eq. (1) and by the choice to update first all the even sites and subsequently all the odd ones, once in each lattice sweep. A site is even or odd according to whether $\|\mathbf{x}\|\equiv\sum_\mu x_\mu$ is even or odd. Since L is even there are as many even sites as odd ones. The order of the updates within the odd or even classes can be chosen at will without affecting the subsequent analysis. Sweeps through the entire lattice are indexed by $n=0,1,2,\ldots$. $n=0$ denotes the starting configuration. In equations, we have

$$-\sigma\eta_{n+1}(\mathbf{x})=(2d+m^2)[\Phi_{n+1}(\mathbf{x})-(1-\omega)\Phi_n(\mathbf{x})]-\tfrac{1}{2}\omega\sum_\mu\{[1+(-)^{\|\mathbf{x}\|}][\Phi_n(\mathbf{x}+\hat{\boldsymbol{\mu}})+\Phi_n(\mathbf{x}-\hat{\boldsymbol{\mu}})]$$
$$+[1-(-)^{\|\mathbf{x}\|}][\Phi_{n+1}(\mathbf{x}+\hat{\boldsymbol{\mu}})+\Phi_{n+1}(\mathbf{x}-\hat{\boldsymbol{\mu}})]\}. \tag{3}$$

The $\eta_n(\mathbf{x})$, $n=1,2,3,\ldots$, are normally distributed independent variables with

$$\langle\eta_n(\mathbf{x})\eta_{n'}(\mathbf{x}')\rangle=\delta_{nn'}\delta_{\mathbf{x}\mathbf{x}'} \tag{4}$$

and $\sigma=[\beta^{-1}\omega(2-\omega)(2d+m^2)]^{1/2}$.

Introducing the Fourier transforms of $\Phi(\mathbf{x})$ by

$$\Phi_n(\mathbf{x})=L^{-d}\sum_r\exp(2\pi i\mathbf{r}\cdot\mathbf{x}/L)\tilde{\Phi}_n(\mathbf{r}), \tag{5}$$

with $r_\mu=0,1,2,\ldots,L-1$, and similarly for $\eta(\mathbf{x})$, and taking the inverse transform of Eq. (3), we obtain

$$-\sigma\tilde{\eta}_{n+1}(\mathbf{r})=(2d+m^2)[\tilde{\Phi}_{n+1}(\mathbf{r})-(1-\omega)\tilde{\Phi}_n(\mathbf{r})]-\omega\sum_\mu[\cos(2\pi r_\mu/L)]$$
$$\times[\tilde{\Phi}_n(\mathbf{r})+\tilde{\Phi}_{n+1}(\mathbf{r})-\tilde{\Phi}_n(\mathbf{r}+\mathbf{T})+\tilde{\Phi}_{n+1}(\mathbf{r}+\mathbf{T})], \tag{6}$$

where $T=(N,N,N,\ldots,N)$ so that $(-1)^{\|\mathbf{x}\|}=\exp(2\pi i\mathbf{T}\cdot\mathbf{x}/L)$ and the notation assumes periodicity under $\mathbf{r}\to\mathbf{r}+\hat{\boldsymbol{\mu}}L$ $\forall\mu$. The $\tilde{\eta}(\mathbf{r})$'s are normally distributed stochastic variables: $\langle\tilde{\eta}_n(\mathbf{r})\tilde{\eta}^*_{n'}(s)\rangle=L^d\delta_{\mathbf{r}\mathbf{s}}\delta_{nn'}$.

Defining $D=2d+m^2$ and $f(\mathbf{r})=(\omega/D)\sum_\mu[\cos(2\pi r_\mu/L)]$ and replacing $\mathbf{r}$ by $\mathbf{r}+\mathbf{T}$ in Eq. (6) we obtain

$$\mathbf{A}(\mathbf{r})\begin{bmatrix}\tilde{\Phi}_{n+1}(\mathbf{r})\\ \tilde{\Phi}_{n+1}(\mathbf{r}+\mathbf{T})\end{bmatrix}=\mathbf{B}(\mathbf{r})\begin{bmatrix}\tilde{\Phi}_n(\mathbf{r})\\ \tilde{\Phi}_n(\mathbf{r}+\mathbf{T})\end{bmatrix}-\frac{\sigma}{D}\begin{bmatrix}\tilde{\eta}_{n+1}(\mathbf{r})\\ \tilde{\eta}_{n+1}(\mathbf{r}+\mathbf{T})\end{bmatrix}. \tag{7}$$

The 2×2 matrices $\mathbf{A}$ and $\mathbf{B}$ are given by

$$\mathbf{A}(\mathbf{r})=\mathbf{I}_2+f(\mathbf{r})\begin{bmatrix}-1&-1\\1&1\end{bmatrix},\quad\mathbf{B}(\mathbf{r})=(1-\omega)\mathbf{I}_2+f(\mathbf{r})\begin{bmatrix}1&-1\\1&-1\end{bmatrix}. \tag{8}$$

The general solution of Eq. (5) is

$$\begin{bmatrix}\tilde{\Phi}_{n+1}(\mathbf{r})\\ \tilde{\Phi}_{n+1}(\mathbf{r}+\mathbf{T})\end{bmatrix}=\mathbf{C}^{n+1}(\mathbf{r})\begin{bmatrix}\tilde{\Phi}_0(\mathbf{r})\\ \tilde{\Phi}_0(\mathbf{r}+\mathbf{T})\end{bmatrix}-\frac{\sigma}{D}\sum_{m=0}^{m=n}\mathbf{C}^{n-m}(\mathbf{r})\mathbf{A}^{-1}(\mathbf{r})\begin{bmatrix}\tilde{\eta}_{m+1}(\mathbf{r})\\ \tilde{\eta}_{m+1}(\mathbf{r}+\mathbf{T})\end{bmatrix}, \tag{9}$$

where $\mathbf{C}=\mathbf{A}^{-1}\mathbf{B}$.

Relaxation and autocorrelation times are governed by $\lambda_{\pm}(\mathbf{r},\omega)$, the eigenvalues of $\mathbf{C}(\mathbf{r})$:

$$\lambda_{\pm}(\mathbf{r},\omega)=(1-\omega)\{g(\mathbf{r},\omega)\pm[g^2(\mathbf{r},\omega)-1]^{1/2}\},$$
$$g(\mathbf{r},\omega)=1+2f^2(\mathbf{r})/(1-\omega). \tag{10}$$

Without overrelaxation we have $\omega=1$ and

$$\max_{\mathbf{r}}[|\lambda_+(\mathbf{r},1)|,|\lambda_-(\mathbf{r},1)|]\approx\exp(-m^2/d)\approx 1-m^2/d$$

for small m. The optimal value of ω is defined to be the one which attains

$$\min_{\omega}\{\max_{\mathbf{r}}[|\lambda_+(\mathbf{r},\omega)|,|\lambda_-(\mathbf{r},\omega)|]\}.$$

It is obtained if all the eigenvalues are made to contain an imaginary part, that $g^2(\mathbf{r};\omega)<1\ \forall\mathbf{r}$. Oscillatory behavior has indeed been observed by Creutz.[4] Its appearance for $\omega>1$ can be understood qualitatively: When ω approaches 2 the algorithm becomes microcanonical and the transition matrix becomes the measure-preserving map discussed in the Appendix of Bhanot, Creutz, and Neuberger.[7] The latter can be represented by a unitary operator in a Hilbert space.

If $g^2(\mathbf{r};\omega)<1\ \forall\mathbf{r}$ we have $\forall\mathbf{r}\ |\lambda_{\pm}(\mathbf{r},\omega)|=|\omega-1|$ and therefore an absolutely uniform convergence rate for all modes (independent of wavelength, as predicted by Adler.[6] To optimize we have to find the smallest $|\omega-1|$ which gives $g^2(\mathbf{r};\omega)<1\ \forall\mathbf{r}$. Clearly

$$g_{\min}(\mathbf{r},\omega)=g(0,\omega)=1-(2\omega^2/D^2)[d^2/(\omega-1)].$$

The optimum is obtained for $\omega=\bar{\omega}$ with $g(0,\bar{\omega})=-1$, that is $\bar{\omega}^2-1=\bar{\omega}^2[d/(2d+m^2)]^2$. Writing $\bar{\omega}=2/(1+C)$ as in Ref. 6 and $m=2d^{1/2}\tan(\theta/2)$, with $0<\theta<\pi$, we obtain

$$C=2\sin(\theta/2)[1+\sin^2(\theta/2)]^{-1},$$
$$|\lambda_{\pm}(\mathbf{r},\bar{\omega})|=\tan^4[(\pi-\theta)/4]. \tag{11}$$

For small masses $|\lambda_{\pm}(\mathbf{r},\bar{\omega})|\approx\exp(-2m/d^{1/2})$ and, in comparison with $\omega=1$, where $|\lambda_{\pm}|_{\max}\approx\exp(-m^2/d)$, the improvement factor in the relaxation time is $2d^{1/2}/m$ in complete agreement with Adler's result of $2d^{1/2}/k_1a$, when his k_1a is replaced by m. For large masses the optimal value of ω approaches 1, the heat-bath value, but $|\lambda_{\pm}(\mathbf{r},\bar{\omega})|\approx d^2m^{-1}$, which is better by a factor of 4 than the heat-bath case $\omega=1$ because there we have $|\lambda_{\max}(\mathbf{r},1)|=4d^2m^{-4}$. Therefore, even without potential critical slowing down, AOA is faster than standard heat bath.

Suppose we are not interested in relaxing the zero mode; this may be the case in a Goldstone system where the zero mode has a typical size $L^{d/2}$, whereas the nonzero, but soft, modes have an amplitude of order L. This is so because, in a finite volume, the two-point correlation

$$\langle\Phi(x)\Phi(0)\rangle\equiv L^{-d}\sum_{\mathbf{r}}\exp(2\pi i\mathbf{x}\cdot\mathbf{r}/L)/A_L(\mathbf{r})$$

must reproduce a nonzero expectation value for Φ at infinite L [$A(\mathbf{r})$ is the inverse full propagator in momentum space]. For small $\mathbf{r}$ we expect

$$A_L(\mathbf{r})\approx 4\sum_{\mu}\sin^2(\pi r_{\mu}/L)+m_{\text{eff}}^2(L)$$

and therefore $m_{\text{eff}}^2(L)\approx L^{-d}/\langle\Phi\rangle^2$ reflecting the rapid vanishing of the mass for $L\to\infty$ as required by Goldstone's theorem. The use of this approximation for $A(\mathbf{r})$ leads to order-or-magnitude estimates for the amplitudes of the appropriate modes as stated above. We therefore may wish to just forget about trying to effectively average over the zero mode even at finite, but large, L, or alternatively, we can deal with the zero mode by another method.

For the other modes the massless version of Eq. (2) is applicable. The fact that $\exp(-\beta S)$ is not integrable in the massless case simply means that, in the free model, the zero mode will not relax no matter what we do with ω. We now require $g^2(\mathbf{r},\omega)<1\ \forall\mathbf{r}\neq 0,\mathbf{T}$ obtaining

$$C=2[\rho_L(1-\rho_L)]^{1/2},\quad \rho_L=d^{-1}\sin^2(\pi/L),$$
$$|\lambda_{\pm}(\mathbf{r},\bar{\omega})|=(1-C)/(1+C)\quad\forall\mathbf{r}\neq 0,\mathbf{T}, \tag{12}$$
$$|\lambda_{\pm}(0,\bar{\omega})|=|\lambda_{\pm}(\mathbf{T},\bar{\omega})|=\begin{cases}[(1-C)/(1+C)]^2,\\ 1.\end{cases}$$

We see that, with the exception of one mode, we have uniform boundedness with a relaxation time growing as L for large L. A heat-bath algorithm would lead to an L^2 growth instead.

A physical situation to which this analysis is particularly relevant is the simulation of the pure Higgs sector of the Weinberg-Salam model which is important for the study of the effects of triviality on the Higgs mass.[8] The true distributions of the modes representing the Weinberg-Salam analogs of soft pions must be well approximated by *weakly* interacting plane waves (i.e., almost free waves); otherwise soft-pion theorems could not be right. So the averaging over the degrees of freedom which best represent the pions in the continuum and infinite volume limits should benefit from AOA and the analysis presented in the present Letter should be of direct relevance.

This research was supported in part by the U.S. Department of Energy under the Outstanding Junior Investigator Program, Contract No. DE-FG05-86-ER40265. Comments by S. Adler are greatly appreciated. He also found and corrected an algebraic error in the original preprint.

[1]M. Creutz, Brookhaven National Laboratory Report No.

BNL-39747, 1987 (to be published).

[2]G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. D **32**, 2736 (1985).

[3]J. Goodman and A. D. Sokal, Phys. Rev. Lett. **56**, 1015 (1986).

[4]M. Creutz, Phys. Rev. D **36**, 515 (1987); F. R. Brown and T. J. Woch, Phys. Rev. Lett. **58**, 2394 (1987).

[5]S. L. Adler, Phys. Rev. D **23**, 2901 (1981); C. Whitmer, Phys. Rev. D **29**, 306 (1984).

[6]S. L. Adler, Phys. Rev. D (to be published).

[7]G. Bhanot, M. Creutz, and H. Neuberger, Nucl. Phys. **B235 [FS11]**, 417 (1984).

[8]R. Dashen and H. Neuberger, Phys. Rev. Lett. **50**, 1897 (1983).