

Optimization of trial wave functions for Hamiltonian lattice models

Matteo Beccaria

*Dipartimento di Fisica, Università di Lecce, Via Arnesano, 73100 Lecce, Italy
and Istituto Nazionale di Fisica Nucleare, Lecce, Italy*

(Received 21 December 1999; published 1 May 2000)

We propose a new Monte Carlo algorithm for the numerical study of general lattice models in Hamiltonian form. The algorithm is based on an initial ansatz for the ground state wave function depending on a set of free parameters which are dynamically optimized. The method is discussed in detail and the results are reported from explicit simulations of a U(1) N chain, a degenerate limit of U(1) lattice gauge theory in 2+1 dimensions.

PACS number(s): 12.38.Gc, 11.15.Ha

The lattice formulation of quantum field theory is a rigorous theoretical framework for the study of nonperturbative phenomena and its applications to phenomenologically relevant models, such as QCD, are expected to play a major role in the near future. In most numerical studies, the lattice version of a given model is built in the Lagrangian formulation [1]. However, for certain “static” problems, such as hadron spectroscopy, the Hamiltonian approach [2] is more suited because time is kept as continuous as possible and the spectrum is deformed in a minimal way. Actually, even the most accurate Lagrangian calculations [3] evidence the advantages of using anisotropic lattices with smaller spacings in the temporal direction, a typical feature of the Hamiltonian approach.

Another advantage of the Hamiltonian formulation is the possibility of easily exploiting any knowledge about the ground state wave function Ψ_0 in order to improve the performance of numerical calculations. In particular, in this paper, we are concerned with the popular and all purpose Green function Monte Carlo algorithm (GFMC) [4] where an approximate version of Ψ_0 , the trial wave function Ψ_T , allows us to speed up the algorithm convergence by means of importance sampling techniques [5].

In a realistic problem, Ψ_T depends on several free parameters $\mathbf{a}=(a_1, \dots, a_N)$ that parametrize interaction terms responsible for correlations expected to show up in the ground state. The development of methods for the optimization of the free parameters \mathbf{a} is a key problem and an active field of research [6]. The simplest approach is to perform a variational calculation of the mean energy $E_V(\mathbf{a})=\langle\Psi_T(\mathbf{a})|H|\Psi_T(\mathbf{a})\rangle$ where H is the Hamiltonian. The quantity $E_V(\mathbf{a})$ is then minimized with respect to \mathbf{a} . Since each evaluation of $E_V(\mathbf{a})$ requires a separate Monte Carlo calculation, this is an expensive procedure for which special tricks have been devised (e.g., correlated sampling [7]). A particularly interesting approach is that of Ref. [8] where the authors propose an algorithm for the automatic dynamical optimization of the free *variational* parameters \mathbf{a} .

In this paper, we abandon the variational approximation and consider instead a full Monte Carlo calculation for which we propose a new strategy to optimize \mathbf{a} . We describe an adaptive algorithm which converges online to an optimal set of parameters \mathbf{a}^* which minimizes the statistical error of the full (*not* variational) Monte Carlo simulation. The nonadap-

tive core of the algorithm is a GFMC algorithm with importance sampling and stochastic reconfiguration [9].

Let us consider a quantum mechanical point particle with Hamiltonian

$$H=H_0+V(\mathbf{q}), \quad H_0=\frac{1}{2}\mathbf{p}^2, \quad (1)$$

where $\mathbf{q}=(q_1, \dots, q_d)$ is the position in \mathbf{R}^d and $\mathbf{p}=(p_1, \dots, p_d)$ is the associated momentum satisfying the canonical commutation rules $[q_i, q_j]=0=[p_i, p_j]$ and $[q_i, p_j]=i\delta_{ij}$. Let $\Psi_T(\mathbf{q}, \mathbf{a})=\exp F(\mathbf{q}, \mathbf{a})$ be a positive approximation of the ground state wave function depending on some parameters $\mathbf{a}=(a_1, \dots, a_N)$. We perform a unitary transformation on H based on Ψ_T and build the new isospectral Hamiltonian

$$\tilde{H}=\Psi_T H_0 \frac{1}{\Psi_T} + V(\mathbf{q}). \quad (2)$$

The Feynman propagator $\tilde{D}=\langle\mathbf{q}''|\exp(-t\tilde{H})|\mathbf{q}'\rangle$ associated with \tilde{H} can be written

$$\begin{aligned} \tilde{D}(\mathbf{q}'', \mathbf{q}', \mathbf{a}, \varepsilon) &= \frac{1}{(2\pi\varepsilon)^{d/2}} \frac{1}{\sqrt{1+\varepsilon\nabla^2 F}} \\ &\times \exp\{-x_i[2\varepsilon(\delta_{ij}+\varepsilon\partial_{ij}^2 F)]^{-1}x_j-\varepsilon\tilde{V}\} + \mathcal{O}(\varepsilon^2), \end{aligned} \quad (3)$$

$$x_i=q_i''-q_i'-\varepsilon\partial_i F,$$

where

$$\tilde{V}(\mathbf{q}, \mathbf{a})=V(\mathbf{q})-\frac{1}{2}[\nabla_{\mathbf{q}}F(\mathbf{q}, \mathbf{a})]^2-\frac{1}{2}\nabla_{\mathbf{q}}^2 F(\mathbf{q}, \mathbf{a}). \quad (4)$$

By standard arguments, we conclude that \tilde{D} may be expressed in terms of a weighted average

$$\tilde{D}(\mathbf{q}'', \mathbf{q}', \mathbf{a}, t)=\int_{\mathbf{q}(0)=\mathbf{q}', \mathbf{q}(t)=\mathbf{q}''} \tilde{D}(\mathbf{q}(t)) e^{-\int_0^t d\tau \tilde{V}[\mathbf{q}(\tau), \mathbf{a}]}, \quad (5)$$

where \tilde{D} is the formal limiting measure (when $\varepsilon \rightarrow 0$) defined on the stochastic paths generated according to the Langevin equation

$$\mathbf{q}_{n+1,i} = \mathbf{q}_{n,i} + \varepsilon \partial_i F(\mathbf{q}_n, \mathbf{a}) + \sqrt{\varepsilon} \left(\mathbf{z}_{n,i} + \frac{1}{2} \varepsilon \partial_{ij}^2 F(\mathbf{q}_n, \mathbf{a}) \mathbf{z}_{n,j} \right), \quad (6)$$

where $\mathbf{q}_{n,i}$ denotes the i th component of \mathbf{q}_n and \mathbf{z}_n are uncorrelated Gaussian random points in \mathbf{R}^d with unit variance. The GFMC algorithm is expressed in a concise and formal way by Eq. (5), but the actual calculation of the weighted averages must keep under control the variance of the path-dependent weights $\exp\{-\int_0^t d\tau \tilde{V}[\mathbf{q}(\tau), \mathbf{a}]\}$ which explodes as $t \rightarrow +\infty$. An efficient technique to solve this problem is stochastic reconfiguration [9] which implements a kill and branch selection on the paths with the desirable feature of dealing always with a fixed size population of walkers. To this aim, a finite collection of K walkers, an ensemble, is introduced

$$\mathcal{E} = \{[\mathbf{q}^{(n)}(t), \omega^{(n)}(t)]\}_{1 \leq n \leq K}, \quad (7)$$

where the weights $\omega^{(n)}$ are defined by

$$\omega^{(n)}(t) = \exp - \int_0^t d\tau \tilde{V}[\mathbf{q}^{(n)}(\tau), \mathbf{a}]. \quad (8)$$

The variance of the weights over the ensemble \mathcal{E} is

$$W(t) = \text{Var} \omega(t) = \frac{1}{K} \sum_{k=1}^K [\omega^{(k)}(t)]^2 - \left(\frac{1}{K} \sum_{k=1}^K \omega^{(k)}(t) \right)^2, \quad (9)$$

and the average of a function $f(\mathbf{q})$ over \mathcal{E} is

$$\langle f \rangle_{\mathcal{E}} = \frac{1}{K} \sum_{k=1}^K f(\mathbf{q}^{(k)}) \omega^{(k)}. \quad (10)$$

When $W(t)$ becomes too large, \mathcal{E} is transformed into a new ensemble \mathcal{E}' with zero variance W and the same averages, at least in the $K \rightarrow \infty$ limit. Simulations can be extended to arbitrarily large times with the drawback of a systematic error vanishing with K and an extrapolation to $K \rightarrow \infty$ is required. More details on this procedure can be found in Ref. [9].

Statistical errors in such a Monte Carlo simulation are related to the fluctuations of $\tilde{V}[\mathbf{q}(\tau), \mathbf{a}]$. To see this, let us consider for instance the calculation of the ground state energy E_0 . A simple algorithm estimates the limit $E_0 = \lim_{t \rightarrow +\infty} E(t, \mathbf{q}', \mathbf{a})$ where, for arbitrary \mathbf{q}' and \mathbf{a} , we define

$$E(t, \mathbf{q}', \mathbf{a}) = - \frac{d}{dt} \log \int_{-\infty}^{\infty} \tilde{D}(\mathbf{q}'', \mathbf{q}', \mathbf{a}, t) d\mathbf{q}'' = \langle \tilde{V}[\mathbf{q}(t), \mathbf{a}] \rangle, \quad (11)$$

$\langle \cdot \rangle$ is the average over the weighted Langevin trajectories as in Eq. (5). If Ψ_T is an exact eigenstate of H with eigenvalue E , then we obtain $\tilde{V} \equiv E$ and the above formula gives $E(t) = E$ with zero variance, namely, no statistical error.

In the more general case of nonoptimal $\Psi_T \neq \Psi_0$ a simulation performed with a sufficiently small ε and a population of K walkers will provide after S Monte Carlo steps only an approximate estimator $\hat{E}_0(S, K, \mathbf{a})$ of E_0 , that is a random variable with the asymptotic properties

$$\langle \hat{E}_0(S, K, \mathbf{a}) \rangle = E_0 + \frac{c_1(\mathbf{a})}{K^\alpha} + o(K^{-\alpha}), \quad \alpha > 0, \quad (12)$$

$$\text{Var} \hat{E}_0(S, K, \mathbf{a}) = \frac{c_2(K, \mathbf{a})}{\sqrt{S}},$$

where averages are over Monte Carlo realizations. The average of \hat{E}_0 extrapolated at $K \rightarrow \infty$ is exact and independent on the trial parameters \mathbf{a} whereas $c_2(K, \mathbf{a})$ is in general strongly dependent on them and is expected to vanish when $K \rightarrow \infty$.

If the family $\Psi_T(\mathbf{q}, \mathbf{a})$ includes the exact ground state at the special point $\mathbf{a} = \mathbf{a}^*$, then we know that $c_1(\mathbf{a}^*) = c_2(K, \mathbf{a}^*) = 0$; in a less optimal situation, motivated by Eq. [10], we seek a minimum of c_2 .

To establish an adaptive algorithm, we let $\mathbf{a} \rightarrow \{\mathbf{a}_n\}$ be a dynamical parameter of the simulation and propose to update it together with the ensemble according to the equations

$$\begin{aligned} \mathbf{q}_{n+1,i} &= \mathbf{q}_{n,i} + \varepsilon \partial_i F(\mathbf{q}_n, \mathbf{a}_n) \\ &+ \sqrt{\varepsilon} \left(\mathbf{z}_{n,i} + \frac{1}{2} \varepsilon \partial_{ij}^2 F(\mathbf{q}_n, \mathbf{a}_n) \mathbf{z}_{n,j} \right), \\ \mathbf{a}_{n+1} &= \mathbf{a}_n - \eta \nabla_{\mathbf{a}} \mathcal{F}_n(\mathbf{a}_n), \\ \mathcal{F}_n(\mathbf{a}_n) &= \text{Var}_{\mathcal{E}} \tilde{V}(\mathbf{q}_n, \mathbf{a}_n), \end{aligned} \quad (13)$$

where η is a constant parameter and \mathbf{a}_n is the value of \mathbf{a} at the n th update. In other words, we implement a local minimization of the weight variance as a driving mechanism for the free parameters.

The coupled set of equations (13) for the evolution of \mathbf{a} and the random walkers is nonlinear and discrete. The optimization of \mathbf{a} and the GFMC are thus linked together. This procedure can be successfully checked in trivial quantum mechanical examples. Here we discuss a nontrivial application to test convergence and stability. We study the model introduced in Ref. [11], the U(1) N chain, a degenerate limit of U(1) gauge theory in 2+1 dimensions. Following the notation of Ref. [11] (see also Ref. [12] for other applications of GFMC to lattice gauge theory), the Hamiltonian of the model is

$$H = \sum_{p=1}^L \left[\left(\sum_{l_p=1}^3 - \frac{1}{2\beta} \frac{\partial^2}{\partial \theta_{l_p}^2} \right) + \beta(1 - \cos \phi_p) \right], \quad (14)$$

where β is the coupling constant, L is the spatial lattice size, θ_{l_p} are link phases around the p th plaquette and $\phi_p = \theta_{1,p} + \theta_{2,p+1} - \theta_{3,p} - \theta_{2,p}$ is the plaquette gauge invariant angle. An accurate variational estimate of the ground state energy

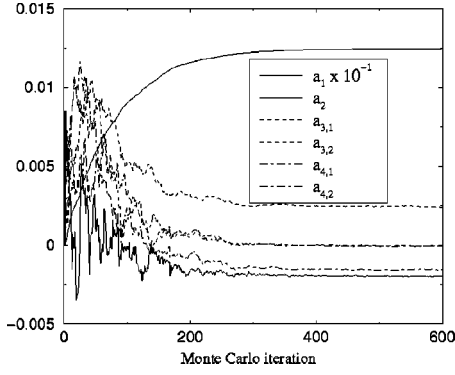


FIG. 1. Monte Carlo time history of the free parameters \mathbf{a} for the run with $\beta=0.5$ and $K=30$.

per plaquette E_0/L is obtained with the gauge invariant ansatz $\Psi_0(\phi_1, \dots, \phi_L) = \exp(\lambda \sum_{p=1}^L \cos \phi_p)$ leading to

$$\frac{E_0^{(\text{var})}(\beta)}{L} = \beta + \frac{1}{\beta} \min_{\lambda} \left[(\lambda - \beta^2) \frac{I_1(2\lambda)}{I_0(2\lambda)} \right], \quad (15)$$

where I_n is the n th modified Bessel function.

We choose a rather general gauge invariant trial wave function $\Psi_T = \exp F$ with interplaquette correlations of the following form:

$$F = \sum_{p=1}^L \left\{ a_1 \cos \phi_p + a_2 \cos 2\phi_p + \sum_{k=1}^2 [a_{3,k} \cos(\phi_p + \phi_{p+k}) + a_{4,k} \cos(\phi_p - \phi_{p+k})] \right\}, \quad (16)$$

which depends on six free parameters $a_1, a_2, a_{3,1}, a_{3,2}, a_{4,1}, a_{4,2}$. The particular case $a_{3,2} = a_{4,2} = 0$ with only on-site and nearest-neighbor terms is discussed in Ref. [11] where the $\mathcal{O}(\beta^4)$ perturbative optimal values are given: $a_1^{\text{pert}} = \beta^2/2$, $a_2^{\text{pert}} = -\beta^4/32$, $a_{3,1}^{\text{pert}} = \beta^4/24$, and $a_{4,1}^{\text{pert}} = -\beta^4/40$; the ground state energy is measured by using only a_1 which is optimized by trials and errors. In this paper, we shall discuss the nonperturbative behavior of the full six parameters at the equilibrium point reached automatically by the algorithm.

We perform simulations with ensembles of $K=10, 30$, and 50 walkers on a system with spatial lattice size $L=8$ and time step $\varepsilon=0.015$. In the Lagrangian formulation this would correspond to a simulation on a lattice with a very

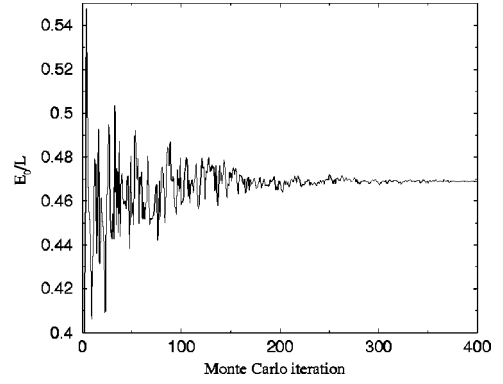


FIG. 2. Monte Carlo time history of the energy measurements during the first steps of the run with $\beta=0.5$ and $K=30$.

large extension in the temporal direction. The coupling β is varied through the values 0.5, 1.0, 1.5, 2.0, and 2.5. We execute stochastic reconfiguration of the ensemble each $r(\beta)$ temporal steps with $r(\beta)$ ranging between a maximum 40 at $\beta=0.5$ and a minimum 20 at $\beta=2.5$. The learning parameter is $\eta=0.0005$ and all the free parameters are zero at the first iteration. The number of Monte Carlo iterations is around 5×10^4 depending slightly on β .

In Fig. 1 we show the Monte Carlo time history of the free parameters \mathbf{a} for the run with $\beta=0.5$ and $K=30$. After about 300 iterations, the parameters reach an optimum value around which they oscillate with small fluctuations. If these fluctuations were too noisy, they could be removed by stochastic gradient approximation techniques, namely, by letting η be a time-dependent positive sequence $\{\eta_n\}_{n \geq 0}$ vanishing with $n \rightarrow \infty$ under the constraints $\sum_n \eta = \infty$ and $\sum_n \eta^2 < \infty$ [8]. The corresponding energy time history is shown in Fig. 2 where the fluctuations of the local measurements are shown to decrease very rapidly.

The optimal values of the six parameters as functions of β at the largest K used are shown in Table I. The leading perturbative expansion of \mathbf{a} can be checked to be correct only at the smallest β and overestimates \mathbf{a}^* at larger β . The next-to-nearest-neighbor terms are rather small being at $\beta=2.5$ roughly 1% of the nearest-neighbor ones. The qualitative picture discussed in Ref. [11] is confirmed with dominating plaquette anticorrelation in the ground state and next-to-nearest neighbor effects below the percent level. Here, we stress again, the determination of the optimal set of parameters \mathbf{a}^* is completely automatic.

About the ground state energy, we show in Table II the

TABLE I. Optimal value of the six free parameters $a_1, a_2, a_{3,1}, a_{3,2}, a_{4,1}, a_{4,2}$ and corresponding minimum standard deviation σ_E of the ground state energy estimator. We show the numerical results obtained with 50 walkers.

| β | a_1 | a_2 | $a_{3,1}$ | $a_{3,2}$ | $a_{4,1}$ | $a_{4,2}$ | σ_E |
|---------|--------------|---------------|--------------|---------------|---------------|---------------|------------|
| 0.5 | 0.1245271(7) | -0.0019590(2) | 0.0024544(4) | -0.0000374(3) | -0.0015535(3) | -0.0000609(3) | 0.0002 |
| 1.0 | 0.46287(4) | -0.022487(6) | 0.02597(1) | 0.00074(1) | -0.015262(8) | -0.00108(1) | 0.003 |
| 1.5 | 0.8327(2) | -0.05406(3) | 0.06255(5) | 0.00319(4) | -0.02999(4) | -0.00414(4) | 0.009 |
| 2.0 | 1.1541(6) | -0.07794(5) | 0.0997(1) | 0.00618(7) | -0.03921(7) | -0.00866(6) | 0.008 |
| 2.5 | 1.4447(5) | -0.09663(7) | 0.1370(1) | 0.01065(9) | -0.0459(1) | -0.0128(1) | 0.008 |

TABLE II. Comparison among the different estimates (the proposed method, standard Monte Carlo and variational) of the ground state energy.

| β | $E_{MC,K=10}^{\text{adaptive}}$ | $E_{MC,K=30}^{\text{adaptive}}$ | $E_{MC,K=50}^{\text{adaptive}}$ | $E_{MC, \text{Barnes et al.}}$ | $E_{\text{variational}}$ |
|---------|---------------------------------|---------------------------------|---------------------------------|--------------------------------|--------------------------|
| 0.5 | 0.468930(2) | 0.468932(1) | 0.4689327(8) | 0.4690(1) | 0.4690 |
| 1.0 | 0.77032(3) | 0.77020(2) | 0.77024(2) | 0.7697(2) | 0.7746 |
| 1.5 | 0.88226(8) | 0.88209(5) | 0.88188(4) | 0.8823(8) | 0.9005 |
| 2.0 | 0.91718(9) | 0.91694(5) | 0.91705(4) | 0.916(1) | 0.9435 |
| 2.5 | 0.93245(8) | 0.93242(5) | 0.93230(3) | 0.929(2) | 0.9594 |

estimates computed with three values of K to show the very small systematic residual error associated with the finite size walker population. We also show the Monte Carlo results from Ref. [11] with which we agree within errors as well as the variational bound. It is remarkable that a good estimate of the ground state energy is obtained with the very small number of walkers $K=10$. This can be interpreted as a signal that for this admittedly simple model the proposed six parameters wave function is rather accurate.

To conclude, let us remark that the above method is rather general and is applicable to dynamically optimize the free parameters of a many body trial wave function for any model that can be studied by GFMC with stochastic reconfiguration. This includes for instance pure gauge SU(2) and SU(3) lattice gauge theory in any dimension.

-
- [1] K. Wilson, Phys. Rev. D **10**, 2445 (1974).
[2] J. B. Kogut and L. I. Susskind, Phys. Rev. D **11**, 395 (1975); J. B. Kogut, Rev. Mod. Phys. **51**, 659 (1979).
[3] C. J. Morningstar and M. Peardon, Phys. Rev. D **56**, 4043 (1997); C. J. Morningstar and M. Peardon, *ibid.* **60**, 034509 (1999).
[4] W. von der Linden, Phys. Rep. **220**, 53 (1992).
[5] D. M. Ceperley and M. H. Kalos, in *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer-Verlag, Heidelberg, 1992).
[6] E. Koch, O. Gunnarsson, and R. M. Martin, Phys. Rev. B **59**, 15 632 (1999).
[7] D. Ceperley, G. Chester, and M. Kalos, Phys. Rev. B **16**, 3081 (1977).
[8] A. Harju, B. Barbiellini, S. Siljamäki, and R. M. Nieminen, Phys. Rev. Lett. **79**, 1173 (1997).
[9] M. C. Buonaura and S. Sorella, Phys. Rev. B **57**, 11 446 (1998); J. A. Hetherington, Phys. Rev. A **30**, 2713 (1984).
[10] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988); P. R. C. Kent, R. J. Needs, and G. Rajagopal, Phys. Rev. B **59**, 12 344 (1999).
[11] T. Barnes and D. Kotchan, Phys. Rev. D **35**, 1947 (1987).
[12] C. Best and A. Schäfer, Nucl. Phys. B (Proc. Suppl.) **42**, 216 (1995); N. J. Davidson and R. F. Bishop, *ibid.* **42**, 817 (1995); S. A. Chin, C. Long, and D. Robson, Phys. Rev. D **37**, 3001 (1988); C. Long, D. Robson, and S. A. Chin, *ibid.* **37**, 3006 (1988).