Energy and Variance Optimization of Many-Body Wave Functions

C. J. Umrigar
Theory Center and Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA

Claudia Filippi
Instituut Lorentz, Universiteit Leiden, Niels Bohrweg 2, Leiden, NL-2333 CA, The Netherlands

(Received 22 December 2004; published 21 April 2005)

We present a simple, robust, and efficient method for varying the parameters in a many-body wave function to optimize the expectation value of the energy. The effectiveness of the method is demonstrated by optimizing the parameters in flexible Jastrow factors that include 3-body electron-electron-nucleus correlation terms for the NO_2 and decapentaene ($C_{10}H_{12}$) molecules. The basic idea is to add terms to the straightforward expression for the Hessian of the energy that have zero expectation value, but that cancel much of the statistical fluctuations for a finite Monte Carlo sample. The method is compared to what is currently the most popular method for optimizing many-body wave functions, namely, minimization of the variance of the local energy. The most efficient wave function is obtained by optimizing a linear combination of the energy and the variance.

DOI: 10.1103/PhysRevLett.94.150201 PACS numbers: 02.70.Ss, 31.10.+z, 31.25.Qm

Quantum Monte Carlo methods [1,2] are some of the most accurate and efficient methods for treating many-body systems. The success of these methods is in large part due to the flexibility in the form of the trial wave functions that results from doing integrals by Monte Carlo calculations. Since the capability to efficiently optimize the parameters in trial wave functions is crucial to the success of both the variational Monte Carlo (VMC) and the diffusion Monte Carlo (DMC) methods, a lot of effort has been put into inventing better optimization methods.

The variance minimization [3,4] method has become the most frequently used method for optimizing many-body wave functions because it is far more efficient than *straightforward* energy minimization. The reason is that, for a sufficiently flexible variational wave function, it is possible to lower the energy on the finite set of Monte Carlo (MC) configurations on which the optimization is performed, while in fact raising the true expectation value of the energy. On the other hand, if the variance of the local energy is minimized, each term in the sum over MC configurations is bounded from below by zero and the problem is far less severe [4].

Nevertheless, in recent years several clever methods have been invented that optimize the energy rather than the variance [5–14]. The motivations for this are fourfold. First, one typically seeks the lowest energy in either a variational or a diffusion Monte Carlo calculation, rather than the lowest variance. Second, although the variance minimization method has been used to optimize both the Jastrow coefficients and the determinantal coefficients (the coefficients in front of the determinants, and in the expansion of the orbitals in a basis, and the exponents in the Slater or Gaussian basis functions) [4,15,16], it takes many iterations to optimize the latter and the optimization can get stuck in multiple local minima. So, most authors have used variance minimization for the Jastrow parameters only where these problems are absent. Third, for a given form of the trial wave function, energy-minimized wave functions on average yield more accurate values of other expectation values than variance-minimized wave functions do [17]. Fourth, the Hellmann-Feynman theorem, combined with a variance reduction technique [18], can be used with energy-minimized wave functions to compute forces on nuclei.

The various energy minimization methods are successful in varying degrees. The generalized eigenvalue method of Nightingale and Melik-Alaverdian [8] is the most efficient choice for optimizing linear parameters, but for nonlinear parameters they use variance minimization. The effective fluctuation potential method [9–13] is the most successful method for nonlinear parameters and has been applied to optimizing the orbitals [10,13] and the linear coefficients in a multideterminantal wave function [11,13], and, has been extended to excited states [13]. It is not straightforward to use this method to optimize Jastrow factors, but Prendergast *et al.* [12] have formulated a version for periodic systems and have optimized an impressively large number of parameters. However, the method is complex and needs to be reformulated for finite systems. The stochastic reconfiguration method [14] is related to the effective fluctuation potential method and is simpler but less efficient [13]. The Newton method as implemented in Ref. [7] is the most straightforward method but is inefficient and unstable. The earlier methods [5,6] have been applied only to very small systems or very few parameters.

The purpose of this Letter is to show that it is possible to devise an energy minimization method that is simple, robust, and efficient. The method can be applied to optimizing many-body wave functions, for both continuum and lattice problems. The trick to doing this is to modify the straightforward expression for the Hessian of the energy by adding a term that has zero expectation value for an infinite MC sample, but that is nonzero and cancels much of the statistical fluctuations for a finite MC sample. First, we review the variance minimization method.

Variance minimization.—The parameters c_i in a real-valued trial wave function ψ are varied to minimize the variance of the local energy,

$$\sigma^2 = \frac{\int d^{3N}R\, \psi^2 (E_{\rm L} - \bar{E})^2}{\int d^{3N}R\, \psi^2} = \langle (E_{\rm L} - \bar{E})^2 \rangle, \qquad (1)$$

where $E_{\rm L} = H\psi/\psi$ is the local energy, $\langle \cdot \rangle$ denotes a ψ^2-weighted expectation value, and $\bar{E} = \langle E_{\rm L} \rangle$ is the expectation value of the energy. The derivative of σ^2 with respect to the ith parameter, c_i, is given by

$$(\sigma^2)_i = 2\bigg[\langle E_{{\rm L},i}(E_{\rm L} - \bar{E}) \rangle + \bigg\langle \frac{\psi_i}{\psi} E_{\rm L}^2 \bigg\rangle - \bigg\langle \frac{\psi_i}{\psi} \bigg\rangle \langle E_{\rm L}^2 \rangle - 2\bar{E} \bigg\langle \frac{\psi_i}{\psi}(E_{\rm L} - \bar{E}) \bigg\rangle \bigg], \qquad (2)$$

where subscript i denotes derivative with respect to c_i.

Since the variance minimization method can be viewed as a fit of the local energy on a fixed set of Monte Carlo configurations [4], an alternative expression follows from ignoring the change of the wave function:

$$(\sigma^2)_i = 2\langle E_{{\rm L},i}(E_{\rm L} - \bar{E}) \rangle = 2\langle (E_{{\rm L},i} - \bar{E}_i)(E_{\rm L} - \bar{E}) \rangle. \quad (3)$$

Then the usual (positive definite) Levenberg-Marquardt approximation [19] to the Hessian matrix is

$$(\sigma^2)_{ij} = 2\langle (E_{{\rm L},i} - \bar{E}_i)(E_{{\rm L},j} - \bar{E}_j) \rangle. \qquad (4)$$

Energy minimization.—The elements of the gradient are

$$\bar{E}_i = \bigg\langle \frac{\psi_i}{\psi} E_{\rm L} + \frac{H\psi_i}{\psi} - 2\bar{E}\frac{\psi_i}{\psi} \bigg\rangle, \qquad (5)$$

$$= 2\bigg\langle \frac{\psi_i}{\psi}(E_{\rm L} - \bar{E}) \bigg\rangle \quad \text{(by Hermiticity)}. \qquad (6)$$

We note that the step from Eq. (5) to Eq. (6) was made not just in the interest of simplicity, but more importantly because the expression in Eq. (6) has zero fluctuations in the limit that ψ is an exact eigenstate, whereas the expression in Eq. (5) has large fluctuations.

Taking the derivative of Eq. (6), the Hessian is

$$\bar{E}_{ij} = 2\bigg[\bigg\langle \bigg(\frac{\psi_{ij}}{\psi} + \frac{\psi_i \psi_j}{\psi^2} \bigg)(E_{\rm L} - \bar{E}) \bigg\rangle - \bigg\langle \frac{\psi_i}{\psi} \bigg\rangle \bar{E}_j - \bigg\langle \frac{\psi_j}{\psi} \bigg\rangle \bar{E}_i + \bigg\langle \frac{\psi_i}{\psi} E_{{\rm L},j} \bigg\rangle \bigg]. \qquad (7)$$

This is nothing more than a rearrangement of terms in the Hessian in Ref. [7]. We now make two changes to the above expression. First, we note that the last term is not symmetric in i and j when approximated by a finite sample, whereas the true Hessian of course is symmetric. So, we symmetrize it. This change does not significantly alter the efficiency of the method, but it does have the advantage that the eigensystem is real. Next, we note that Eq. (6) and all except the last term in Eq. (7) are in the form of a covariance, $(\langle ab \rangle - \langle a \rangle \langle b \rangle)$. The fluctuations of $\langle ab \rangle - \langle a \rangle \langle b \rangle$ are in most cases smaller than those of $\langle ab \rangle$, (e.g., if a and b are weakly correlated), and, they are much smaller if $\sqrt{\langle a^2 \rangle - \langle a \rangle^2} \ll |\langle a \rangle|$ and a is not strongly correlated with $1/b$. Since the Hamiltonian is Hermitian it follows, as also noted in Ref. [7], that $\langle E_{{\rm L},j} \rangle = 0$. Hence, an alternative symmetric expression [20] for the Hessian, written entirely in terms of covariances, is

$$\bar{E}_{ij} = 2\bigg[\bigg\langle \bigg(\frac{\psi_{ij}}{\psi} + \frac{\psi_i \psi_j}{\psi^2} \bigg)(E_{\rm L} - \bar{E}) \bigg\rangle - \bigg\langle \frac{\psi_i}{\psi} \bigg\rangle \bar{E}_j - \bigg\langle \frac{\psi_j}{\psi} \bigg\rangle \bar{E}_i \bigg] + \bigg\langle \frac{\psi_i}{\psi} E_{{\rm L},j} \bigg\rangle - \bigg\langle \frac{\psi_i}{\psi} \bigg\rangle \langle E_{{\rm L},j} \rangle + \bigg\langle \frac{\psi_j}{\psi} E_{{\rm L},i} \bigg\rangle - \bigg\langle \frac{\psi_j}{\psi} \bigg\rangle \langle E_{{\rm L},i} \rangle. \qquad (8)$$

The additional terms have zero expectation value for an infinite sample, but cancel most of the fluctuations in the existing terms for a finite sample, making the method vastly more efficient. In fact, away from the minimum, the terms outside the square parentheses fluctuate less than those within. Note also that $\bar{E}_{ij}$ in Eq. (7), evaluated on a finite sample, is not invariant under renormalization of ψ by a parameter-dependent constant, but $\bar{E}_{ij}$ in Eq. (8) is.

We note that Eqs. (6) and (8) are *not* the gradient and the Hessian of the energy estimated on the particular finite set of sampled points. In fact, any method that attempts to minimize the energy, by minimizing the energy evaluated on a finite sample of Monte Carlo points, is bound to require a very large sample and therefore be highly inefficient for the reason discussed in the introduction. Nightingale and Melik-Alaverdian [8] discuss a method in which a zero-variance principle can be realized that is destroyed if one employs expressions derived by minimizing the energy on a finite Monte Carlo sample.

Newton method.—In both the energy and the variance minimization methods, the gradient, $\mathbf{b}$, and the Hessian, $\mathbf{A}$, are used to update the variational parameters, $\mathbf{c}$, using Newton's method, $\mathbf{c}_{\rm next} = \mathbf{c}_{\rm current} - \mathbf{A}^{-1}\mathbf{b}$.

Note that if we are far away from the minimum, or if the number of Monte Carlo samples, $N_{\rm MC}$, is small, then the Hessian of Eq. (8) need not be positive definite, whereas the approximate Hessian of Eq. (4) is always positive definite. Further, even for positive definite Hessians, the new parameter values may make the wave function worse if one is not sufficiently close to the minimum for the quadratic approximation to hold or if the approximate Hessian of Eq. (4) is not sufficiently accurate. Hence, we determine the eigenvalues of the Hessian and add to the diagonal of the Hessian the negative of the most negative eigenvalue (if one exists) plus a constant $a_{\rm diag}$. This shifts the eigenvalues by the added constant. As $a_{\rm diag}$ is in-

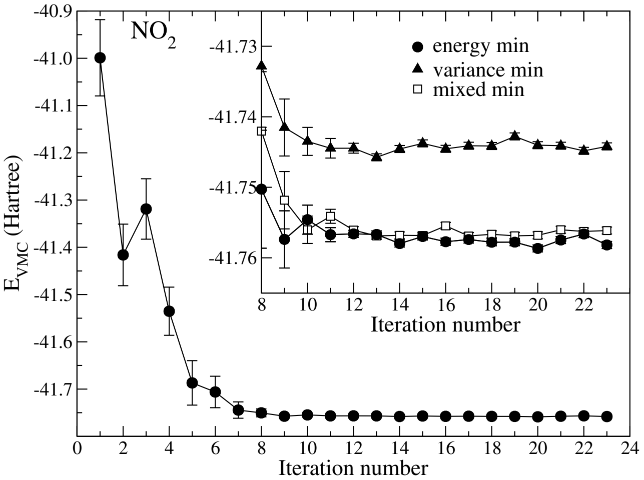

FIG. 1. Energy of NO$_2$ versus iteration number for energy minimization. Inset: the later iterations on an expanded scale and also the energies from minimizing the variance and minimizing the linear combination. The linear combination yields almost as good an energy as energy minimization.

creased, the parameter changes become smaller and rotate from the Newtonian direction to the steepest descent direction. As an aside, we note that for the form of the wave functions used and the molecules studied, we find that the eigenvalues of the Hessians of Eqs. (4) and (8) span 11 orders of magnitude when the parameters are close to optimal.

Results.—We have tested the methods on NO$_2$ and the excited ^1B_u state of decapentaene ($C_{10}H_{12}$) using a nonlocal pseudopotential to remove core electrons. We optimize the parameters in a flexible Jastrow factor [15] that contains electron-electron, electron-nucleus, and electron-electron-nucleus terms, making a total of 43 free parameters. The starting Jastrow is a crude electron-electron Jastrow of the form $\exp(\frac{br}{1+r})$, where b is set by the cusp conditions for antiparallel- and parallel-spin electrons.

In Fig. 1, we plot the energy, and, in Fig. 2 the root mean square fluctuations of the local energy, σ, of NO$_2$ as a function of the iteration number as we energy optimize the 43 free parameters in the Jastrow. The first six iterations employ a very small MC sample, $N_{\mathrm{MC}} = 1000$, and $a_{\mathrm{diag}} = 0.2$. For each of the next six iterations we increase N_{MC} by a multiplicative factor of 4 and decrease a_{diag} by a multiplicative factor of 0.1. The remaining 11 iterations are performed with the values at the end of this process, namely, $N_{\mathrm{MC}} = 4\,096\,000$, and $a_{\mathrm{diag}} = 2 \times 10^{-7}$. (Setting $a_{\mathrm{diag}} = 0$ would work equally well for these iterations.) The first few iterations are extremely fast due to the small value of N_{MC} and achieve most of the optimization. In the insets we show the later iterations on an expanded scale, and also the energies and σ from minimizing the variance [using Eqs. (2) and (4)] and from minimizing a linear combination, with the variance having a weight of 0.05 and the energy a weight of 0.95. Of course, the variance-minimized wave functions have a lower σ and the energy-minimized wave functions a lower energy. The mixed-minimization wave functions have an energy that is almost as good as that of the energy-minimized wave functions, and, a σ that is in between.

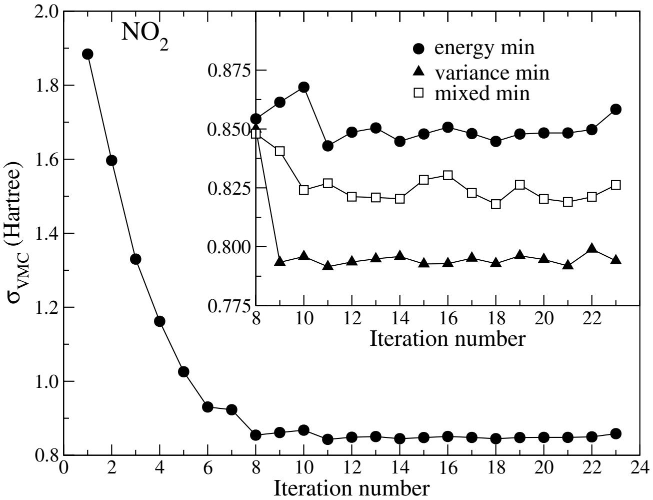

FIG. 2. Same as Fig. 1 but for the rms fluctuations of the local energy σ rather than the energy. The linear combination σ is half way between those from energy minimization and variance minimization.

The computer time, for a given statistical error, is proportional to $\sigma^2 T_{\mathrm{corr}}$, where T_{corr} is the autocorrelation time of the energy as defined in Ref. [21]. In DMC, energy-minimized wave functions will have a smaller T_{corr} than variance-minimized wave functions, since both σ and T_{corr} serve to lower the DMC energy relative to the variational

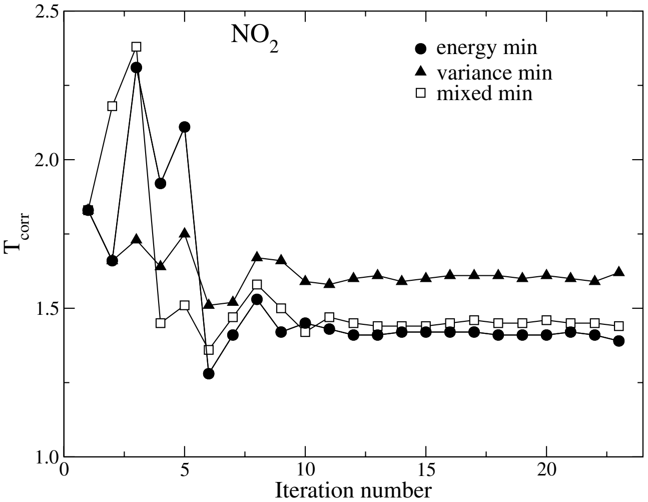

FIG. 3. The autocorrelation time, T_{corr}, of NO$_2$ versus iteration number. Energy minimization gives the smallest T_{corr}, variance minimization the largest, and, the mixed minimization a value that is close to that from energy minimization.

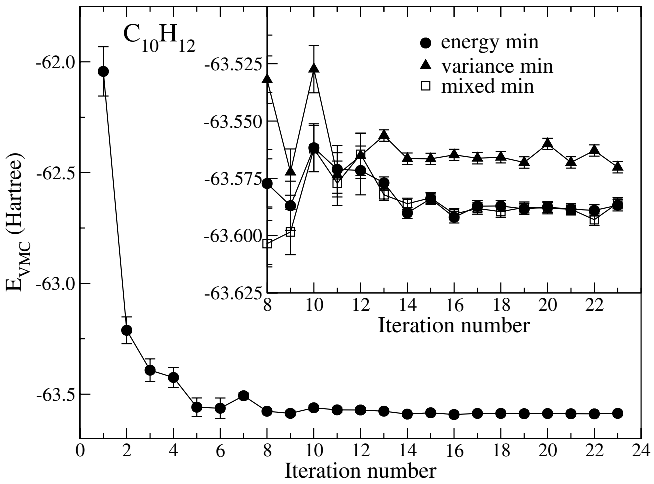

FIG. 4. Same as Fig. 1 but for decapentaene ($C_{10}H_{12}$).

energy. In Fig. 3, we show T_{corr} for each of the three methods. The energy-minimized wave function has a smaller value of T_{corr} than the variance-minimized wave function, even in the VMC method. The mixed-minimization wave function has a T_{corr} that is close to that of the energy-minimized wave function. For the variance, energy, and mixed optimizations, $\sigma^2 T_{\text{corr}}$ is 1.04(1), 1.03(1), and 0.98(1) H in VMC, and, 3.05(3), 2.87(3), and 2.75(3) H in DMC using a time step of 0.05 H^{-1}. Hence, the wave functions obtained from the mixed optimization are the most efficient ones.

We note that E, σ, and T_{corr} converge in 12 iterations. In fact, they converge in two iterations if we use from the outset larger values for N_{MC} and optimal values for a_{diag} (zero after first iteration).

In Fig. 4 we plot the energy of the excited ^1B_u state of a larger molecule, decapentaene ($C_{10}H_{12}$), as a function of the iteration number. For the first six iterations we optimize just the 13 parameters in the electron-nucleus and the electron-electron Jastrows, and, optimize the full set of 43 parameters starting from iteration 7. As in the case of NO_2, we employ $N_{\text{MC}} = 1000$ and $a_{\text{diag}} = 0.2$ during the first six iterations. The next six are performed with $N_{\text{MC}} = 16\,000$ and the final 11 iterations are performed with $N_{\text{MC}} = 256\,000$ and $a_{\text{diag}} = 2 \times 10^{-5}$. The results are similar to those for NO_2, and so in the interest of brevity we omit plots for σ and T_{corr}.

It is remarkable that the optimization is stable with as few as 1000 MC configurations. In contrast, if Eq. (7) is used for the Hessian, then the fluctuations are much larger and the method becomes unstable for the molecules treated here even if we increase the number of Monte Carlo configurations, N_{MC}, by a factor of a thousand to 10^6 configurations. (We can make it stable by increasing substantially also the value of a_{diag}, but this increases the number of iterations needed to converge.) Hence, the simple change going from Eq. (7) to Eq. (8) that entails no additional computational cost, results in a gain in efficiency of at least 3 orders of magnitude.

We thank Peter Nightingale for valuable discussions. Supported by NSF (DMR-0205328), NASA, Sandia National Laboratory, and Stichting voor Fundamenteel Onderzoek der Materie (FOM).

[1] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[2] *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by M. P. Nightingale and C. J. Umrigar, NATO ASI Ser. C. 525 (Kluwer, Dordrecht, 1999).
[3] J. H. Bartlett, J. J. Gibbons, and C. G. Dunn, Phys. Rev. **47**, 679 (1935); R.L. Coldwell, Int. J. Quantum Chem., Quantum Chem. Symp. **11**, 215 (1977).
[4] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988); C. J. Umrigar, K. G. Wilson, and J.W. Wilkins, in *Computer Simulation Studies in Condensed Matter Physics: Recent Developments*, edited by D. P. Landau, K. K. Mon, and H. B. Schüttler, Springer Proc. Phys. (Springer, Berlin, 1988); C. J. Umrigar, Int. J. Quantum Chem., Quantum Chem. Symp. **23**, 217 (1989).
[5] A. Harju, B. Barbiellini, S. Siljamäki, R. M. Nieminen, and G. Ortiz, Phys. Rev. Lett. **79**, 1173 (1997).
[6] Martin Snajdr, Jason R. Dwyer, and Stuart M. Rothstein, J. Chem. Phys. **111**, 9971 (1999); **114**, 6960(E) (2001).
[7] Xi Lin, Hongkai Zhang, and Andrew M. Rappe, J. Chem. Phys. **112**, 2650 (2000); Myung Won Lee, Massimo Mella, and Andrew M. Rappe, physics/0411209.
[8] M. P. Nightingale and Melik-Alaverdian, Phys. Rev. Lett. **87**, 043401 (2001).
[9] S. Fahy, in Ref. [2].
[10] Claudia Filippi and Stephen Fahy, J. Chem. Phys. **112**, 3523 (2000).
[11] Friedemann Schautz and Stephen Fahy, J. Chem. Phys. **116**, 3533 (2002).
[12] David Prendergast, David Bevan, and Stephen Fahy, Phys. Rev. B **66**, 155104 (2002).
[13] Friedemann Schautz and Claudia Filippi, J. Chem. Phys. **120**, 10 931 (2004).
[14] Sandro Sorella, Phys. Rev. B **64**, 024512 (2001); Michele Casula and S. Sorella, J. Chem. Phys. **119**, 6500 (2003).
[15] Claudia Filippi and C. J. Umrigar, J. Chem. Phys. **105**, 213 (1996). Our Jastrow is related to the one here.
[16] Chien-Jung Huang, C. J. Umrigar, and M. P. Nightingale, J. Chem. Phys. **107**, 3007 (1997).
[17] Martin Snajdr and Stuart M. Rothstein, J. Chem. Phys. **112**, 4935 (2000).
[18] R. Assaraf and M. Caffarel, J. Chem. Phys. **113**, 4028 (2000); **119**, 10 536 (2003).
[19] See, e.g., W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, England 1992).
[20] After submission we learned that Eq. (8) has been independently derived by S. Sorella, cond-mat/0502553.
[21] C.J. Umrigar, Phys. Rev. Lett. **71**, 408 (1993); C.J. Umrigar, in Ref. [2].