ARTICLES

Robust wave function optimization procedures in quantum Monte Carlo methods

Dario Bressanini[a)] and Gabriele Morosi[b)]
Dipartimento di Scienze Chimiche, Fisiche e Matematiche, Universita' dell'Insubria Sede di Como, via Lucini 3, 22100 Como, Italy

Massimo Mella[c)]
Dipartimento di Chimica Fisica ed Elettrochimica, Universita' di Milano, via Golgi 19, 20133 Milano, Italy

The energy variance optimization algorithm over a fixed ensemble of configurations in variational Monte Carlo often encounters problems of convergence. Being formally identical to a problem of fitting data, we re-examine it from a statistical maximum-likelihood point of view. We show that the assumption of an underlying Gaussian distribution of the local energy, implicit in the standard variance minimization scheme, is not theoretically nor practically justified, and frequently generates convergence problems. We propose alternative procedures for optimization of trial wave functions in quantum Monte Carlo and successfully test them by optimizing a trial wave function for the helium trimer. © 2002 American Institute of Physics. [DOI: 10.1063/1.1455618]

INTRODUCTION

The problem of the optimization of a trial function with many nonlinear parameters for a quantum system is still an open issue. This is especially true in the field of quantum Monte Carlo simulations, where usually one uses trial wave functions for which the analytical evaluation of the energy is impossible. For this reason, the variational Monte Carlo (VMC) method is used to numerically evaluate the energy and other properties of the trial wave functions, and to optimize them. While in standard quantum mechanical calculations it is common to optimize a trial wave function by minimizing its variational energy, in VMC simulations it is much more common to minimize the variance of the local energy $\sigma^2(H) = \langle H^2 \rangle - \langle H \rangle^2$, rather than the energy itself. The reason is that the minimization of the energy by a Monte Carlo method is much more troublesome from a numerical point of view than the minimization of its variance, which is however far from being problem free, as discussed below.

The minimization of the variance, as an alternative to the minimization of the energy, has been first proposed, to the best of our knowledge, by Weinstein[1] in the context of lower bound calculations of eigenvalues of the Schrödinger equation, and by Bartlett *et al.*,[2] who suggested the use of the variance of the local energy as a criterion of goodness for approximate wave functions. Frost and co-workers[3–7] and Conroy[8–10] implemented the minimization of the variance within a numerical scheme to optimize trial wave functions for small atoms and molecules. Variance minimization was introduced in VMC by Coldwell[11] and Umrigar and

co-workers.[12] The main reason this method is rarely applied nowadays in standard computational methods is that its use involves the expectation value of the square of the Hamiltonian, a very difficult quantity to compute analytically. A second problem is that for some (albeit, very poor) trial wave functions this quantity might diverge. A third problem is that sometimes, again for very poor wave functions, a local minimum cannot be found [for example, it is easy to check that using a single Gaussian function to describe the ground state of the hydrogen atom, a minimum of $\sigma^2(H)$ does not exist]. The main point, i.e., the difficulty in its computation, is not an issue in VMC, while the other two problems are not met usually using realistic trial wave functions. However the issue of the quality of the optimized wave function remains to be settled. For an infinitely flexible trial wave function, both energy and variance minimization procedures recover the exact ground state, in practice however, using an incomplete basis, the resulting wave functions are different. Whether or not the variance optimized wave function is an overall better wave function than the energy optimized one has been discussed in the QMC literature.[13–15] Initially the variance optimized wave function was thought to be at least as good as the energy optimized one, if not superior for certain properties. At present there are suggestions that this is not so,[13–16] especially if the trial function is not flexible enough. These conclusions seem to confirm the results that James and Coolidge[17] and later Goodisman[18] obtained a long time ago in the field of *ab initio* methods. For this reason, several papers have recently focused on the problem of minimizing the energy using VMC.[14,19] However, the modern energy optimization algorithms can approach only few electron systems and in any case the methods do not appear to be as fast as the standard variance minimization algorithm. For these reasons the variance minimization method is still generally

a)Electronic mail: dario.bressanini@uninsubria.it
b)Electronic mail: gabriele.morosi@uninsubria.it
c)Electronic mail: massimo.mella@unimi.it

used in VMC and it is easy to predict that its use will last. In spite of this large diffusion, there is a problem that the quantum Monte Carlo practitioner is sometimes faced with while trying to optimize a trial wave function using a fixed ensemble of configurations. Varying the parameters in the trial wave function to minimize the variance of the local energy, the optimization algorithm reaches a minimum and produces the best set of parameters for that fixed ensemble of configurations. However, when the new trial wave function is used in a VMC simulation, often it appears even worse than the previous function, in the sense that both the energy and the variance are worse.

The problem we are addressing in this paper we call *false convergence*. We need to understand its origin in order to make the algorithm more reliable and able to generate better wave functions. The origin of the problem can be traced back to the presence in the fixed ensemble of some configurations that have values of the local energy very different from the average. We call these walkers ''bad'' as their presence in the fixed ensemble spoils the minimization process. A practical, and sometimes effective, solution is to eliminate these bad walkers from the fixed ensemble in some *ad hoc* way. For example, retaining only those walkers whose local energy is within a fixed chosen window, or discarding all the walkers whose local energy is off by $L\sigma$ from the average value, for some value of L, as suggested by Kent et al.[20] These solutions are far from being satisfactory from the theoretical point of view, since these empirical criteria are somehow arbitrary. A better solution would be to design a method that automatically deals with these bad walkers in a nonarbitrary way.

The variance minimization algorithm searches the minimum

$$\min_{\mathbf{a}} \sum_i^N \ (E_L(\mathbf{R}_i,\mathbf{a})-E_R)^2 \qquad (1)$$

given N data points $(\mathbf{R}_i, E_L(\mathbf{R}_i))$, where $\mathbf{R}_i$ represents the walker and $E_L(\mathbf{R}_i)$ its local energy associated to the trial wave function $\Psi(\mathbf{R};\mathbf{a})$ that depends on the adjustable parameters vector $\mathbf{a}$. E_R is a reference energy, close to the exact one. In order to understand why things sometimes go wrong in this process and how to correctly deal with bad walkers, we first need to understand what we are *really* doing when we are minimizing the variance of the local energy. We will show that the origin of the problem of false convergence might not reside in the bad walkers, but rather in the optimization procedure itself.

We now derive Eq. (1) in the standard way, and show that an alternative derivation, focused on the discrete algorithm used to search the minimum, involving a sum over discrete points and not a continuous integration, can shed light on the problem of false convergence.

The integral point of view

The usual way to derive Eq. (1) is to start from the variance minimum principle[3,8,12,21]

$$\sigma^2(H)=\langle H^2\rangle-\langle H\rangle^2\geqslant 0. \qquad (2)$$

For an eigenstate, the variance of the Hamiltonian is at a local minimum, and equal to zero.

Optimizing a wave function using this principle means solving the problem

$$\min_{\mathbf{a}} \sigma^2(H(\mathbf{a})). \qquad (3)$$

Explicitly writing the integrals

$$\min_{\mathbf{a}} \sigma^2(H(\mathbf{a}))=\min_{\mathbf{a}} \frac{\int \Psi(\mathbf{R};\mathbf{a})^2(E_L(\mathbf{R};\mathbf{a})-\langle H\rangle)^2\,d\mathbf{R}}{\int \Psi(\mathbf{R};\mathbf{a})^2\,d\mathbf{R}}. \qquad (4)$$

In practice, since the integrals cannot be evaluated analytically, they are estimated using a finite (small) number of integration points, generated by VMC. Suppose the VMC process generated N points $\mathbf{R}_i$ distributed according to $\Psi(\mathbf{R};\mathbf{a})^2$, the discrete approximation of the variance is

$$\sigma^2(H(\mathbf{a}))\cong \frac{1}{N}\sum_i^N \ (E_L(\mathbf{R}_i;\mathbf{a})-\langle H\rangle)^2. \qquad (5)$$

Instead of minimizing $\sigma^2(H(\mathbf{a}))$, sometimes it is preferable to minimize a related quantity, namely the second moment of the local energy with respect to a fixed (or reference) energy E_R,

$$\mu_{E_R}^2(H(\mathbf{a}))= \frac{\int \Psi(\mathbf{R};\mathbf{a})^2(E_L(\mathbf{R};\mathbf{a})-E_R)^2\,d\mathbf{R}}{\int \Psi(\mathbf{R};\mathbf{a})^2\,d\mathbf{R}}$$

$$\cong \frac{1}{N}\sum_i^N \ (E_L(\mathbf{R}_i;\mathbf{a})-E_R)^2. \qquad (6)$$

The constant E_R should be close to the energy of the state being sought, although the optimization does not depend strongly on its value. Minimizing this quantity [which many authors call $\sigma^2(H(\mathbf{a}))$ without making any distinction] is almost equivalent to minimizing the variance. A little algebra shows that

$$\mu_{E_R}^2(H)=\sigma^2(H)+(\langle H\rangle-E_R)^2 \qquad (7)$$

from which Eq. (1) is recovered, using the minimum principle of Eq. (2).

This is, in short, what we might call *the integral point of view*. Adopting this view, one uses Eq. (5) or Eq. (6). Taking the integral point of view, it is very easy to invent different functionals upon which to base the minimization process. For examples,

$$\frac{\int (E_L(\mathbf{R}_i;\mathbf{a})-E_R)^2\Psi(\mathbf{R})^4\,d\mathbf{R}}{(\int \Psi(\mathbf{R})^2\,d\mathbf{R})^2} \qquad (8)$$

and

$$\frac{\int |E_L(\mathbf{R}_i;\mathbf{a})-E_R|\Psi(\mathbf{R})^2\,d\mathbf{R}}{\int \Psi(\mathbf{R})^2\,d\mathbf{R}} \qquad (9)$$

and several others have been suggested and discussed by Alexander et al.,[22] in the context of the biased-selection Monte Carlo method. In general, any functional of the general form

$$\frac{\int P(\mathbf{R})f(E_L(\mathbf{R}_i;\mathbf{a})-E_R)d\mathbf{R}}{\int P(\mathbf{R})d\mathbf{R}} \tag{10}$$

can be proposed, as long as P is a probability distribution and f is a non-negative function such that $f(0)=0$.

Although this derivation of the optimization process is clean and simple, it does not shed any light on the problems of false convergence that, as mentioned before, one often encounters in practice. In the derivation, the only role played by the discretization process is to approximate the integrals by finite sums, while the focus is really on the integral functional. However it is really the discretization process that is the source of the problem. So, if we want to clarify this issue, we must start from a point of view that includes the discrete nature of the algorithm from the very beginning.

Furthermore, why should one use the variance functional [Eqs. (5) and (6)] rather than the functionals of Eq. (8) and (9), or one of the many possible others? Is there any *a priori* reason to prefer one to the others?

The fitting point of view

If the discretization process would be only a way to approximate the integrals involved in the different functionals, we should be very surprised by the usually good quality of the results obtained with the generally rather small number of points used (usually from some hundreds to some thousands). As correctly pointed out by Umrigar and co-workers,[12] there is really no need to invoke any kind of integral approximation to justify the algorithm. The key observation is that the correct way to look at the optimization algorithm is as a fitting process, and if the true wave function were representable by an n-parameter trial wave function, then only n walkers would be necessary to determine the exact wave function. So the minimization of Eq. (1) should be considered in its own right, with no connection to integrals. This crucial point was already pointed out by Frost and co-workers[4] who observed that, if Ψ is the exact wave function and E_R is the exact energy, the sum in Eq. (1) is zero for *any* distribution of points, whether or not they are chosen to yield good approximations to integrals. Since the discrete nature of the algorithm is intrinsic in the *fitting point of view*, from this perspective we can hope to clarify the problem of false convergence. A related analysis, starting from a different point of view, of the effects of approximating the required integrals by finite sums has been recently carried out by Kent and co-workers.[20]

If we do not start by invoking the variance of the Hamiltonian, we must ask ourselves why should we minimize the square of the deviation of the local energy rather than its absolute deviation, or its fourth power or other formulas. Which criterion should drive us to select one among them? The answer, well known to statisticians, is the *maximum likelihood estimator*.[23] We can view the above formulas as fitting criteria of N measurements, affected by errors. Each algorithm generates the vector $\mathbf{a}$ that best reproduces these values. There are vectors $\mathbf{a}$ that are very "unlikely" and others that are more "likely" since they give an average energy closer to E_R. We now *assume* that these values are affected

by normally distributed independent errors with a fixed standard deviation σ. The *likelihood* of a single observation is

$$P_i \propto e^{-1/2(E_L(\mathbf{R}_i;\mathbf{a})-E_R)^2/\sigma^2}. \tag{11}$$

The total probability, assuming the same value of σ for each point, is

$$P_{\text{tot}} \propto \prod_i^N e^{-1/2(E_L(\mathbf{R}_i;\mathbf{a})-E_R)^2/\sigma^2}. \tag{12}$$

The usual (and often implicit) choice at this point is to select the set of parameters $\mathbf{a}$ such that the total probability (or likelihood) is maximized. This is equivalent to maximizing the logarithm of Eq. (12), leading to Eq. (1).

So, the familiar least squares fitting in Eq. (1) is a maximum likelihood estimation of the parameters if the underlying distribution is assumed to be a Gaussian. Implicit is the assumption that points far from the average value are very unlikely, since the tails of a Gaussian distribution decay very fast. These points, when they are present, influence the least-squares procedure more than they should. In the statistical jargon these points are called *outliers*.[23] If they occur they might bias the fitting procedure to produce meaningless values of the parameters. The problem lies in the fact that the probability of these points in the assumed Gaussian model is so small that the maximum likelihood estimator tries to distort the whole model to take them into account. As a result, using Eq. (1) to find the optimum set of parameters $\mathbf{a}$ might not guarantee the convergence to the set that minimizes the functional in Eq. (4).

Now that we have identified the problem, we need to understand why we encounter these outliers in VMC and how to deal with them. This last point is the subject of *robust estimation*, a well-established field of statistics. Let us concentrate on the first point.

Is the local energy distribution really Gaussian?

Given a trial wave function Ψ_T, we can define the distribution of the local energy as

$$\rho(E)=\frac{\int \Psi_T(\mathbf{R})\delta(E-E_L(\mathbf{R}))\Psi_T(\mathbf{R})d\mathbf{R}}{\int \Psi_T^2(\mathbf{R})d\mathbf{R}}. \tag{13}$$

If the trial wave function is an exact eigenstate, the local energy is constant and the local energy distribution is a Dirac delta. This of course is not the usual case, but if the trial wave function is sufficiently good, the local energy distribution is very sharp, located around the exact energy, and it can be well approximated by a Gaussian. However, in the general case there really are no reasons why $\rho(E)$ should be well approximated by a Gaussian.

Let us consider two analytical examples first. Consider the trial wave function $\Psi_T=e^{-ar}$ for the hydrogen atom. The local energy associated to this function is

$$E_L=-\frac{a^2}{2}+\frac{a-1}{r}. \tag{14}$$

Note that if $a>1$, $E_L>-a^2/2$, while if $a<1$, $E_L<-a^2/2$. This means that $\rho(E)$ is zero below or above
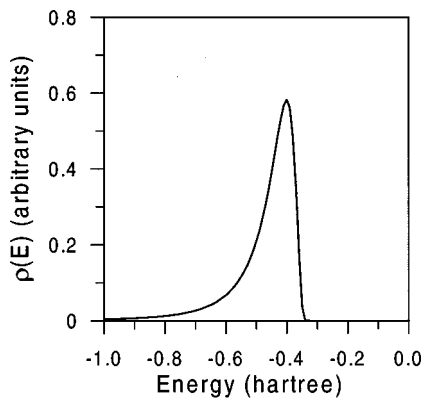
FIG. 1. Local energy distribution for the hydrogen atom wave function $\Psi_T = e^{-0.8r}$ (see text).

$-a^2/2$. For $a=1$, $E_L = -1/2$, the local energy distribution $\rho(E)$ is a Dirac delta and the trial wave function becomes the ground state eigenfunction.

Suppose that $a < 1$. Then Eq. (13) can be easily integrated, giving (not normalized)

$$\rho(E) = \begin{cases} \dfrac{(1-a)^3 a^3 e^{-\{[4(a-1)a]/(a^2+2E)\}}}{(a^2+2E)^4}, & E < -a^2/2, \\ 0, & E \geq -a^2/2. \end{cases} \tag{15}$$

A similar formula can be obtained in the case $a > 1$. The plot for $a = 0.8$ is shown in Fig. 1. First note that this function is neither a Gaussian, nor peaked on the exact eigenvalue of $-1/2$. The distribution is quite skewed, and it even contains an essential singularity. For values greater than $-a^2/2$, the distribution is zero. On the other hand, the left tail decays as E^{-4}: substantially slower than Gaussian decay. This means that the number of outliers observed will be higher than predicted by the Gaussian model. Of course for $a=1$ the trial wave function becomes the exact eigenstate and the local energy distribution becomes a Dirac delta.

Another interesting case is the 3D harmonic oscillator. Let us consider the trial function $\Psi_T = e^{-ar^2}$ where for $a = 1/2$ the exact wave function is recovered. For the case $0 < a < \frac{1}{2}$, Eq. (13) can be integrated, giving

$$\rho(E) = \begin{cases} 0, & E \leq 3a, \\ \dfrac{8a}{1-4a^2} \sqrt{\dfrac{4a(E-3a)}{\pi(1-4a^2)}} e^{-\{[4a(E-3a)]/(1-4a^2)\}}, & \\ & E > 3a. \end{cases} \tag{16}$$

Even in this case the local energy distribution is far from being a Gaussian. The tail decays exponentially, so even in this case, in a Monte Carlo simulation, we should expect a number of bad walkers greater than what a simple Gaussian local energy distribution predicts.

Only for very simple cases the analytical form of the local energy distribution can be obtained. In the case of the helium atom, even for the correlated wave function $\Psi_T = e^{-2(r_1+r_2)+r_{12}/2}$, satisfying the cusp conditions, the local energy distribution has a tail decaying as E^{-4}. This means that the satisfaction of the cusp conditions is not directly

involved in the shape of the energy distribution. Unfortunately it is not possible to make a general statement on its form. However, it seems clear, even from numerical evidence[20,24,25] that the assumption of a Gaussian tail, implicit in the variance minimization, might not be a good choice, since in this case the bad walkers can spoil the entire optimization process, especially if the trial function is not flexible enough.

Of course, the local energy distribution associated to the trial wave function $\Psi(\mathbf{R}; \mathbf{a})$ changes when the parameters $\mathbf{a}$ are modified. Here we make the reasonable assumption that the generic analytical form of this distribution does not change significantly with $\mathbf{a}$.

A robust estimation procedure

A possible solution to this problem comes from the field of *Robust Estimation*. If the outlier points are troublesome, a sensible way to proceed could be to take into account their presence from the very beginning, assuming a different functional form for the tails of the local energy distribution. Instead of starting from a simple Gaussian distribution [Eq. (11)] we can assume a distribution with higher tails. This approach is called the *M-estimates* approach.

For example, we could assume that the local energy distribution decays exponentially

$$P_i \propto e^{-|E_L(\mathbf{R}_i; \mathbf{a}) - E_R|}. \tag{17}$$

Here the tails of the distribution, although exponentially decaying, are asymptotically much higher than those of a Gaussian. Using simple algebra it is easy to show that this assumption leads to the minimization of the mean absolute deviation, rather than the mean square deviation,

$$\min_{\mathbf{a}} \sum_i^N |E_L(\mathbf{R}_i, \mathbf{a}) - E_R|. \tag{18}$$

So a lower weight is given to the outliers and the risk of spoiling the optimization is reduced. In this way we have derived one of the functionals studied by Alexander *et al.*[22] and empirically found to cure the problem of false convergence by Lester and co-workers[26] in electronic structure calculations.

Of course it would be possible to start with a distribution with an even higher tail, like the Cauchy or Lorentzian distribution, obtaining

$$\min_{\mathbf{a}} \sum_i^N \log[1 + (E_L(\mathbf{R}_i; \mathbf{a}) - E_R)^2/2]. \tag{19}$$

Even better would be to use the "real" local energy distribution for a given trial wave function model. For example, we have seen [Eq. (15)] that the tails of the distribution for the hydrogen atom should behave as E^{-4}. A distribution with a similar tail might be

$$P_i \propto \frac{s}{1 + s^4(E_L(\mathbf{R}_i; \mathbf{a}) - E_R)^4}, \tag{20}$$

where s is an adjustable parameter, which leads to the minimization of the quantity
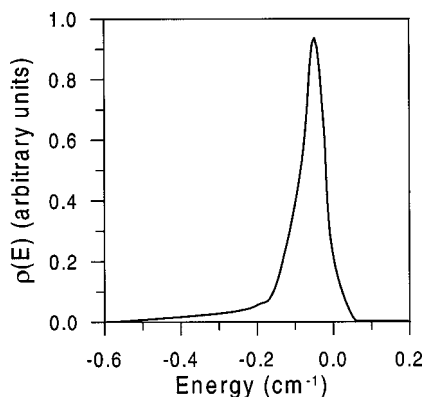
FIG. 2. Local energy distribution for the helium trimer wave function (see text).

FIG. 3. Sequence of optimization cycles for the helium trimer wave function.

$$\sum_{i}^{N} \ \log\left(\frac{1 + s^4 (E_L(\mathbf{R}_i ; \mathbf{a}) - E_R)^4}{s} \right). \tag{21}$$

However, one rarely knows the functional form of the local energy distribution for the trial function employed, but it should be sufficient to use any distribution with sufficiently high tails to reduce or eliminate the problems caused by the outliers.

A REAL EXAMPLE: THE HELIUM TRIMER

The optimization of good trial functions is a very important issue in the study of pure and doped helium clusters, and in the field of weakly bound clusters in general. The functional forms usually employed[25,27–30] are not very flexible and this frequently generates problems like those we have described.[25,28,29] A particularly problematic system is the helium trimer, due to its very diffuse nature and high anharmonicity. The optimization of trial functions for the helium trimer has been reported to cause problems,[25,28] so it is a good testing ground for our investigation. The local energy distribution for a common trial wave function[27] for $^4\mathrm{He}_3$, computed using VMC, is shown in Fig. 2. We employed the LM2M2 potential,[31] but the results do not depend on the particular form of the potential used. The local energy distribution has a noticeably non-Gaussian shape. Although most of the curve lies in the negative energy region, there is a sizable portion in the positive part. More problematic is the fact that there is a very slowly decaying tail for energies very far from the average. In this case, the optimization of the variance of the local energy is likely to cause problems and probably it would be better to assume a more slowly decaying distribution. To test our proposal, we optimized the sum of the absolute deviations and the sum of the square deviations.

We performed two sets of five optimization cycles, starting from the same trial wave function and from the same ensemble of 5000 walkers. In the first set we optimized the variance of the local energy, while in the second set we optimized the mean absolute deviation. After each optimization step the energy was computed by a VMC simulation. The newly generated ensemble was used as fixed sample for the next optimization.
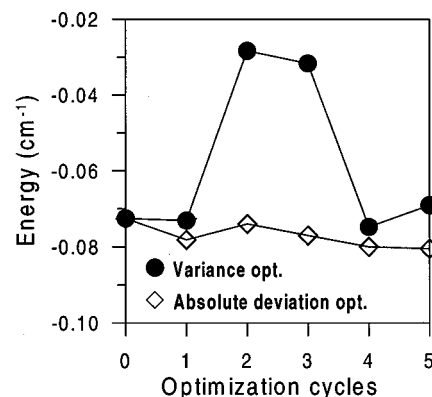
The calculated energies are reported in Fig. 3. As expected, the variance minimization is troublesome. The variational energy of the starting trial wave function is $-0.0725(1) \ \mathrm{cm}^{-1}$. Although the variance minimization procedure is able to produce, after four optimization cycles, a trial wave function slightly better than the starting one, giving $-0.0747(1) \ \mathrm{cm}^{-1}$, its trend is erratic, and even generates very bad trial wave functions during the process. On the other end, minimizing the mean absolute deviation seems to be a much more reliable procedure, showing a much smoother trend. Furthermore, the final wave function gives a much better energy of $-0.0805(1) \ \mathrm{cm}^{-1}$. So a useful side benefit of minimizing the absolute deviation rather than the variance is that the resulting wave function is closer to the one that minimizes the energy. We tried optimizing other objective functions, including Eq. (19), all sharing a less fast decaying tail than a Gaussian, obtaining similar results. This means that minimizing the mean absolute deviation is not the only possible choice here, and the good behavior is not directly related to some peculiar property of the absolute deviation function, but rather to the assumption of a less fast decaying tail. There might be cases where even an exponentially decaying tail might be too fast, and some benefit could be gained in assuming a powerlike decaying tail.

CONCLUSIONS

In this paper we have re-examined the discrete nature of the energy variance minimization algorithm in quantum Monte Carlo methods from the *maximum-likelihood* point of view, without regards to any integral approximation. This allowed us to unveil the origin of the problem of false convergence and to suggest alternative procedures designed to reduce or eliminate the problem. We also showed that the assumption of a Gaussian distribution of the local energy, implicit in the standard variance optimization algorithm, is not theoretically justified. We tested our proposal optimizing a trial wave function for the helium trimer using several algorithms. The minimization of the mean absolute deviation is shown to be more reliable than the variance minimization, void of convergence problems, and able to generate trial wave functions with better variational energies.

ACKNOWLEDGMENTS

[1] D. H. Weinstein, Proc. Natl. Acad. Sci. U.S.A. **20**, 529 (1934).

[2] J. H. Bartlett, Jr., J. J. Gibbons, Jr., and C. G. Dunn, Phys. Rev. **47**, 679 (1935).

[3] A. A. Frost, J. Chem. Phys. **10**, 240 (1942).

[4] A. A. Frost, R. E. Kellogg, and E. C. Curtis, Rev. Mod. Phys. **32**, 313 (1960).

[5] A. A. Frost, R. E. Kellogg, B. M. Gimarc, and J. D. Scargle, J. Chem. Phys. **35**, 827 (1961).

[6] B. M. Gimarc and A. A. Frost, J. Chem. Phys. **39**, 1698 (1963).

[7] D. K. Harriss and A. A. Frost, J. Chem. Phys. **40**, 204 (1964).

[8] H. Conroy, J. Chem. Phys. **41**, 1331 (1964).

[9] H. Conroy, J. Chem. Phys. **41**, 1336 (1964).

[10] H. Conroy, J. Chem. Phys. **47**, 930 (1967).

[11] R. L. Coldwell, Int. J. Quantum Chem., Quantum Chem. Symp. **11**, 215 (1977).

[12] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).

[13] M. Snajdr and S. M. Rothstein, J. Chem. Phys. **112**, 4935 (2000).

[14] X. Lin, H. K. Zhang, and A. M. Rappe, J. Chem. Phys. **112**, 2650 (2000).

[15] F. J. Galvez, E. Buendia, and A. Sarsa, J. Chem. Phys. **115**, 1166 (2001).

[16] R. Bianchi, D. Bressanini, P. Cremaschi, M. Mella, and G. Morosi, Int. J. Quantum Chem. **57**, 321 (1996).

[17] H. M. James and A. S. Coolidge, Phys. Rev. **51**, 860 (1937).

[18] J. Goodisman, J. Chem. Phys. **45**, 3659 (1966).

[19] M. Snajdr, J. R. Dwyer, and S. M. Rothstein, J. Chem. Phys. **114**, 6960 (2001).

[20] P. R. C. Kent, R. J. Needs, and C. Rajagopal, Phys. Rev. B **59**, 12344 (1999).

[21] D. Bressanini and P. J. Reynolds, in *Advances in Chemical Physics*, edited by D. M. Ferguson, J. I. Siepmann, and D. G. Truhlar (Wiley, New York, 1999), Vol. 105, pp. 37.

[22] S. A. Alexander, R. L. Coldwell, H. J. Monkhorst, and J. D. Morgan III, J. Chem. Phys. **95**, 6622 (1991).

[23] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, 2nd ed. (Cambridge University Press, Cambridge, 1992).

[24] R. Bianchi, P. Cremaschi, G. Morosi, and C. Puppi, Chem. Phys. Lett. **148**, 86 (1988).

[25] M. Lewerenz, J. Chem. Phys. **106**, 4596 (1997).

[26] A. A. Aspuru-Guzik, O. Couronne, I. Ovcharenko, and W. A. Lester, Jr. (unpublished).

[27] S. W. Rick, R. D. Lynch, and J. D. Doll, J. Chem. Phys. **95**, 3506 (1991).

[28] D. Bressanini, M. Zavaglia, M. Mella, and G. Morosi, J. Chem. Phys. **112**, 717 (2000).

[29] M. Casalegno, M. Mella, G. Morosi, and D. Bressanini, J. Chem. Phys. **112**, 69 (2000).

[30] K. B. Whaley, Int. Rev. Phys. Chem. **13**, 41 (1994).

[31] R. A. Aziz and M. J. Slaman, J. Chem. Phys. **94**, 8047 (1991).