Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions

C. J. Umrigar,[1] Julien Toulouse,[1] Claudia Filippi,[2] S. Sorella,[3] and R. G. Hennig[4]

[1]*Theory Center and Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA*
[2]*Instituut Lorentz, Universiteit Leiden, Niels Bohrweg 2, Leiden, NL-2333 CA, The Netherlands*
[3]*INFM-Democritos National Simulation Centre, and SISSA, Trieste, Italy*
[4]*Materials Science and Engineering, Cornell University, Ithaca, New York 14853, USA*

(Received 3 November 2006; published 15 March 2007)

We present a simple, robust, and highly efficient method for optimizing all parameters of many-body wave functions in quantum Monte Carlo calculations, applicable to continuum systems and lattice models. Based on a strong zero-variance principle, diagonalization of the Hamiltonian matrix in the space spanned by the wave function and its derivatives determines the optimal parameters. It systematically reduces the fixed-node error, as demonstrated by the calculation of the binding energy of the small but challenging C_2 molecule to the experimental accuracy of 0.02 eV.

DOI: 10.1103/PhysRevLett.98.110201 PACS numbers: 02.70.Ss, 31.10.+z, 31.25.Qm

Many important problems in computational physics and chemistry can be reduced to the computation of the dominant eigenvalues of matrices or integral kernels. For problems with many degrees of freedom, Monte Carlo approaches such as diffusion, reptation, auxilliary-field, and transfer-matrix Monte Carlo are among the most accurate and efficient methods [1]. Their efficiency, and in most cases accuracy, rely crucially on good approximate eigenstates. For fermionic systems, the antisymmetry constraint leads to the Fermion-sign problem that forces practical implementations to fix the nodes to those of an approximate trial wave function. The resulting fixed-node error is the main uncontrolled error in fermionic quantum Monte Carlo (QMC) calculations. Despite great effort aimed at eliminating this error entirely [2], to date there is no generally applicable solution; therefore, systematic improvement of the wave function by optimization of an increasing number of variational parameters is the most practical approach for reducing this error.

In this Letter we develop a robust and highly efficient method for optimizing all the parameters in approximate wave functions. Application to the small but challenging C_2 molecule demonstrates that the method systematically eliminates the fixed-node error. The method is based on energy minimization in variational Monte Carlo (VMC) calculations and extends the zero-variance linear optimization method [3] to nonlinear parameters. The approach is simpler than the Newton method [4] as it does not require second derivatives of the wave function. Moreover, in contrast to the perturbative optimization method [5], it enables the simultaneous parameter optimization of both the Jastrow and the determinantal parts of the wave function with equal efficiency.

Form of wave functions.—We employ N-electron wave functions which depend on $N_{\rm opt}$ variational parameters collectively denoted by $\mathbf{p} = (\mathbf{c}, \boldsymbol{\alpha}, \boldsymbol{\lambda})$ and $3N$ electronic coordinates, $\mathbf{R}$,

$$\Psi(\mathbf{p}, \mathbf{R}) = J(\boldsymbol{\alpha}, \mathbf{R}) \sum_{i=1}^{N_{\rm CSF}} c_i C_i(\boldsymbol{\lambda}, \mathbf{R}), \tag{1}$$

where $J(\boldsymbol{\alpha}, \mathbf{R})$ is a Jastrow factor that contains electron-nuclear, electron-electron, and electron-electron-nuclear terms. Each of the $N_{\rm CSF}$ configuration state functions (CSFs), $C_i(\boldsymbol{\lambda}, \mathbf{R})$, is a symmetry-adapted linear combination of Slater determinants of single-particle orbitals $\phi_\mu(\mathbf{r})$, which are themselves expanded in terms of $N_{\rm basis}$ basis functions $\chi_\nu(\mathbf{r})$: $\phi_\mu(\mathbf{r}) = \sum_{\nu=1}^{N_{\rm basis}} \lambda_{\mu\nu}\chi_\nu(\mathbf{r})$. The variational parameters are the linear CSF coefficients $\mathbf{c}$, the nonlinear Jastrow parameters $\boldsymbol{\alpha}$, and the nonlinear expansion coefficients of the orbitals $\boldsymbol{\lambda}$. In practice, we optimize only $N_{\rm CSF} - 1$ CSF coefficients (as the normalization of the wave function is irrelevant) and a set of nonredundant orbital rotation parameters [6,7].

Optimization of wave functions.—We extend the zero-variance method of Nightingale and Melik-Alaverdian [3] for linear parameters to nonlinear parameters. At each optimization step, the wave function is expanded to linear order in $\Delta\mathbf{p} = \mathbf{p} - \mathbf{p}^0$ around the current parameters $\mathbf{p}^0$,

$$\Psi_{\rm lin}(\mathbf{p}, \mathbf{R}) = \Psi_0(\mathbf{R}) + \sum_{i=1}^{N_{\rm opt}} \Delta p_i \Psi_i(\mathbf{R}), \tag{2}$$

where $\Psi_0 = \Psi(\mathbf{p}_0)$ is the current wave function and $\Psi_i = (\partial\Psi(\mathbf{p})/\partial p_i)_{\mathbf{p}=\mathbf{p}_0}$ are the $N_{\rm opt}$ derivatives of the wave function with respect to the parameters. On an *infinite* Monte Carlo (MC) sample, the parameter variations $\Delta\mathbf{p}$ minimizing the energy calculated with the linearized wave function of Eq. (2) are the lowest eigenvalue solution of the generalized eigenvalue equation

$$\mathbf{H}\,\Delta\mathbf{p} = E\mathbf{S}\Delta\mathbf{p}, \tag{3}$$

where $\mathbf{H}$ and $\mathbf{S}$ are the Hamiltonian and overlap matrices in the $(N_{\rm opt} + 1)$-dimensional basis formed by the current wave function and its derivatives $\{\Psi_0, \Psi_1, \Psi_2, \cdots, \Psi_{N_{\rm opt}}\}$

and $\Delta p_0 = 1$. On a *finite* MC sample, following Ref. [3], we estimate these matrices by

$$H_{ij} = \left\langle \frac{\Psi_i}{\Psi_0} \frac{H\Psi_j}{\Psi_0} \right\rangle_{\Psi_0^2}, \qquad S_{ij} = \left\langle \frac{\Psi_i}{\Psi_0} \frac{\Psi_j}{\Psi_0} \right\rangle_{\Psi_0^2}, \quad (4)$$

where H is the electronic Hamiltonian. We employ the notation that $\langle \frac{\Psi_i}{\Psi_0} \frac{A\Psi_j}{\Psi_0} \rangle_{\Psi_0^2}$ denotes the statistical estimate of $\langle \Psi_i | \hat{A} | \Psi_j \rangle / \langle \Psi_0 | \Psi_0 \rangle$ evaluated using MC configurations sampled from Ψ_0^2. The estimator for the matrix $\mathbf{H}$ in Eq. (4) is *nonsymmetric* and is *not* the symmetric Hamiltonian matrix that one would obtain by minimizing the energy of the finite MC sample. In fact, as shown in Ref. [3], it is only this nonsymmetric estimator for $\mathbf{H}$ that leads to a strong zero-variance principle: the variance of the parameter changes $\Delta\mathbf{p}$ in Eq. (3) vanishes not only in the limit that Ψ_0 is the exact eigenstate but even in the limit that Ψ_{lin} with optimized parameters is an exact eigenstate. For the wave functions that we employ, the asymmetric Hamiltonian matrix results in 1 to 2 orders of magnitude smaller fluctuations in the parameter changes $\Delta\mathbf{p}$.

Quite generally, methods that minimize the energy of a finite MC sample are stable only if a much larger number of MC configurations is employed than is necessary for the variance minimization method [8] because the energy of a finite sample is not bounded from below but the variance is. However, it is possible to devise simple modifications of these energy-minimization methods that require orders of magnitude fewer MC configurations. Both the zero-variance linear method employed here and the modified Newton method [4] are examples of such modifications.

Having obtained the parameter variations $\Delta\mathbf{p}$ by solving Eq. (3), there is no unique way to update the parameters in the wave function. The simplest procedure of incrementing the current parameters by $\Delta\mathbf{p}$, $\mathbf{p}_0 \rightarrow \mathbf{p}_0 + \Delta\mathbf{p}$ works for the linear parameters but is not guaranteed to work for the nonlinear parameters if the linear approximation of Eq. (2) is not good. A better but more complicated procedure is to fit the original wave function form to the optimal linear combination. We next discuss a simple prescription that avoids doing this fit.

At first, it may appear that nothing can be done to make the linear approximation better, but this is in fact not the case. One can exploit the freedom of the normalization of the wave function to alter the dependence on the nonlinear parameters as follows. Consider the differently normalized wave function $\bar{\Psi}(\mathbf{p}, \mathbf{R}) = N(\mathbf{p})\Psi(\mathbf{p}, \mathbf{R})$ such that $\bar{\Psi}(\mathbf{p}_0, \mathbf{R}) = \Psi(\mathbf{p}_0, \mathbf{R}) \equiv \Psi_0(\mathbf{R})$ and $N(\mathbf{p})$ depends only on the nonlinear parameters. Then, the derivatives of $\bar{\Psi}(\mathbf{p})$ at $\mathbf{p} = \mathbf{p}_0$ are

$$\bar{\Psi}_i = \Psi_i + N_i \Psi_0, \quad \text{where } N_i = (\partial N(\mathbf{p})/\partial p_i)_{\mathbf{p}=\mathbf{p}_0}, \quad (5)$$

with $N_i = 0$ for linear parameters. The first-order expansion of $\bar{\Psi}(\mathbf{p})$ after optimization is

$$\bar{\Psi}_{\text{lin}} = \Psi_0 + \sum_{i=1}^{N_{\text{opt}}} \Delta\bar{p}_i \bar{\Psi}_i. \quad (6)$$

Since $\bar{\Psi}_{\text{lin}}$ and Ψ_{lin} were obtained by optimization in the same variational space, they must be proportional to each other, so $\Delta\bar{\mathbf{p}}$ are related to $\Delta\mathbf{p}$ by a uniform rescaling

$$\Delta\bar{\mathbf{p}} = \frac{\Delta\mathbf{p}}{1 - \sum_{i=1}^{N_{\text{opt}}} N_i \Delta p_i}. \quad (7)$$

Since the rescaling factor can be anywhere between $-\infty$ and ∞ the choice of normalization can affect not only the magnitude of the parameter changes but even the sign.

For the nonlinear parameters, we have found that, in all cases considered here, a fast and stable optimization is achieved if N_i are determined by imposing the condition that each derivative $\bar{\Psi}_i$ is orthogonal to a linear combination of Ψ_0 and Ψ_{lin}, i.e., $\langle \xi \frac{\Psi_0}{\|\Psi_0\|} + (1-\xi) \frac{\Psi_{\text{lin}}}{\|\Psi_{\text{lin}}\|} | \bar{\Psi}_i \rangle = 0$, where ξ is a constant between 0 and 1, resulting in

$$N_i = -\frac{\xi D S_{0i} + (1-\xi)(S_{0i} + \sum_j^{\text{nonlin}} S_{ij}\Delta p_j)}{\xi D + (1-\xi)(1 + \sum_j^{\text{nonlin}} S_{0j}\Delta p_j)}, \quad (8)$$

with $D = \sqrt{1 + 2\sum_j^{\text{nonlin}} S_{0j}\Delta p_j + \sum_{j,k}^{\text{nonlin}} S_{jk}\Delta p_j \Delta p_k}$, where the sums are over only the nonlinear parameters [9]. The simple choice $\xi = 1$ first used by Sorella in the context of the stochastic reconfiguration method [10] often leads to good parameter variations, but can result in arbitrarily large parameter variations that may or may not be desirable. The safer choice, $\xi = 0$, minimizes the norm of the linear wave function change $\|\bar{\Psi}_{\text{lin}} - \Psi_0\|$ in the case that only the nonlinear parameters are varied, but, it can yield arbitrarily small parameter changes even far from the energy minimum. In contrast, the choice $\xi = 1/2$, imposing $\|\bar{\Psi}_{\text{lin}}\| = \|\Psi_0\|$, guarantees finite parameter changes, until the energy minimum is reached.

If instead of finding the parameter changes from Eqs. (3) and (7) one expands the Rayleigh quotient $\langle \bar{\Psi}_{\text{lin}} | \hat{H} | \bar{\Psi}_{\text{lin}} \rangle / \langle \bar{\Psi}_{\text{lin}} | \bar{\Psi}_{\text{lin}} \rangle$ with $\xi = 1$ to second order in $\Delta\bar{\mathbf{p}}$, one recovers the stochastic reconfiguration with Hessian acceleration (SRH) method with $\beta = 0$ of Ref. [11]. It turns out that the SRH method is much less stable and converges more slowly, particularly for large systems with many parameters.

Stabilization of the optimization.—When the parameter values are far from optimal or if the MC sample used to evaluate $\mathbf{H}$ and $\mathbf{S}$ is very small, the updated parameters may be worse than the original ones. However, it is possible to devise a scheme to stabilize the method in a manner similar to that used for the modified Newton method [4]. Stabilization is achieved by adding a positive constant, $a_{\text{diag}} \geq 0$, to the diagonal of the Hamiltonian matrix except for the first element: $H_{ij} \rightarrow H_{ij} + a_{\text{diag}}\delta_{ij}(1 - \delta_{i0})$. As a_{diag} becomes larger, the parameter variations $\Delta\mathbf{p}$ become smaller and rotate towards the steepest descent direction. In practice, the value of a_{diag} is automatically adjusted at

each optimization step. Once **H** and **S** have been computed, three values of a_{diag} differing from each other by factors of 10 are used to predict three new wave functions. A short MC run is then performed using correlated sampling to compute energy differences of these wave functions more accurately than the energy itself. The optimal value of a_{diag} is then calculated by parabolic interpolation on these three energies, with some bounds imposed.

Results.—We demonstrate the performance of our optimization method on the $^1\Sigma_g^+$ ground state of the C_2 molecule at the experimental equilibrium interatomic distance of 2.3481 Bohr, employing a scalar-relativistic Hartree-Fock (HF) pseudopotential [12] with a large Gaussian polarized valence quintuple-zeta one-electron basis (12s, 10p, and 4d functions contracted to 5s, 5p, and 4d functions) to ensure basis-set convergence. The quantum chemistry package GAMESS [13] is used to obtain multiconfiguration self-consistent field (MCSCF) wave functions in complete active spaces (CAS) generated by distributing n valence electrons in m orbitals [CAS(n, m)]. We also employ restricted active space RAS(n, m) wave functions consisting of truncations of the CAS(n, m) wave functions to quadruple excitations. We retain only those CSFs with absolute coefficients larger than some cutoff; smaller cutoffs give larger numbers of CSFs. The initial trial wave function is this linear combination of CSFs multiplied by a Jastrow factor with all free parameters chosen to be zero.

Figure 1 illustrates the convergence of the VMC energy during the simultaneous optimization of 24 Jastrow, 73 CSF and 174 orbital parameters for a truncated CAS(8,14) wave function. The energy converges to an accuracy of about 1 mHa in 5 iterations, making it the most rapidly convergent method for optimizing all the parameters in Jastrow-Slater wave functions.

Figure 2 shows the total VMC and DMC energies of the C_2 molecule as a function of the number of determinants retained in VMC-optimized truncated Jastrow-Slater CAS(8,14) wave functions. For comparison, the convergence of the $^3\Sigma_g^+$ ground state energy of Si_2 is also shown. For each VMC and DMC energy, there are three curves. For the upper curve, only the Jastrow parameters are optimized and the CSF and orbital coefficients are fixed at their MCSCF values. For the middle curve, the Jastrow and CSF parameters are optimized simultaneously while, for the lower curve, the Jastrow, CSF and orbital parameters are optimized simultaneously. With fixed CSF coefficients, the energy does not improve monotonically with the number of determinants. In contrast, if the CSF coefficients are reoptimized then of course the VMC energy improves monotonically. For this example the DMC energy also improves monotonically, implying that the nodal hypersurface of the wave function improves monotonically, even though this is not guaranteed by the method.

The difference in the behaviors of the C_2 and Si_2 molecules is striking. While for Si_2 the energy shows a small gradual decrease upon increasing the number of determinants, for C_2 there is a very rapid initial decrease, a manifestation of strong correlation. The fixed-node error of a single-determinant wave function using MCSCF natural orbitals is about 6 mHa for Si_2 and about 46 mHa for C_2, almost an order of magnitude larger.

Retaining all the determinants of a CAS(8,14) wave function would be costly in QMC but one can estimate the energy in this limit by extrapolation. Figure 3 shows a quadratic fit of the VMC and DMC energies obtained with truncated, fully reoptimized wave functions with respect to the sum of the squares of the MCSCF CSF coefficients

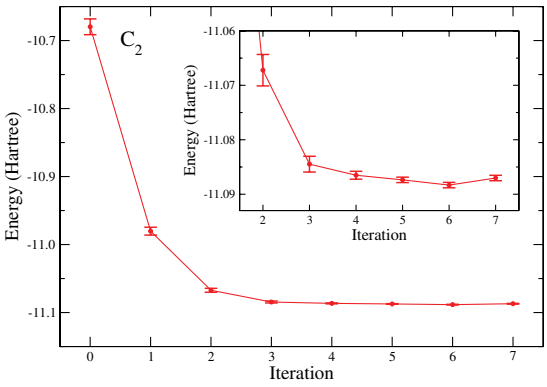

FIG. 1 (color online). Convergence of the VMC total energy of the C_2 molecule when simultaneously optimizing 24 Jastrow, 73 CSF and 174 orbital parameters for a truncated CAS(8,14) wave function. The number of MC configurations range from 10 000 for the first iteration to 400 000 for the last iterations.

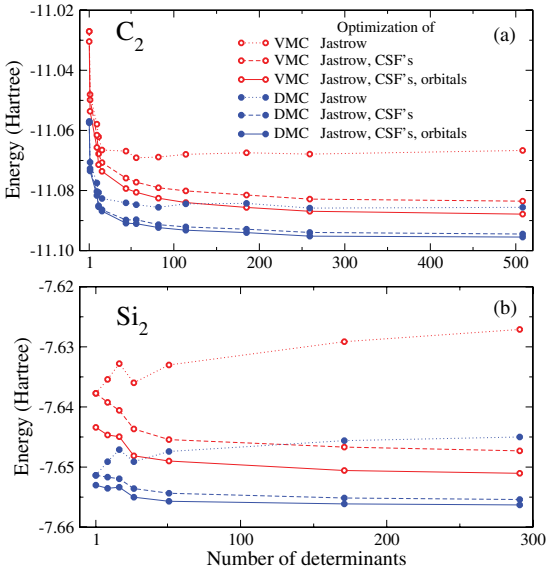

FIG. 2 (color online). VMC and DMC total energies of the C_2 and Si_2 molecules, versus the number of determinants in truncated Jastrow-Slater CAS(8,14) wave functions, for different levels of optimization. Both VMC and DMC energies decrease monotonically if the CSF coefficients are reoptimized in VMC but not if only the Jastrow factor is optimized. Statistical errors are smaller than the plotted symbols.

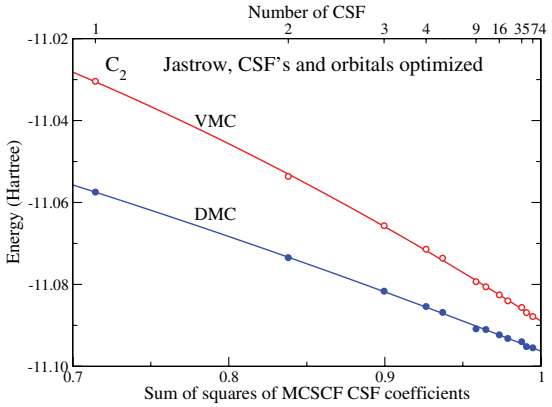

FIG. 3 (color online). Parabolic extrapolation (to 1) of the VMC and DMC energies obtained with the truncated, fully reoptimized Jastrow-Slater CAS(8,14) wave functions with respect to the sum of the squares of the MCSCF CSF coefficients.

retained, $\sum_{i=1}^{N_{\mathrm{CSF}}}(c_i^{\mathrm{MCSCF}})^2$. This sum is equal to 1 in the limit that all the CSFs are included.

Table I reports the extrapolated DMC energies and well depths (dissociation energy + zero-point energy) for a series of fully optimized Jastrow-Slater CAS and RAS wave functions. Increasing the size of the active space results in a monotonic improvement of the total energy and the well depth in column 3 obtained using a good estimate of the exact atomic energy from a DMC calculation with a CAS(4,13) wave function. Chemical accuracy (1 kcal/mol = 0.04 eV) is reached for the largest active space. Alternatively, as shown in column 4, good well depths can be obtained using much smaller active spaces by relying on a partial cancellation of error employing atomic wave functions consistent with the molecular ones: for the molecular single-determinant wave function, an atomic single-determinant wave function is also used; for the molecular CAS(8, m) wave functions, atomic CAS(4, $m/2$) wave functions are used. In this case, while the use of a single-determinant wave function yields an error of 0.67 eV, all the CAS wave functions yield well depths that agree with the exact one to better than chemical accuracy. This behavior parallels the standard quantum chemistry approaches where single-determinant reference methods such as coupled cluster are in error by as much as 0.2 eV while multireference configuration interaction yields chemical accuracy [14]. Density functional methods perform rather poorly, giving a well depth of 6.69 and 4.69 eV within LDA and B3LYP, respectively.

TABLE I. Total energy and well depth of the C_2 molecule obtained in DMC with a series of fully optimized Jastrow-Slater wave functions. The well depth in column 3 is obtained using a well-converged DMC atomic energy, −5.43403(3) whereas that in column 4 uses atomic CAS wave functions consistent with the molecular ones. The DMC time step, τ, is 0.01 Ha^{-1}. Using $\tau =$ 0.1 Ha^{-1}, alters the total energies by at most 0.4 mHa and the well depths by at most 0.01 eV.

Wave function	Total energy (Ha)	Well depth (eV)	
1 determinant	−11.0575(3)	5.15(1)	5.69(1)
CAS(8,8)	−11.0925(3)	6.11(1)	6.39(1)
CAS(8,10)	−11.0942(2)	6.15(1)	6.37(1)
CAS(8,14)	−11.0962(2)	6.21(1)	6.38(1)
CAS(8,18)	−11.0985(3)	6.27(1)	6.36(1)
RAS(8,26)	−11.1007(3)	6.33(1)	-
Estimated exact	−11.1020(3)[b]	6.36(2)[a]	6.36(2)[a]

[a]Scalar-relativistic, valence-corrected estimate of Ref. [14].
[b]From atomic $E_{\mathrm{DMC}} = 5.434\,03$ Ha and well depth 6.365 eV.

Conclusions.—The method presented provides a systematic means of eliminating the main limitation of present day QMC calculations, namely, the fixed-node error, and supercedes variance minimization [8] as the method of choice for optimizing many-body wave functions. Extension of the method to optimization of the DMC energy is possible and extension to geometry optimization will overcome the other major limitation of QMC methods.

We thank M. P. Nightingale, R. Assaraf, and A. Savin for discussions. Computations were performed at NERSC, NCSA, OSC, and CNF. Supported by NSF (No. DMR-0205328, No. EAR-0530301), Sandia Natl. Lab., FOM (Netherlands), and MIUR-COFIN05.

[1] W. M. C. Foulkes *et al.*, Rev. Mod. Phys. **73**, 33 (2001); *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); S. Baroni and S. Moroni, Phys. Rev. Lett. **82**, 4745 (1999); S. Zhang and H. Krakauer, *ibid.* **90**, 136401 (2003); M. P. Nightingale and H. Blöte, *ibid.* **60**, 1562 (1988).
[2] D. M. Ceperley and B. J. Alder, J. Chem. Phys. **81**, 5833 (1984); Shiwei Zhang and M. H. Kalos, Phys. Rev. Lett. **67**, 3074 (1991).
[3] M. P. Nightingale and V. Melik-Alaverdian, Phys. Rev. Lett. **87**, 043401 (2001).
[4] C. J. Umrigar and C. Filippi, Phys. Rev. Lett. **94**, 150201 (2005).
[5] A. Scemama and C. Filippi, Phys. Rev. B **73**, 241101 (2006).
[6] J. Toulouse and C. J. Umrigar, J. Chem. Phys. **126**, 084102 (2007).
[7] F. Schautz and C. Filippi, J. Chem. Phys. **120**, 10 931 (2004).
[8] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).
[9] Defining optimal parameter variations $\Delta\mathbf{p}_{\mathrm{opt}}$ as maximizing the overlap of $\Psi(\mathbf{p}_0 + \Delta\mathbf{p}_{\mathrm{opt}})$ and Ψ_{lin}, for any N_i, $\Delta\mathbf{p}_{\mathrm{opt}} = \Delta\bar{p} + \mathcal{O}(|\Delta\bar{p}|^2)$. For the N_i of Eq. (8) and parameters linear in the exponent, the coefficients of the second order terms are tricovariances and therefore small.
[10] S. Sorella, Phys. Rev. B **64**, 024512 (2001).
[11] S. Sorella, Phys. Rev. B **71**, 241103 (2005).
[12] M. Burkatzki, C. Filippi, and M. Dolg (unpublished).
[13] M. W. Schmidt *et al.*, J. Comput. Chem. **14**, 1347 (1993).
[14] L. Bytautas and K. Ruedenberg, J. Chem. Phys. **122**, 154110 (2005).