Optimization of quantum Monte Carlo wave functions by energy minimization

Julien Toulouse[a)]
Cornell Theory Center, Cornell University, Ithaca, New York 14853

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

(Received 13 October 2006; accepted 4 January 2007; published online 22 February 2007)

We study three wave function optimization methods based on energy minimization in a variational Monte Carlo framework: the Newton, linear, and perturbative methods. In the Newton method, the parameter variations are calculated from the energy gradient and Hessian, using a reduced variance statistical estimator for the latter. In the linear method, the parameter variations are found by diagonalizing a nonsymmetric estimator of the Hamiltonian matrix in the space spanned by the wave function and its derivatives with respect to the parameters, making use of a strong zero-variance principle. In the less computationally expensive perturbative method, the parameter variations are calculated by approximately solving the generalized eigenvalue equation of the linear method by a nonorthogonal perturbation theory. These general methods are illustrated here by the optimization of wave functions consisting of a Jastrow factor multiplied by an expansion in configuration state functions (CSFs) for the C_2 molecule, including both valence and core electrons in the calculation. The Newton and linear methods are very efficient for the optimization of the Jastrow, CSF, and orbital parameters. The perturbative method is a good alternative for the optimization of just the CSF and orbital parameters. Although the optimization is performed at the variational Monte Carlo level, we observe for the C_2 molecule studied here, and for other systems we have studied, that as more parameters in the trial wave functions are optimized, the diffusion Monte Carlo total energy improves monotonically, implying that the nodal hypersurface also improves monotonically. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2437215]

I. INTRODUCTION

Quantum Monte Carlo (QMC) methods (see, e.g., Refs. 1–3) constitute an alternative to standard *ab initio* methods of quantum chemistry for accurate calculations of the electronic structure of atoms, molecules, and solids. The two most commonly used variants, variational Monte Carlo (VMC) and diffusion Monte Carlo (DMC), rely on an explicitly correlated trial wave function, generally consisting for atoms and molecules of a Jastrow factor multiplied by a short expansion in configuration state functions (CSFs), each consisting of a linear combination of Slater determinants, a form capable of encompassing most of the electron correlation effects. To fully benefit from the considerable flexibility in the form of the wave function, it is crucial to be able to efficiently optimize the parameters in these wave functions.

Variance minimization in correlated sampling[4–6] has become the most frequently used method in QMC for optimizing wave functions because it is far more efficient than *straightforward* energy minimization on a finite Monte Carlo sample. However, while the method works relatively well for the optimization of the Jastrow factor, it is much less effective for the optimization of the determinantal part of the wave function (though still possible[4,7,8]). Furthermore, there is some evidence that energy-optimized wave functions give on average better expectation values for other observables than variance-optimized ones (see, e.g., Refs. 9 and 10). As a result, a lot of effort has recently been devoted to developing efficient methods for the optimization of QMC wave functions by energy minimization. On the other hand, it should be mentioned that variance-minimized wave functions often have a smaller time-step error in DMC.

We now summarize some of the major approaches that have been proposed for energy minimization in VMC. The most efficient method to minimize the energy with respect to linear parameters, such as the CSF coefficients, is to solve the associated generalized eigenvalue equation using a nonsymmetric estimator of the Hamiltonian matrix.[11] The energy fluctuation potential (EFP) method[12–16] is very efficient for optimizing some nonlinear parameters and has been applied very successfully to the optimization of the orbitals[13,16] and CSF coefficients.[15,16] It has also been applied to the optimization of Jastrow factors in periodic solids.[14] The perturbative EFP method, a simplification of the EFP method, retains the same convergence rate for the optimization of the orbitals and CSF coefficients while decreasing the computational cost.[17] The stochastic reconfiguration (SR) method, originally developed for lattice systems,[18] has been applied to the full optimization of atomic and molecular wave functions consisting of an antisymmetrized geminal power part multi-

[a)]Electronic mail: toulouse@tc.cornell.edu
[b)]Electronic mail: cyrus@tc.cornell.edu

plied by a Jastrow factor.[19,20] It is related to the perturbative EFP method and is simpler but less efficient.[16,17] The Newton method is a conceptually simple and general optimization method but a straightforward implementation of it in QMC is rather inefficient.[21,22] However, an improved version of it, making use of a reduced variance estimator of the Hessian matrix,[23] is very efficient for the optimization of Jastrow factors. Another modified version of the Newton method with an approximate Hessian, named stochastic reconfiguration with Hessian acceleration (SRH), has been applied to lattice models.[24]

In this work, we investigate the three best energy minimization methods for the optimization of the Jastrow, CSF, and orbital parameters of QMC wave functions: the Newton, linear and perturbative methods. The Newton method has already been applied very successfully to the optimization of Jastrow factors by Umrigar and Filippi,[23] and in this paper it is also applied to the optimization of the determinantal part of the wave function. The linear method is an extension of the zero-variance generalized eigenvalue equation approach of Nightingale and Melik-Alaverdian[11] to arbitrary nonlinear parameters: at each step of the iterative procedure, the wave function is linearized with respect to the parameters and the optimal values of the parameters are found by diagonalizing the Hamiltonian in the space spanned by the current wave function and its derivatives with respect to the parameters. This method is briefly presented in Ref. 25. The perturbative method coincides with the perturbative EFP method of Scemama and Filippi[17] for the optimization of the CSF and orbital parameters. Here, we put this approach on more general grounds by recasting it as a simplification of the linear method where the generalized eigenvalue equation is solved approximately by a nonorthogonal perturbation theory. The Newton and linear methods are very efficient for the optimization of the Jastrow, CSF, and orbital parameters. The perturbative method is a good alternative for the optimization of just the CSF and orbital parameters.

The paper is organized as follows. In Sec. II, the parametrization of the trial wave function is presented. The energy minimization procedures are discussed in Sec. III, and their realizations in VMC are discussed in Sec. IV. Section V contains computational details of the calculations performed on the C_2 molecule to test the optimization methods, and in Sec. VI we present the results. Section VII contains our conclusions.

Hartree atomic units are used throughout this work.

II. WAVE FUNCTION PARAMETRIZATION AND DERIVATIVES

We begin by describing the form of the wave function used, the actual parametrization chosen for the optimization, and the corresponding derivatives of the wave function with respect to the parameters.

A. Form of the wave function

We use an N-electron wave function of the usual Jastrow-Slater form that is denoted at each iteration of the optimization procedure by

$$|\Psi_0\rangle = \hat{J}(\boldsymbol{\alpha}^0)|\Phi_0\rangle, \tag{1}$$

where $\hat{J}(\boldsymbol{\alpha}^0)$ is a Jastrow operator depending on the current parameters α_i^0 and $|\Phi_0\rangle$ is a multideterminantal wave function. For notational convenience, we assume that the wave function $|\Psi_0\rangle$ is always normalized to unity, i.e., $\langle\Psi_0|\Psi_0\rangle = 1$. In practice, $|\Psi_0\rangle$ can have arbitrary normalization.

The wave function $|\Phi_0\rangle$ is a linear combination of N_{CSF} orthonormal configuration state functions, $|C_I\rangle$, with current coefficients c_I^0,

$$|\Phi_0\rangle = \sum_{I=1}^{N_{\mathrm{CSF}}} c_I^0 |C_I\rangle. \tag{2}$$

Each CSF is a short linear combination of products of spin-up and spin-down Slater determinants, $|D_{\mathbf{k}}^{\uparrow}\rangle$ and $|D_{\mathbf{k}}^{\downarrow}\rangle$, $|C_I\rangle = \Sigma_{\mathbf{k}} d_{I,\mathbf{k}} |D_{\mathbf{k}}^{\uparrow}\rangle |D_{\mathbf{k}}^{\downarrow}\rangle$, where the coefficients $d_{I,\mathbf{k}}$ are fully determined by the spatial and spin symmetries of the state considered (see, e.g., Ref. 26). The use of CSFs is important to decrease the number of coefficients to be optimized and to ensure the correct symmetry of the wave function after optimization in the presence of statistical noise. The $N_{\uparrow}$-electron and $N_{\downarrow}$-electron spin-assigned Slater determinants are generated from a set of current orthonormal orbitals, $|D_{\mathbf{k}}^{\uparrow}\rangle = \hat{a}_{k_1\uparrow}^{\dagger}\hat{a}_{k_2\uparrow}^{\dagger}\cdots\hat{a}_{k_{N_\uparrow}\uparrow}^{\dagger}|\mathrm{vac}\rangle$ and $|D_{\mathbf{k}}^{\downarrow}\rangle = \hat{a}_{k_{N_\uparrow+1}\downarrow}^{\dagger}\hat{a}_{k_{N_\uparrow+2}\downarrow}^{\dagger}\cdots\hat{a}_{k_N\downarrow}^{\dagger}|\mathrm{vac}\rangle$, where $\hat{a}_{k\sigma}^{\dagger}$ (with $\sigma = \uparrow, \downarrow$) is the fermionic creation operator for the spatial orbital $|\phi_k^0\rangle$ in the spin-σ determinant, and $|\mathrm{vac}\rangle$ is the vacuum state of second quantization. The (occupied and virtual) spatial orbitals are written as linear combinations of N_{bas} basis functions $|\chi_\mu\rangle$ (e.g., Slater or Gaussian functions) with current coefficients $\lambda_{k,\mu}^0$, $|\phi_k^0\rangle = \Sigma_{\mu=1}^{N_{\mathrm{bas}}} \lambda_{k,\mu}^0 |\chi_\mu\rangle$.

The N-electron Jastrow operator, $\hat{J}(\boldsymbol{\alpha}^0)$, is defined by its matrix elements in the N-electron position basis $|\mathbf{R}\rangle = |\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\rangle$,

$$\langle\mathbf{R}|\hat{J}(\boldsymbol{\alpha}^0)|\mathbf{R}'\rangle = J(\boldsymbol{\alpha}^0;\mathbf{R})\,\delta(\mathbf{R}-\mathbf{R}'), \tag{3}$$

where $J(\boldsymbol{\alpha}^0;\mathbf{R})$ is the spin-assigned Jastrow factor, a real positive function of $\mathbf{R}$ which is symmetric under the exchange of two same-spin electrons. Its action on an arbitrary N-electron state $|\Phi\rangle$ is given by $\langle\mathbf{R}|\hat{J}(\boldsymbol{\alpha}^0)|\Phi\rangle = J(\boldsymbol{\alpha}^0;\mathbf{R})\Phi(\mathbf{R})$, where $\Phi(\mathbf{R}) = \langle\mathbf{R}|\Phi\rangle$. The Jastrow operator is Hermitian, $\hat{J}(\boldsymbol{\alpha}^0)^{\dagger} = \hat{J}(\boldsymbol{\alpha}^0)$. We use flexible Jastrow factors consisting of the exponential of the sum of electron-nucleus, electron-electron, and electron-electron-nucleus terms, written as systematic polynomial or Padé expansions[27] (see also Refs. 7 and 28).

B. Wave function parametrization

We want to optimize the Jastrow parameters α_i, the CSF coefficients c_I, and the orbital coefficients $\lambda_{k,\mu}$. Some parameters in the Jastrow factor are fixed by imposing the electron-nucleus and electron-electron cusp conditions[29] on the wave function; the other Jastrow parameters are varied freely. Due to the arbitrariness of the overall normalization of the wave function, only $N_{\mathrm{CSF}}-1$ CSF coefficients need be varied, e.g., the coefficient of the first configuration can be kept fixed. The situation is more involved for the orbital coefficients

which are not independent due to the invariance properties of determinants under elementary row operations. To easily retain only unconstrained, nonredundant orbital parameters in the optimization, it is convenient to vary the orbital coefficients by performing rotations among the (occupied and virtual) orbitals with a unitary operator parametrized as an exponential of an anti-Hermitian operator. This parametrization is used in multiconfiguration self-consistent-field (MCSCF) calculations (for a recent and general review of MCSCF theory, see Ref. 30). More specifically, we use the following parametrization of the wave function depending on $N_{\mathrm{Jas}}^{\mathrm{opt}}$ Jastrow parameters $\boldsymbol{\alpha}$, $N_{\mathrm{CSF}}^{\mathrm{opt}}=N_{\mathrm{CSF}}-1$ free CSF coefficients $\mathbf{c}$ (c_1 is fixed), and $N_{\mathrm{orb}}^{\mathrm{opt}}$ orbital rotation parameters $\boldsymbol{\kappa}$,

$$|\Psi(\boldsymbol{\alpha},\mathbf{c},\boldsymbol{\kappa})\rangle = \hat{J}(\boldsymbol{\alpha})e^{\hat{\kappa}(\boldsymbol{\kappa})}\sum_{I=1}^{N_{\mathrm{CSF}}} c_I|C_I\rangle, \tag{4}$$

where $e^{\hat{\kappa}(\boldsymbol{\kappa})}$ is the unitary operator that performs rotations in orbital space (see, e.g., Refs. 31 and 32). More elaborate parametrizations of the CSF coefficients, such as a unitary parametrization,[33] are often used in the MCSCF theory (see, e.g., Ref. 32), but we have not found any decisive advantage to using them for our purpose.

The rotations in orbital space are generated by the anti-Hermitian real singlet orbital excitation operator[34]

$$\hat{\kappa}(\boldsymbol{\kappa}) = \sum_{k<l} \kappa_{kl}\hat{E}_{kl}^{-}, \tag{5}$$

where the sum is over all nonredundant orbital pairs, $\hat{E}_{kl}^{-}=\hat{E}_{kl}-\hat{E}_{lk}$, and $\hat{E}_{kl}=\hat{a}_{k\uparrow}^{\dagger}\hat{a}_{l\uparrow}+\hat{a}_{k\downarrow}^{\dagger}\hat{a}_{l\downarrow}$ is the singlet excitation operator from orbital l to orbital k. In Eq. (4), the action of the operator $e^{\hat{\kappa}(\boldsymbol{\kappa})}$ is to rotate each occupied orbital in the Slater determinants as

$$|\phi_k\rangle = e^{\hat{\kappa}(\boldsymbol{\kappa})}|\phi_k^0\rangle = \sum_l (e^{\boldsymbol{\kappa}})_{kl}|\phi_l^0\rangle, \tag{6}$$

where the sum is over all (occupied and virtual) orbitals, and $(e^{\boldsymbol{\kappa}})_{kl}$ are the elements of the orthogonal matrix $e^{\boldsymbol{\kappa}}$ constructed from the real antisymmetric matrix $\boldsymbol{\kappa}$ with elements κ_{kl}. More generally, any unitary matrix can be written as an exponential of an anti-Hermitian matrix, the off-diagonal upper triangular part of the anti-Hermitian matrix realizing a nonredundant parametrization of the unitary matrix. To maintain the orthonormality of the entire set of orbitals, the operator $e^{\hat{\kappa}(\boldsymbol{\kappa})}$ is applied to the virtual orbitals as well. For a single Slater determinant wave function, the orbitals can be partitioned into three sets referred to as *closed* (i.e., doubly occupied), *open* (i.e., singly occupied), and *virtual* (i.e., unoccupied). The nonredundant excitations to consider are then closed→open, closed→virtual, and open→virtual. For a multiconfiguration wave function, the orbitals can be partitioned into three sets referred to as *inactive* (i.e., occupied in all determinants), *active* (i.e., occupied in some determinants and unoccupied in the others), and *secondary* (i.e., unoccupied in all determinants). For a multiconfiguration complete active space (CAS) wave function,[35] the nonredundant excitations are then inactive→active, inactive→secondary, and active→secondary. For a single-determinant and multideterminant CAS wave function, the action of the reverse excitation from orbital k to l ($\hat{E}_{lk}$) in $\hat{E}_{kl}^{-}=\hat{E}_{kl}-\hat{E}_{lk}$ is always zero. For a general multiconfiguration wave function (not CAS), some active→active excitations must also be included. Consequently, the action of the reverse excitation $\hat{E}_{lk}$ in $\hat{E}_{kl}^{-}=\hat{E}_{kl}-\hat{E}_{lk}$ does not generally vanish. Only excitations between orbitals of the same spatial symmetry have to be considered. In the superconfiguration interaction approach[36] where the orbitals are optimized by adding the single excitations of the (multiconfiguration) reference wave function to the variational space, pioneered in QMC by Filippi and co-workers,[16,17] an alternative linear parametrization of the orbital space is chosen, $|\phi_k\rangle=(\hat{1}+\hat{\kappa}(\boldsymbol{\kappa}))|\phi_k^0\rangle$, instead of the unitary parametrization of Eq. (6). In that case, the optimized orbitals are not orthonormal.

In the following, we will collectively refer to the Jastrow, CSF, and orbital parameters as $\mathbf{p}=(\boldsymbol{\alpha},\mathbf{c},\boldsymbol{\kappa})$. The wave function of Eq. (1) is thus simply $|\Psi_0\rangle=|\Psi(\mathbf{p}^0)\rangle$, where $\mathbf{p}^0=(\boldsymbol{\alpha}^0,\mathbf{c}^0,\boldsymbol{\kappa}^0=\mathbf{0})$ are the current parameters. We will designate by $N^{\mathrm{opt}}=N_{\mathrm{Jas}}^{\mathrm{opt}}+N_{\mathrm{CSF}}^{\mathrm{opt}}+N_{\mathrm{orb}}^{\mathrm{opt}}$ the total number of parameters to be optimized.

C. First-order wave function derivatives

We now give the expressions for the first-order derivatives of the wave function $|\Psi(\mathbf{p})\rangle$ of Eq. (4) with respect to the parameters p_i at $\mathbf{p}=\mathbf{p}^0$,

$$|\Psi_i\rangle = \left(\frac{\partial|\Psi(\mathbf{p})\rangle}{\partial p_i}\right)_{\mathbf{p}=\mathbf{p}^0}, \tag{7}$$

which collectively designate the derivatives with respect to the Jastrow parameters

$$|\Psi_{\alpha_i}\rangle = \frac{\partial\hat{J}(\boldsymbol{\alpha}^0)}{\partial\alpha_i}|\Phi_0\rangle, \tag{8}$$

with respect to the CSF parameters

$$|\Psi_{c_I}\rangle = \hat{J}(\boldsymbol{\alpha}^0)|C_I\rangle, \tag{9}$$

and with respect to the orbital parameters

$$|\Psi_{\kappa_{kl}}\rangle = \hat{J}(\boldsymbol{\alpha}^0)\hat{E}_{kl}^{-}|\Phi_0\rangle. \tag{10}$$

The first-order orbital derivatives are thus generated by the single excitations of orbitals out of the state $|\Phi_0\rangle$.

D. Second-order wave function derivatives

The second-order derivatives with respect to the parameters p_i at $\mathbf{p}=\mathbf{p}^0$, which are needed only for the Newton method, are

$$|\Psi_{ij}\rangle = \left(\frac{\partial^2|\Psi(\mathbf{p})\rangle}{\partial p_i\partial p_j}\right)_{\mathbf{p}=\mathbf{p}^0}, \tag{11}$$

which collectively designate the Jastrow-Jastrow derivatives

$$|\Psi_{\alpha_i\alpha_j}\rangle = \frac{\partial^2\hat{J}(\boldsymbol{\alpha}^0)}{\partial\alpha_i\partial\alpha_j}|\Phi_0\rangle, \tag{12}$$

the Jastrow-CSF derivatives

$$|\Psi_{\alpha_i c_I}\rangle = \frac{\partial \hat{J}(\boldsymbol{\alpha}^0)}{\partial \alpha_i}|C_I\rangle, \tag{13}$$

the Jastrow-orbital derivatives

$$|\Psi_{\alpha_i \kappa_{kl}}\rangle = \frac{\partial \hat{J}(\boldsymbol{\alpha}^0)}{\partial \alpha_i}\hat{E}^-_{kl}|\Phi_0\rangle, \tag{14}$$

the CSF-orbital derivatives

$$|\Psi_{c_I \kappa_{kl}}\rangle = \hat{J}(\boldsymbol{\alpha}^0)\hat{E}^-_{kl}|C_I\rangle, \tag{15}$$

and the orbital-orbital derivatives

$$|\Psi_{\kappa_{kl}\kappa_{mn}}\rangle = \hat{J}(\boldsymbol{\alpha}^0)\hat{E}^-_{kl}\hat{E}^-_{mn}|\Phi_0\rangle. \tag{16}$$

Notice that the wave function form of Eq. (4) is linear in the CSF parameters and therefore the CSF-CSF derivatives are zero, $|\Psi_{c_I c_J}\rangle = 0$. The orbital-orbital derivatives correspond to double excitations of orbitals out of the state $|\Phi_0\rangle$. Since we usually start the optimization with reasonably good initial orbitals coming from a standard MCSCF calculation, we set these second derivatives to zero, $|\Psi_{\kappa_{kl}\kappa_{mn}}\rangle = 0$, in order to reduce the computational cost per iteration during Newton minimization. Nevertheless, it takes only a few steps to optimize the orbitals as discussed in Sec. VI.

III. ENERGY MINIMIZATION PROCEDURES

In this section, we present the three methods investigated in this work to minimize the variational energy with respect to the wave function parameters $\mathbf{p}$,

$$E = \min_{\mathbf{p}} E(\mathbf{p}), \tag{17}$$

where $E(\mathbf{p}) = \langle \Psi(\mathbf{p})|\hat{H}|\Psi(\mathbf{p})\rangle / \langle \Psi(\mathbf{p})|\Psi(\mathbf{p})\rangle$ and $\hat{H} = \hat{T} + \hat{W}_{\mathrm{ee}} + \hat{V}_{\mathrm{ne}}$ is the electronic Hamiltonian, including the kinetic, electron-electron interaction, and nuclei-electron interaction terms. The Hamiltonian can also include a nonlocal pseudopotential, enabling one to avoid the explicit treatment of core electrons. The energy corresponding to the current parameters $\mathbf{p}^0$ will be denoted by $E_0 = E(\mathbf{p}^0)$.

A. Newton optimization method

The Newton method was first applied to the optimization of QMC wave functions by Rappe and co-workers.[21,22] It has been considerably improved by Umrigar and Filippi,[23] and by Sorella,[24] by making use of a lower variance statistical estimator of the Hessian matrix and by employing stabilization techniques. In Ref. 23 the correct Hessian was used, whereas in Ref. 24 an approximate Hessian, which reduces to the exact Hessian for parameters that are linear in the exponent, was used. We now recall the basic working equations.

The energy $E(\mathbf{p})$ is expanded to second order in the parameters $\mathbf{p}$ around $\mathbf{p}^0$,

$$E^{[2]}(\mathbf{p}) = E_0 + \sum_{i=1}^{N^{\mathrm{opt}}} g_i \Delta p_i + \frac{1}{2}\sum_{i=1}^{N^{\mathrm{opt}}}\sum_{j=1}^{N^{\mathrm{opt}}} h_{ij}\Delta p_i \Delta p_j, \tag{18}$$

where the sums are over all the parameters to be optimized, $\Delta p_i = p_i - p_i^0$ are the components of the vector of parameter variations $\Delta\mathbf{p}$,

$$g_i = \left(\frac{\partial E(\mathbf{p})}{\partial p_i}\right)_{\mathbf{p}=\mathbf{p}^0} \tag{19}$$

are the components of the energy gradient vector $\mathbf{g}$, and

$$h_{ij} = \left(\frac{\partial^2 E(\mathbf{p})}{\partial p_i \partial p_j}\right)_{\mathbf{p}=\mathbf{p}^0} \tag{20}$$

are the elements of the energy Hessian matrix $\mathbf{h}$. Imposition of the stationary condition on the expanded energy expression, $\partial E^{[2]}(\mathbf{p})/\partial p_i = 0$, gives the following standard solution for the parameter variations:

$$\Delta\mathbf{p} = -\mathbf{h}^{-1}\cdot\mathbf{g}, \tag{21}$$

where $\mathbf{h}^{-1}$ is the inverse of the Hessian matrix. In practice, the energy gradient and Hessian are calculated in VMC with the statistical estimators given in Sec. IV A, yielding the parameter variations $\Delta\mathbf{p}$ of Eq. (21) that are used to update the current wave function, $|\Psi_0\rangle \rightarrow |\Psi(\mathbf{p}^0 + \Delta\mathbf{p})\rangle$. It simply remains to iterate until convergence.

Stabilization. As explained in Ref. 23, the stabilization of the Newton method is achieved by adding a positive constant, $a_{\mathrm{diag}} \geqslant 0$, to the diagonal of the Hessian matrix $\mathbf{h}$, i.e., $h_{ij} \rightarrow h_{ij} + a_{\mathrm{diag}}\delta_{ij}$. As a_{diag} is increased, the parameter variations $\Delta\mathbf{p}$ become smaller and rotate from the Newtonian direction to the steepest descent direction. A good value of a_{diag} is automatically determined at each iteration by performing three very short Monte Carlo calculations using correlated sampling with wave function parameters obtained with three trial values of a_{diag} and predicting by parabolic interpolation the value of a_{diag} that minimizes the energy[25] with some bounds imposed. The use of correlated sampling makes it possible to calculate energy differences with much smaller statistical error than the energies themselves. This procedure helps convergence if one is far from the minimum or if the statistical noise is large in the Monte Carlo evaluation of the gradient and Hessian.

We have found that adding in a multiple of the unit matrix to the Hessian as described above works well, but there exist other possible choices of positive definite matrices that could be added in. For instance, Sorella[24] adds in a multiple of the overlap matrix of the first-order derivatives of the wave function. Another possible choice is a multiple of the Levenberg-Marquardt approximation to the Hessian of the variance of the local energy.

B. Linear optimization method

The most straightforward way to energy optimize linear parameters in wave functions, such as the CSF parameters, is to diagonalize the Hamiltonian in the variational space that they define, leading to a generalized eigenvalue equation. This has been done in QMC, for example, in Refs. 11 and 37.

The linear method that we present now is an extension of the approach of Ref. 11 to arbitrary nonlinear parameters. This method is also presented in Ref. 25, using slightly different but equivalent conventions.

For notational convenience, we first introduce the normalized wave function

$$|\bar{\Psi}(\mathbf{p})\rangle = \frac{|\Psi(\mathbf{p})\rangle}{\sqrt{\langle\Psi(\mathbf{p})|\Psi(\mathbf{p})\rangle}}. \tag{22}$$

The idea is then to expand this normalized wave function $|\bar{\Psi}(\mathbf{p})\rangle$ to first order in the parameters $\mathbf{p}$ around the current parameters $\mathbf{p}^0$,

$$|\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})\rangle = |\Psi_0\rangle + \sum_{i=1}^{N^{\mathrm{opt}}} \Delta p_i |\bar{\Psi}_i\rangle, \tag{23}$$

where the wave function at $\mathbf{p}=\mathbf{p}^0$ is simply $|\bar{\Psi}(\mathbf{p}^0)\rangle = |\bar{\Psi}_0\rangle = |\Psi_0\rangle$ (chosen to be normalized to 1) and, for $i \geqslant 1$, $|\bar{\Psi}_i\rangle$ are the derivatives of $|\bar{\Psi}(\mathbf{p})\rangle$ that are orthogonal to $|\Psi_0\rangle$,

$$|\bar{\Psi}_i\rangle = \left(\frac{\partial|\bar{\Psi}(\mathbf{p})\rangle}{\partial p_i}\right)_{\mathbf{p}=\mathbf{p}^0} = |\Psi_i\rangle - S_{0i}|\Psi_0\rangle, \tag{24}$$

where $S_{0i}=\langle\Psi_0|\Psi_i\rangle$. The minimization of the energy calculated with this linear wave function

$$E_{\mathrm{lin}} = \min_{\mathbf{p}} E_{\mathrm{lin}}(\mathbf{p}), \tag{25}$$

where

$$E_{\mathrm{lin}}(\mathbf{p}) = \frac{\langle\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})|\hat{H}|\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})\rangle}{\langle\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})|\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})\rangle}, \tag{26}$$

leads to the stationary condition of the associated Lagrange function

$$\nabla_{\mathbf{p}}[\langle\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})|\hat{H}|\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})\rangle - E_{\mathrm{lin}}\langle\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})|\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})\rangle] = 0, \tag{27}$$

where E_{lin} acts as a Lagrange multiplier for the normalization condition. The Lagrange function is quadratic in $\mathbf{p}$ and Eq. (27) leads to the following generalized eigenvalue equation:

$$\bar{\mathbf{H}}\cdot\Delta\mathbf{p} = E_{\mathrm{lin}}\bar{\mathbf{S}}\cdot\Delta\mathbf{p}, \tag{28}$$

where $\bar{\mathbf{H}}$ is the matrix of the Hamiltonian $\hat{H}$ in the $(N^{\mathrm{opt}}+1)$-dimensional basis consisting of the current normalized wave function and its derivatives $\{|\bar{\Psi}_0\rangle, |\bar{\Psi}_1\rangle, |\bar{\Psi}_2\rangle, \ldots, |\bar{\Psi}_{N^{\mathrm{opt}}}\rangle\}$, with elements $\bar{H}_{ij}=\langle\bar{\Psi}_i|\hat{H}|\bar{\Psi}_j\rangle$, $\bar{\mathbf{S}}$ is the overlap matrix of this $(N^{\mathrm{opt}}+1)$-dimensional basis, with elements $\bar{S}_{ij}=\langle\bar{\Psi}_i|\bar{\Psi}_j\rangle$ (note that $\bar{S}_{00}=1$ and $\bar{S}_{i0}=\bar{S}_{0i}=0$ for $i\geqslant 1$), and $\Delta\mathbf{p}$ is the $(N^{\mathrm{opt}}+1)$-dimensional vector of parameter variations with $\Delta p_0=1$. The linear method consists of solving the generalized eigenvalue equation of Eq. (28), for the lowest (physically reasonable) eigenvalue and associated eigenvector denoted by $\Delta\bar{\mathbf{p}}$. The overlap and (nonsymmetric) Hamiltonian matrices are computed in VMC using the statistical estimators given in Sec. IV B. Although we focus here on the optimization of the ground-state wave function, solving Eq. (28) also gives upper bound estimates of excited state energies of states with the same spatial and spin symmetries.

However, there is an arbitrariness in the previously described procedure: we have found the parameter variations $\Delta\bar{\mathbf{p}}$ from the expansion of the wave function $|\bar{\Psi}(\mathbf{p})\rangle$ of Eq. (22), but another choice of the normalization of the wave function will lead to different parameter variations. To see that, consider a differently normalized wave function

$$|\bar{\bar{\Psi}}(\mathbf{p})\rangle = N(\mathbf{p})|\bar{\Psi}(\mathbf{p})\rangle, \tag{29}$$

where the normalization function $N(\mathbf{p})$ is chosen to satisfy $N(\mathbf{p}^0)=1$ so as to leave unchanged the normalization at $\mathbf{p}=\mathbf{p}^0$, i.e., $|\bar{\bar{\Psi}}(\mathbf{p}^0)\rangle = |\Psi_0\rangle$. The derivatives of this new wave function are

$$|\bar{\bar{\Psi}}_i\rangle = \left(\frac{\partial|\bar{\bar{\Psi}}(\mathbf{p})\rangle}{\partial p_i}\right)_{\mathbf{p}=\mathbf{p}^0} = |\bar{\Psi}_i\rangle + N_i|\Psi_0\rangle, \tag{30}$$

where $N_i=(\partial N(\mathbf{p})/\partial p_i)_{\mathbf{p}=\mathbf{p}^0}$, i.e., their projections onto the current wave function $|\Psi_0\rangle$ depend on the normalization. Consequently, the first-order expansion of this new wave function

$$|\bar{\bar{\Psi}}_{\mathrm{lin}}(\mathbf{p})\rangle = |\Psi_0\rangle + \sum_{i=1}^{N^{\mathrm{opt}}} \Delta p_i |\bar{\bar{\Psi}}_i\rangle, \tag{31}$$

leads, after optimization of the energy, to different optimal parameter variations $\Delta\bar{\bar{\mathbf{p}}}$. As the two wave functions $|\bar{\Psi}_{\mathrm{lin}}(\mathbf{p})\rangle$ and $|\bar{\bar{\Psi}}_{\mathrm{lin}}(\mathbf{p})\rangle$ lie in the same variational space, they must be proportional after minimization of the energy, which implies that the new optimal parameter variations $\Delta\bar{\bar{\mathbf{p}}}$ are actually related to the original optimal parameter variations $\Delta\bar{\mathbf{p}}$ by a uniform rescaling

$$\Delta\bar{\bar{\mathbf{p}}} = \frac{\Delta\bar{\mathbf{p}}}{1-\sum_{i=1}^{N^{\mathrm{opt}}} N_i \Delta\bar{p}_i}. \tag{32}$$

Any choice of normalization does not necessarily give good parameter variations. For the CSF parameters, it is obvious that the best choice is the normalization of the wave function of Eq. (4) in order to keep the linear dependence on these parameters, ensuring convergence of the linear method in a single step. This is achieved by choosing $|\bar{\bar{\Psi}}_i\rangle = |\Psi_i\rangle$ which gives

$$N_i = S_{i0} \quad \text{for linear parameters}. \tag{33}$$

For the nonlinear Jastrow and orbital parameters, several criteria are possible. We have found that a good one is to choose the normalization by imposing that, for the variation of the nonlinear parameters, each derivative $|\bar{\bar{\Psi}}_i\rangle$ is orthogonal to a linear combination of $|\Psi_0\rangle$ and $|\bar{\Psi}_{\mathrm{lin}}\rangle$, i.e., $\langle\bar{\bar{\Psi}}_i|\,\xi\Psi_0 + (1-\xi)\bar{\Psi}_{\mathrm{lin}}/\|\bar{\Psi}_{\mathrm{lin}}\|\rangle = 0$, where ξ is a constant between 0 and 1, resulting in

$$N_i = -\frac{(1-\xi)\Sigma_j^{\text{nonlin}}\Delta\bar{p}_j\bar{S}_{ij}}{(1-\xi)+\xi\sqrt{1+\Sigma_{j,k}^{\text{nonlin}}\Delta\bar{p}_j\Delta\bar{p}_k\bar{S}_{jk}}} \tag{34}$$

for nonlinear parameters, where the sums are only over the nonlinear Jastrow and orbital parameters. The simple choice $\xi=1$ first used by Sorella[18] in the context of the SR method leads in many cases to good parameter variations, but in some cases can result in parameter variations that are too large. The choice $\xi=0$ making the norm of the linear wave function change $\|\bar{\tilde{\Psi}}_{\text{lin}}-\Psi_0\|$ minimum is safer but in some cases can yield parameter variations that are too small. In those cases, the choice $\xi=1/2$, imposing $\|\bar{\tilde{\Psi}}_{\text{lin}}\|=\|\Psi_0\|$, avoids both too large and too small parameter variations. In particular, if $\Delta\bar{\mathbf{p}}=\infty$, meaning that $\bar{\Psi}_{\text{lin}}$ is orthogonal to Ψ_0, it follows from Eqs. (32) and (34) that $\Delta\bar{\mathbf{p}}$ is zero for $\xi=0$ but $\Delta\bar{\mathbf{p}}$ is nonzero and finite for $\xi=1/2$. In practice, all these three choices for ξ usually lead to a very rapid convergence of the nonlinear parameters. In contrast, choosing the original derivatives, i.e., $N_i=S_{i0}$, leads to slowly converging or diverging Jastrow parameters.

Stabilization. Similarly to the procedure used for the Newton method, we stabilize the linear method by adding a positive constant, $a_{\text{diag}}\geqslant 0$, to the diagonal of $\bar{\mathbf{H}}$ except for the first element, i.e., $\bar{H}_{ij}\rightarrow\bar{H}_{ij}+a_{\text{diag}}\delta_{ij}(1-\delta_{i0})$. Again, as a_{diag} becomes larger, the parameter variations $\Delta\mathbf{p}$ become smaller and rotate toward the steepest descent direction. The value of a_{diag} is then automatically adjusted in the course of the optimization in the same way as in the Newton method. Note that if instead we were to add a_{diag} to $\bar{\mathbf{S}}^{-1}\cdot\bar{\mathbf{H}}$, then it would be the "level-shift" parameter commonly used in diagonalization procedures. We prefer to add to $\bar{\mathbf{H}}$, in part, because it is not necessary to compute $\bar{\mathbf{S}}^{-1}\cdot\bar{\mathbf{H}}$ in order to solve Eq. (28).

Connection to the EFP method. The generalized eigenvalue equation of Eq. (28) can be rewritten as an eigenvalue equation $\bar{\mathbf{H}}'\cdot\Delta\mathbf{p}=E_{\text{lin}}\Delta\mathbf{p}$, where $\bar{\mathbf{H}}'=\bar{\mathbf{S}}^{-1}\cdot\bar{\mathbf{H}}$, i.e., with matrix elements $\bar{H}'_{ij}=\Sigma_{k=0}^{N^{\text{opt}}}(\bar{\mathbf{S}}^{-1})_{ik}\langle\bar{\Psi}_k|\hat{H}|\bar{\Psi}_j\rangle$. This form is useful to establish the connection with the EFP optimization method for the CSF and orbital parameters.[13,15,16] This latter approach consists of solving at each iteration the effective eigenvalue equation $\bar{\mathbf{H}}^{\text{EFP}}\cdot\Delta\mathbf{p}=E^{\text{EFP}}\Delta\mathbf{p}$, where the EFP effective Hamiltonian has matrix elements $\bar{H}_{ij}^{\text{EFP}}=\langle\Phi_i|\hat{H}|\Phi_i\rangle\delta_{ij}+\Sigma_{k=1}^{N^{\text{opt}}}(\bar{\mathbf{S}}^{-1})_{ik}\langle\bar{\Psi}_k|\hat{H}|\bar{\Psi}_0\rangle[(1-\delta_{i0})\delta_{0j}+\delta_{i0}(1-\delta_{0j})]$, where $|\Phi_i\rangle$ designates the current wave function and its derivatives without the Jastrow factor, i.e., $|\Psi_i\rangle=\hat{J}(\boldsymbol{\alpha}^0)|\Phi_i\rangle$, and $\langle\bar{\Psi}_k|\hat{H}|\bar{\Psi}_0\rangle$ are just the components of half the gradient of the energy. Hence, in the EFP method, only the off-diagonal elements in the first column and first row calculated from the components of the energy gradient are retained in $\bar{\mathbf{H}}^{\text{EFP}}$.

Connection to the Newton and SRH methods. In the linear method, the energy expression that is minimized at each iteration, $E_{\text{lin}}(\mathbf{p})$, contains all orders in the parameter variations because of the presence of the denominator in Eq. (26), though only the zeroth- and first-order terms match those of the expansion of the exact energy $E(\mathbf{p})$. In contrast, in the Newton method, the energy expression of Eq. (18), $E^{[2]}(\mathbf{p})$, is truncated at second order in $\Delta\mathbf{p}$ but is exact up to this order. Now, if instead of solving the generalized eigenvalue equation [Eq. (28)], one expands the energy expression of Eq. (26) to second order in $\Delta\mathbf{p}$, one recovers the Newton method with an approximate (symmetric) Hessian $h_{ij}^{\text{lin}}=\bar{H}_{ij}+\bar{H}_{ji}-2E_0\bar{S}_{ij}$ corresponding exactly to the SRH method with $\beta=0$ of Ref. 24. The SRH method is much less stable and converges more slowly than either our linear method or our Newton method for the systems studied here.

C. Perturbative optimization method

The perturbative method discussed next is identical to the perturbative EFP approach of Scemama and Filippi[17] for the optimization of the CSF and orbital parameters, provided that the same choice is made for the energy denominators (see below). We give here an alternate proof without introducing the concept of energy fluctuations that in principle extends the method to other kinds of parameters as well.

Instead of calculating the optimal linearized wave function $|\bar{\Psi}_{\text{lin}}\rangle$ by diagonalizing the Hamiltonian $\hat{H}$ in the subspace spanned by $\{|\bar{\Psi}_0\rangle,|\bar{\Psi}_1\rangle,|\bar{\Psi}_2\rangle,\ldots,|\bar{\Psi}_{N^{\text{opt}}}\rangle\}$, we formulate a nonorthogonal perturbation theory for $|\bar{\Psi}_{\text{lin}}\rangle$. The textbook formulation of perturbation theory starts from the Hamiltonian $\hat{H}$ whose eigenstates we wish to compute and a zeroth-order Hamiltonian $\hat{H}^{(0)}$ whose eigenstates are known. Instead, here we start with $\hat{H}$ and the states $\{|\bar{\Psi}_0\rangle,|\bar{\Psi}_1\rangle,|\bar{\Psi}_2\rangle,\ldots,|\bar{\Psi}_{N^{\text{opt}}}\rangle\}$ and define a zeroth-order operator $\hat{H}^{(0)}$ for which these states are right eigenstates. To do this, we introduce $\{\langle\tilde{\Psi}_i|\}$, the dual (biorthonormal) basis of the basis $\{|\bar{\Psi}_i\rangle\}$, i.e., $\langle\tilde{\Psi}_i|\bar{\Psi}_j\rangle=\delta_{ij}$, given by (see, e.g., Ref. 38)

$$\langle\tilde{\Psi}_i| = \sum_{j=0}^{N^{\text{opt}}}(\bar{\mathbf{S}}^{-1})_{ij}\langle\bar{\Psi}_j|, \tag{35}$$

where $(\bar{\mathbf{S}}^{-1})_{ij}$ are the elements of the inverse of the overlap matrix $\bar{\mathbf{S}}$, and we introduce the non-Hermitian projector operator onto this subspace

$$\hat{P} = \sum_{i=0}^{N^{\text{opt}}}|\bar{\Psi}_i\rangle\langle\tilde{\Psi}_i|. \tag{36}$$

The optimal linearized wave function, minimizing the energy [Eq. (25)], satisfies the projected Schrödinger equation

$$\hat{P}\hat{H}|\bar{\Psi}_{\text{lin}}\rangle = E_{\text{lin}}\hat{P}|\bar{\Psi}_{\text{lin}}\rangle, \tag{37}$$

with the normalization condition $\langle\tilde{\Psi}_0|\bar{\Psi}_{\text{lin}}\rangle=1$, ensuring that the coefficient of $|\bar{\Psi}_{\text{lin}}\rangle$ on $|\bar{\Psi}_0\rangle=|\Psi_0\rangle$ is 1 as in Eq. (23).

To construct the perturbation theory, we now introduce a fictitious projected Schrödinger equation depending on a coupling constant λ,

$$\hat{P}\hat{H}^{\lambda}|\bar{\Psi}_{\rm lin}^{\lambda}\rangle = E_{\rm lin}^{\lambda}\hat{P}|\bar{\Psi}_{\rm lin}^{\lambda}\rangle, \tag{38}$$

with the normalization condition $\langle\tilde{\Psi}_0|\bar{\Psi}_{\rm lin}^{\lambda}\rangle=1$ for all λ, so that, for $\lambda=1$, Eq. (38) reduces to Eq. (37), $\hat{H}^{\lambda=1}=\hat{H}$, $|\bar{\Psi}_{\rm lin}^{\lambda=1}\rangle=|\bar{\Psi}_{\rm lin}\rangle$, $E_{\rm lin}^{\lambda=1}=E_{\rm lin}$, and we partition the Hamiltonian $\hat{H}^{\lambda}$ as follows:

$$\hat{H}^{\lambda} = \hat{H}^{(0)} + \lambda\hat{H}^{(1)}. \tag{39}$$

In this expression, $\hat{H}^{(0)}$ is a zeroth-order non-Hermitian operator

$$\hat{H}^{(0)} = \sum_{i=0}^{N^{\rm opt}} \mathcal{E}_i|\bar{\Psi}_i\rangle\langle\tilde{\Psi}_i|, \tag{40}$$

where $\mathcal{E}_i$ are *arbitrary* energies. Clearly, $\hat{H}^{(0)}$ admits $|\bar{\Psi}_i\rangle$ as right eigenstate and $\langle\tilde{\Psi}_i|$ as left eigenstate, with common eigenvalue $\mathcal{E}_i$. The non-Hermitian perturbation operator is obviously defined as $\hat{H}^{(1)}=\hat{H}-\hat{H}^{(0)}$. We expand $|\bar{\Psi}_{\rm lin}^{\lambda}\rangle$ and $E_{\rm lin}^{\lambda}$ in powers of λ: $|\bar{\Psi}_{\rm lin}^{\lambda}\rangle=\Sigma_{k=0}^{\infty}\lambda^k|\bar{\Psi}_{\rm lin}^{(k)}\rangle$ and $E_{\rm lin}^{\lambda}=\Sigma_{k=0}^{\infty}\lambda^k E_{\rm lin}^{(k)}$. The zeroth-order (right) eigenstate and energy are simply $|\bar{\Psi}_{\rm lin}^{(0)}\rangle=|\bar{\Psi}_0\rangle$ and $E_{\rm lin}^{(0)}=\mathcal{E}_0$. The first-order correction to the wave function is determined by the equation

$$\hat{P}(\hat{H}^{(0)}-\mathcal{E}_0)|\bar{\Psi}_{\rm lin}^{(1)}\rangle = -\hat{P}(\hat{H}^{(1)}-E_{\rm lin}^{(1)})|\bar{\Psi}_0\rangle. \tag{41}$$

To solve this equation, we define the non-Hermitian projector operator $\hat{R}=\Sigma_{i=1}^{N^{\rm opt}}|\bar{\Psi}_i\rangle\langle\tilde{\Psi}_i|$ which, in comparison with the projector $\hat{P}$, also removes the component parallel to $|\bar{\Psi}_0\rangle$. Note that $\hat{R}\hat{P}=\hat{R}$, $\hat{R}$ commutes with $\hat{H}^{(0)}-\mathcal{E}_0$ and $\hat{R}|\bar{\Psi}_{\rm lin}^{(1)}\rangle=|\bar{\Psi}_{\rm lin}^{(1)}\rangle$ (since $\langle\tilde{\Psi}_0|\bar{\Psi}_{\rm lin}^{\lambda}\rangle=1$ and $\langle\tilde{\Psi}_0|\bar{\Psi}_0\rangle=1$, implying $\langle\tilde{\Psi}_0|\bar{\Psi}_{\rm lin}^{(1)}\rangle=0$), so that applying $\hat{R}$ on Eq. (41) leads to

$$\begin{aligned}|\bar{\Psi}_{\rm lin}^{(1)}\rangle &= -\frac{\hat{R}}{\hat{H}^{(0)}-\mathcal{E}_0}(\hat{H}^{(1)}-E_{\rm lin}^{(1)})|\bar{\Psi}_0\rangle \\ &= -\sum_{i=1}^{N^{\rm opt}}|\bar{\Psi}_i\rangle\frac{\langle\tilde{\Psi}_i|\hat{H}-\mathcal{E}_0-E_{\rm lin}^{(1)}|\bar{\Psi}_0\rangle}{\mathcal{E}_i-\mathcal{E}_0} \\ &= -\sum_{i=1}^{N^{\rm opt}}\sum_{j=1}^{N^{\rm opt}}(\bar{\mathbf{S}}^{-1})_{ij}\frac{\langle\bar{\Psi}_j|\hat{H}|\bar{\Psi}_0\rangle}{\mathcal{E}_i-\mathcal{E}_0}|\bar{\Psi}_i\rangle,\end{aligned} \tag{42}$$

where $\mathcal{E}_0$ and $E_{\rm lin}^{(1)}$ in the numerator and the term $j=0$ have been dropped since, for $i\neq 0$, $\langle\tilde{\Psi}_i|\bar{\Psi}_0\rangle=0$ and $(\bar{\mathbf{S}}^{-1})_{i0}=0$, respectively. Therefore, the parameter variations in this first-order perturbation theory are

$$\Delta\bar{p}_i^{(1)} = -\frac{1}{\Delta\mathcal{E}_i}\sum_{j=1}^{N^{\rm opt}}(\bar{\mathbf{S}}^{-1})_{ij}\bar{H}_{j0}, \tag{43}$$

where $\bar{H}_{j0}=\langle\bar{\Psi}_j|\hat{H}|\bar{\Psi}_0\rangle=\langle\Psi_j|\hat{H}-E_0|\Psi_0\rangle=g_j/2$ is just half the gradient of the energy and $\Delta\mathcal{E}_i=\mathcal{E}_i-\mathcal{E}_0$. The perturbative method consists of calculating the parameter variations $\Delta\bar{\mathbf{p}}^{(1)}$ according to Eq. (43), updating the current wave function, $|\Psi_0\rangle\rightarrow|\Psi(\mathbf{p}^0+\Delta\bar{\mathbf{p}}^{(1)})\rangle$, and iterating until convergence. It is apparent from Eq. (43) that the perturbative method can be viewed as the Newton method with an approximate Hessian, $h_{ij}^{\rm pert}=(\bar{\mathbf{S}}^{-1})_{ij}/\Delta\mathcal{E}_i$, as also noted in Ref. 17.

The energy denominators $\Delta\mathcal{E}_i$ in Eq. (43) remain to be chosen. Since perturbation theory works best when $\hat{H}^{(0)}$ is "close" to $\hat{H}$, we choose $\hat{H}^{(0)}$ to have the same diagonal elements as $\hat{H}$, resulting in

$$\Delta\mathcal{E}_i = \frac{\langle\bar{\Psi}_i|\hat{H}|\bar{\Psi}_i\rangle}{\langle\bar{\Psi}_i|\bar{\Psi}_i\rangle} - E_0 = \frac{\bar{H}_{ii}}{\bar{S}_{ii}} - \bar{H}_{00}. \tag{44}$$

In practice, only rough estimates of the $\Delta\mathcal{E}_i$'s are necessary for the optimization so that one can compute them for just the initial iteration and keep them fixed for the following iterations. Therefore, for these iterations, only the inverse overlap matrix, $\bar{\mathbf{S}}^{-1}$, and the gradient of the energy, $g_j=2\bar{H}_{j0}$, need to be calculated in the perturbative method, leading to an important computational speedup per iteration in comparison with the linear method.

Stabilization. Similarly to the linear method, the perturbative method can be stabilized by adding an adjustable positive constant, $a_{\rm diag}\geqslant 0$, to the energy denominators, i.e., $\Delta\mathcal{E}_i\rightarrow\Delta\mathcal{E}_i+a_{\rm diag}$, which has the effect of decreasing the parameter variations $\Delta\bar{\mathbf{p}}^{(1)}$.

Connection to the perturbative EFP and SR methods. For the CSF and orbital parameters, if the energy denominators are chosen to be $\Delta\mathcal{E}_i=\langle\Phi_i|\hat{H}|\Phi_i\rangle/\langle\Phi_i|\Phi_i\rangle-\langle\Phi_0|\hat{H}|\Phi_0\rangle/\langle\Phi_0|\Phi_0\rangle$ (i.e., without the Jastrow factor), Eq. (43) exactly reduces to the perturbative EFP method.[17] Also, Eq. (43) reduces to the SR optimization method[18–20] if the energy denominators are all chosen equal, $\Delta\mathcal{E}_i=\Delta\mathcal{E}$ for all i.

IV. VARIATIONAL MONTE CARLO REALIZATION

When the previously described energy minimization procedures are implemented in VMC, it is important to pay attention to the statistical fluctuations. Expressions that are equivalent in the limit of an infinite Monte Carlo sample can, in fact, have very different statistical errors for a finite sample. We provide prescriptions for low variance estimators in this section.

We also note that, in order to reduce round-off noise, it can help to rescale the elements of the gradient vector, and the hessian, Hamiltonian and overlap matrices using the square root of the diagonal of overlap matrix.

At each step of the optimization, the quantum-mechanical averages are computed by sampling the probability density of the current wave function $\Psi_0(\mathbf{R})^2$. We will denote the statistical average of a local quantity, $f(\mathbf{R})$, by $\langle f(\mathbf{R})\rangle=(1/M)\Sigma_{k=1}^{M}f(\mathbf{R}_k)$, where the M electron configurations $\mathbf{R}_k$ are sampled from $\Psi_0(\mathbf{R})^2$.

A. Energy gradient and Hessian

In terms of the derivatives $\Psi_i(\mathbf{R})$ of the wave function of Eq. (4), and using the Hermiticity of the Hamiltonian $\hat{H}$, an estimator of the energy gradient is[39]

$$g_i = 2\left[\left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} E_L(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \langle E_L(\mathbf{R}) \rangle \right], \tag{45}$$

where $E_L(\mathbf{R}) = [H(\mathbf{R})\Psi_0(\mathbf{R})]/\Psi_0(\mathbf{R})$ is the local energy. In the limit that $\Psi_0(\mathbf{R})$ is an exact eigenfunction, the local energy becomes constant, $E_L(\mathbf{R}) = E_{\text{exact}}$ for all $\mathbf{R}$, and thus the gradient of Eq. (45) vanishes with zero variance. This leads to the following zero-variance principle for the Newton and perturbative methods: in the limit that $\Psi_0(\mathbf{R})$ is an exact eigenfunction, the parameter variations of Eqs. (21) and (43) vanish with zero variance.

Taking the derivative of Eq. (45) leads to the straightforward estimator of the energy Hessian of Lin, Zhang, and Rappe (LZR),[21]

$$h_{ij}^{\text{LZR}} = A_{ij} + B_{ij} + C_{ij}, \tag{46}$$

where

$$A_{ij} = 2\left[\left\langle \frac{\Psi_{ij}(\mathbf{R})}{\Psi_0(\mathbf{R})} E_L(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_{ij}(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \langle E_L(\mathbf{R}) \rangle - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} E_L(\mathbf{R}) \right\rangle + \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \langle E_L(\mathbf{R}) \rangle \right], \tag{47}$$

involving the second derivatives $\Psi_{ij}(\mathbf{R})$ of the wave function,

$$\begin{aligned} B_{ij} &= 4\left[\left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} E_L(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \times \langle E_L(\mathbf{R}) \rangle \right] - 2\left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle g_j - 2\left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle g_i \\ &= 4\left\langle \left(\frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \right) \times \left(\frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} - \left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \right) (E_L(\mathbf{R}) - \langle E_L(\mathbf{R}) \rangle) \right\rangle, \end{aligned} \tag{48}$$

and

$$C_{ij} = 2\left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} E_{L,j}(\mathbf{R}) \right\rangle, \tag{49}$$

where

$$E_{L,j}(\mathbf{R}) = [H(\mathbf{R})\Psi_j(\mathbf{R})]/\Psi_0(\mathbf{R}) - [\Psi_j(\mathbf{R})/\Psi_0(\mathbf{R})]E_L(\mathbf{R})$$

is the derivative of the local energy with respect to parameter j. In this estimator of the Hessian, the term that fluctuates the most is C_{ij}.

Umrigar and Filippi[23] observed that the fluctuations of a covariance $\langle ab \rangle - \langle a \rangle \langle b \rangle$ are much smaller than those of $\langle ab \rangle$ if the fluctuations of a are much smaller than the average of a, i.e., $\sqrt{\langle a^2 \rangle - \langle a \rangle^2} \ll |\langle a \rangle|$, and a is not strongly correlated with $1/b$. In Eq. (49), $\Psi_i(\mathbf{R})/\Psi_0(\mathbf{R})$ is always of the same sign for parameters in the exponent and in practice its fluctuations are much smaller than its average. Furthermore, it follows from the Hermiticity of the Hamiltonian that $\langle E_{L,j}(\mathbf{R}) \rangle$ vanishes in the limit of an infinite sample.[21] Using these two observations, Umrigar and Filippi[23] provided an estimator of the Hessian,

$$h_{ij}^{\text{UF}} = A_{ij} + B_{ij} + D_{ij}, \tag{50}$$

that fluctuates much less than the straightforward LZR estimator, where the symmetrized estimator,

$$D_{ij} = \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} E_{L,j}(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \langle E_{L,j}(\mathbf{R}) \rangle + \left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} E_{L,i}(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \langle E_{L,i}(\mathbf{R}) \rangle, \tag{51}$$

has the same average as the term C_{ij} in the limit of an infinite sample, but being a covariance has much smaller fluctuations. We note that A_{ij} is already a covariance and B_{ij} is a tricovariance.

Although the A_{ij} and B_{ij} terms vanish with zero variance in the limit that $\Psi_0(\mathbf{R})$ is an exact eigenfunction (the D_{ij} term does not), in practice for the Jastrow parameters, far from the minimum, the B_{ij} fluctuates more than the D_{ij} term for the Jastrow parameters in the Hessian of Eq. (50). With the form of the Jastrow factors that we use, we have observed that the ratio $(B_{ij}+D_{ij})/D_{ij}$ is roughly independent of i and j for most i and j though it changes during the Monte Carlo iterations. It is typically between 1.2 and 2.5 at the initial iteration and between 0.9 and 1.1 at the final iteration. We exploit this to decrease the fluctuations by defining a new, approximate Hessian partially averaged over the Jastrow parameters

$$h_{ij}^{\text{TU}} = A_{ij} + \frac{\langle\langle |B_{ij} + D_{ij}| \rangle\rangle}{\langle\langle |D_{ij}| \rangle\rangle} D_{ij}, \tag{52}$$

where TU are the initials of the present authors, and the average over the Jastrow parameter pairs are defined by $\langle\langle X_{ij} \rangle\rangle = (2/N_{\text{Jas}}^{\text{opt}}(N_{\text{Jas}}^{\text{opt}}+1)) \Sigma_{i=1}^{N_{\text{Jas}}^{\text{opt}}} \Sigma_{j=i}^{N_{\text{Jas}}^{\text{opt}}} X_{ij}$. The average is calculated as $\langle\langle |B_{ij}+D_{ij}| \rangle\rangle / \langle\langle |D_{ij}| \rangle\rangle$ and not as $\langle\langle (|B_{ij}+D_{ij}|)/|D_{ij}| \rangle\rangle$ to avoid possible numerical divergences of this ratio for small D_{ij}. In Eq. (52), i and j refer only to Jastrow parameters. For all the terms related to the other parameters (including all the mixed terms), the Hessian of Eq. (50) is used without further modification.

Exact or approximate wave functions such as $\Psi_0(\mathbf{R})$ go linearly to zero with the distance d between $\mathbf{R}$ and their nodal hypersurface, i.e., $\Psi_0(\mathbf{R}) \sim d$ for $d \to 0$. The local energy $E_L(\mathbf{R})$ generally diverges as $1/d$ for $d \to 0$ for approximate wave functions. In contrast to the case of the Jastrow parameters, the derivatives $\Psi_i(\mathbf{R})$ for the CSF and orbital parameters have a different nodal hypersurface than $\Psi_0(\mathbf{R})$ and the ratio $\Psi_i(\mathbf{R})/\Psi_0(\mathbf{R})$ thus also diverges as $1/d$, even if the wave function $\Psi_0(\mathbf{R})$ is exact. Consequently, the derivative of the local energy $E_{L,i}(\mathbf{R})$ generally diverges as $1/d^2$ for approximate wave functions. In the expression of the Hessian, the leading divergence at the nodes of the approximate wave function $\Psi_0(\mathbf{R})$ thus comes from the terms $(\Psi_i(\mathbf{R})/\Psi_0(\mathbf{R}))(\Psi_j(\mathbf{R})/\Psi_0(\mathbf{R}))E_L(\mathbf{R})$,

$(\Psi_i(\mathbf{R})/\Psi_0(\mathbf{R}))E_{L,j}(\mathbf{R})$, and $(\Psi_j(\mathbf{R})/\Psi_0(\mathbf{R}))E_{L,i}(\mathbf{R})$ that behave as $1/d^3$. It is, however, easy to check that these third-order divergences cancel exactly in Eq. (50).

B. Overlap and Hamiltonian matrices

The elements of the symmetric overlap matrix $\bar{\mathbf{S}}$ are

$$\bar{S}_{00} = 1 \tag{53a}$$

and, for $i,j>0$,

$$\bar{S}_{i0} = \bar{S}_{0j} = 0, \tag{53b}$$

and

$$\bar{S}_{ij} = \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle. \tag{53c}$$

The elements of the Hamiltonian matrix $\bar{\mathbf{H}}$ are

$$\bar{H}_{00} = \langle E_L(\mathbf{R}) \rangle, \tag{54a}$$

and, for $i,j>0$,

$$\bar{H}_{i0} = \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} E_L(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \langle E_L(\mathbf{R}) \rangle, \tag{54b}$$

$$\bar{H}_{0j} = \left[\left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} E_L(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \langle E_L(\mathbf{R}) \rangle \right] + \langle E_{L,j}(\mathbf{R}) \rangle, \tag{54c}$$

which are two estimators of half of the energy gradient, and

$$\bar{H}_{ij} = \left[\left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} E_L(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} E_L(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} E_L(\mathbf{R}) \right\rangle + \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \left\langle \frac{\Psi_j(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \langle E_L(\mathbf{R}) \rangle \right] + \left[\left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} E_{L,j}(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi_0(\mathbf{R})} \right\rangle \langle E_{L,j}(\mathbf{R}) \rangle \right]. \tag{54d}$$

We do not use the Hermiticity of the Hamiltonian $\hat{H}$ to symmetrize the matrix $\bar{\mathbf{H}}$. In fact, as shown by Nightingale and Melik-Alaverdian,[11] using the nonsymmetric matrix $\bar{\mathbf{H}}$ of Eqs. (54) leads to a stronger zero-variance principle than the one previously described for the Newton and perturbative methods: in the limit that the states $\{|\bar{\Psi}_0\rangle, |\bar{\Psi}_1\rangle, |\bar{\Psi}_2\rangle, \ldots, |\bar{\Psi}_{N^{\mathrm{opt}}}\rangle\}$ span an invariant subspace of the Hamiltonian $\hat{H}$, i.e., in the limit that the linear wave function $\bar{\Psi}_{\mathrm{lin}}(\mathbf{R})$ of Eq. (23) after optimization is an exact eigenfunction, the matrix $\bar{\mathbf{S}}^{-1} \cdot \bar{\mathbf{H}}$ and consequently the eigenvector solution $\Delta\bar{\mathbf{p}}$ have zero variance. In practice, even if we do not work in an invariant subspace of $\hat{H}$, using the nonsymmetric matrix $\bar{\mathbf{H}}$ leads to smaller statistical errors on a finite sample than using its symmetrized analog. Although in principle diagonalization of a nonsymmetric matrix leads to complex eigenvalues, in practice the physically reasonable (i.e., with large overlap with the current wave function) lowest eigenvectors have usually real eigenvalues. Of course, in the limit of an infinite sample $M \to \infty$ a symmetric matrix $\bar{\mathbf{H}}$ is recovered.

As noted in the previous subsection for the Hessian, although the terms $(\Psi_i(\mathbf{R})/\Psi_0(\mathbf{R}))(\Psi_j(\mathbf{R})/\Psi_0(\mathbf{R}))E_L(\mathbf{R})$ and $(\Psi_i(\mathbf{R})/\Psi_0(\mathbf{R}))E_{L,j}(\mathbf{R})$ in the expression of the Hamiltonian matrix of Eq. (54d) display a third-order divergence $1/d^3$ as the distance d between $\mathbf{R}$ and the nodal hypersurface of $\Psi_0(\mathbf{R})$ goes to zero, again these divergences cancel exactly.

C. Comparison of computational cost per iteration

At each optimization iteration, besides the calculation the current wave function $\Psi_0(\mathbf{R})$ and the local energy $E_L(\mathbf{R})$, the Newton method requires the computation of the first-order and second-order wave function derivatives, $\Psi_i(\mathbf{R})$ and $\Psi_{ij}(\mathbf{R})$, and the first-order derivatives of the local energy $E_{L,i}(\mathbf{R})$. The linear method requires the calculation of $\Psi_i(\mathbf{R})$ and $E_{L,i}(\mathbf{R})$ but not of the second-order derivatives of the wave function with respect to the parameters. In principle, this decreases the computational cost per iteration, especially if the many orbital-orbital second-order derivatives were to be computed in the Newton method. In practice, since our implementation of the Newton method neglects these orbital-orbital derivatives, the computational cost per iteration of the Newton and linear methods is very similar.

The perturbative method requires the computation of the same quantities as the linear method. However, since the method is not very sensitive to having accurate energy denominators $\Delta\mathcal{E}_i$ in Eq. (43), and since the energy denominators do not undergo large changes from iteration to iteration, we compute these for the first iteration only. Hence it is not necessary to compute $E_{L,i}(\mathbf{R})$ for subsequent iterations. This leads to a computational speedup per iteration in comparison with the linear method. The precise speedup factor depends on the wave function used; typically, for the systems studied here, we have found factors ranging from about 1.5 for a single-determinant wave function to 5.5 for the largest multideterminant wave function considered, for the iterations for which the $\Delta\mathcal{E}_i$'s are not computed.

V. COMPUTATIONAL DETAILS

We illustrate the optimization methods by calculating the ground-state electronic energy of the all-electron C_2 molecule at the experimental equilibrium interatomic distance of 2.3481 bohr.[40] The ground-state wave function is of symmetry $^1\Sigma_g^+$ in the point group $D_{\infty h}$. The estimated exact, infinite nuclear mass, nonrelativistic electronic energy is −75.9265(8) hartree,[41] where the number in parentheses is an estimate of the uncertainty in the last digit. This system has a strong multiconfiguration character due to the energetic near degeneracy of the valence orbitals, making it a challenging system despite its small size.

We start by generating a standard *ab initio* wave function using the quantum chemistry program GAMESS,[42] typically a restricted Hartree-Fock (RHF) wave function or a MCSCF wave function, using the symmetry point group D_{4h} which is the largest subgroup of $D_{\infty h}$ available in GAMESS. We use the uncontracted Slater basis set form of Clementi and Roetti,[43] with exponents reoptimized at the RHF level by Koga *et al.*[44] For carbon, the basis set contains two $1s$, three $2s$, one $3s$, and four $2p$ Slater functions, that are each approximated by a fit to six Gaussian functions[45,46] in GAMESS. Specifically, we consider the following *ab initio* wave functions: a RHF wave function, with orbital occupations $1\sigma_g^2 1\sigma_u^2 2\sigma_g^2 2\sigma_u^2 1\pi_{u,x}^2 1\pi_{u,y}^2$; a CAS(8,5) wave function, containing 6 CSFs in D_{4h} symmetry made of 7 Slater determinants generated by distributing the eight valence electrons over the five active valence orbitals $2\sigma_g 2\sigma_u 1\pi_{u,x} 1\pi_{u,y} 3\sigma_g$; a CAS(8,7) wave function, containing 80 CSFs made of 165 determinants with the seven active orbitals $2\sigma_g 2\sigma_u 1\pi_{u,x} 1\pi_{u,y} 3\sigma_g 1\pi_{g,x} 1\pi_{g,y}$; a CAS(8,8) wave function, containing 264 CSFs made of 660 determinants with the eight active orbitals $2\sigma_g 2\sigma_u 1\pi_{u,x} 1\pi_{u,y} 3\sigma_g 1\pi_{g,x} 1\pi_{g,y} 3\sigma_u$, i.e., all the valence orbitals originating from the $n=2$ shell of the C atoms. In addition, we construct a larger one-electron basis set by adding to the basis of Koga *et al.*, one d function with an exponent of 2.13 optimized in RHF, and we consider a wave function obtained from a restricted active space (RAS) calculation in this basis that would correspond to a CAS(8,26) calculation, using all the 26 orbitals originating from the $n=2$ and $n=3$ shells of the C atoms, but where only single (S), double (D), triple (T), and quadruple (Q) excitations are allowed in the active space. This wave function, that we denote by RAS-SDTQ(8,26), contains 110 481 CSFs made of 411 225 determinants.

The standard *ab initio* wave function is then multiplied by a Jastrow factor, imposing the electron-electron cusp conditions, but with essentially all other free parameters chosen to be 0 to form our starting trial wave function. QMC calculations are performed with the program CHAMP,[47] using this time the true Slater basis set rather than its Gaussian expansion. In comparison with GAMESS, additional symmetries outside the point group D_{4h} are detected numerically which allows one to reduce the numbers of CSFs to 5, 50, and 165 for the CAS(8,5), CAS(8,7), and CAS(8,8) wave functions, respectively. For the large RAS-SDTQ(8,26) wave function, only a fraction of all the CSFs are retained in QMC by applying a variable cutoff on the CSF coefficients and an extrapolation procedure is used to estimate the QMC result if all the CSFs had been included (see Sec. VI E). For the orbital optimization, only the single excitations between orbitals of the same irreducible representation of $D_{\infty h}$ are generated. We, however, impose no restriction inside each of the two-dimensional irreducible representations π_u and π_g. Although one can in principle identify the π_x and π_y components and forbid excitations between these two components to further reduce the number of free parameters, these redundancies appear to cause no problem in practice during the optimization. Also, we impose the electron-nucleus cusp condition on each orbital. The parameters of the trial wave function are optimized by the previously described energy minimization procedures in VMC, using a very efficient accelerated Metropolis algorithm,[48,49] allowing us to simultaneously make large Monte Carlo moves in configuration space and have a high acceptance probability. Once a trial wave function has been optimized, we perform a DMC calculation within the fixed-node and the short-time approximations (see, e.g., Refs. 50–53). We use an imaginary time step of $\tau=0.01$ hartree^{-1} in an efficient DMC algorithm featuring very small time-step errors,[54] so that the accuracy is essentially limited by the quality of the nodal hypersurface of the trial wave function.

VI. RESULTS AND DISCUSSION

A. Optimization of the Jastrow factor

We first study the convergence behavior of the energy minimization methods for the separate optimization of the Jastrow, CSF, and orbital parameters. To facilitate comparisons, we apply the VMC optimization procedures with a common fixed statistical error of the energy at each step, namely, 0.5 mhartree. This is not the usual way in which we routinely perform optimizations which is described later in Sec. VI D.

Figure 1 shows the convergence of the total VMC energy during the optimization of the 24 Jastrow parameters in a wave function composed of the RHF Slater determinant multiplied by a Jastrow factor. The linear, perturbative, and Newton methods are compared. For the Newton method, we present the results obtained with the UF Hessian of Eq. (50), already used in Ref. 23, and with the TU Hessian of Eq. (52). To compare the fluctuations of these two Hessians, we have computed the quantity $\eta=1/N^{\mathrm{opt}}(N^{\mathrm{opt}}+1)\sum_{i=1}^{N^{\mathrm{opt}}}\sum_{j=i}^{N^{\mathrm{opt}}}(\sigma(h_{ij}))^2$, where $(\sigma(h_{ij}))^2$ is the variance of the element h_{ij} of the Hessian averaged over 100 Monte Carlo configurations. For the initial iteration of the optimization, far from the energy minimum, the UF Hessian fluctuates more than the TU Hessian by a factor of $\eta^{\mathrm{UF}}/\eta^{\mathrm{TU}}=3.6$. For comparison, the LZR Hessian of Eq. (46) fluctuates more than the TU Hessian by a factor of $\eta^{\mathrm{LZR}}/\eta^{\mathrm{TU}}=150$, more than two orders of magnitude larger even for this modest system. Near the energy minimum, the factors are $\eta^{\mathrm{UF}}/\eta^{\mathrm{TU}}=3.3$ and $\eta^{\mathrm{LZR}}/\eta^{\mathrm{TU}}=600$. These factors tend to increase with the system size. The Newton method with the UF Hessian converges reasonably fast in about six iterations, which is a little faster than the convergence shown in Figs. 1, 2, and 4 of Ref. 23 due to the previously described correlated sampling adjustment of the

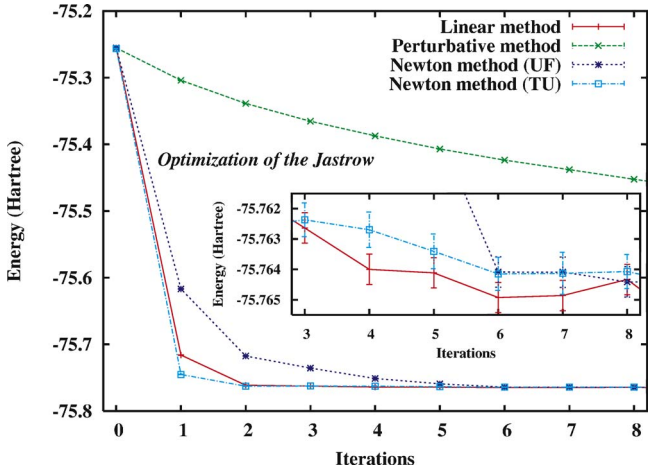

FIG. 1. Convergence of the VMC total energy E_{VMC} of the all-electron C_2 molecule during the optimization of the 24 Jastrow parameters in a wave function composed of the RHF Slater determinant multiplied by a Jastrow factor. The linear, perturbative, and Newton energy minimization methods are compared. For the Newton method, the results obtained with the UF Hessian of Eq. (50) and the TU Hessian of Eq. (52) are shown. The statistical error on the energy at each iteration is 0.5 mhartree. The inset is an enlargement of the last six iterations.

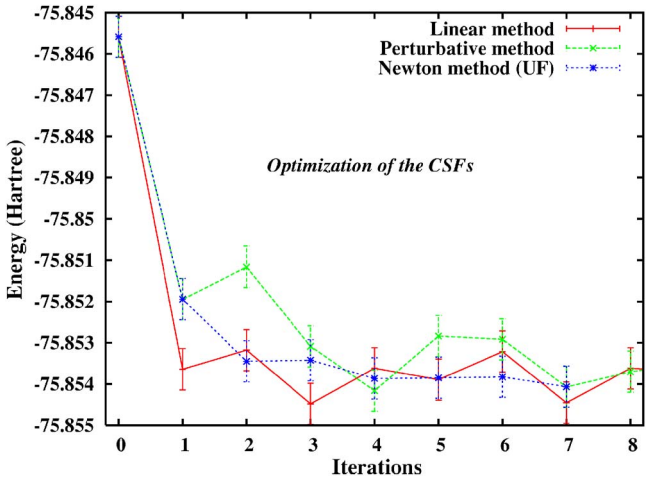

FIG. 2. Convergence of the VMC total energy E_{VMC} of the all-electron C_2 molecule during the optimization of the 49 CSF parameters in a wave function composed of a CAS(8,7) part multiplied by a previously optimized Jastrow factor. The linear, perturbative, and Newton [with the UF Hessian of Eq. (50)] energy minimization methods are compared. The statistical error on the energy at each iteration is 0.5 mhartree.

stabilizing constant a_{diag} in the course of the optimization and despite the fact that we are performing an all-electron rather than a pseudopotential calculation here.[55] The Newton method with the TU Hessian displays an even faster convergence, the energy being essentially converged within the statistical error at iteration 3 or 4. The linear method has a similar convergence rate to the Newton method with the TU Hessian. The Newton method with the TU Hessian and the linear method are both stable even without stabilization if sufficiently large Monte Carlo samples are used. When stabilization is employed, the stabilization constant a_{diag} remains small during the optimization, in this example from 10^{-3} for the initial iteration to 10^{-7} for the last iterations which is two or three orders of magnitude smaller than the values of a_{diag} in the Newton method with the UF Hessian. The perturbative method, in contrast, converges very slowly. In fact, it turns out that the energy denominators for the Jastrow parameters, $\Delta\mathcal{E}_{\alpha_i}$, calculated according to Eq. (44), are all of order unity and a_{diag} needs to be increased to as much as 10^2 to retain stability. In this case, the perturbative method essentially reduces to the inefficient SR optimization technique.

B. Optimization of the CSF coefficients

Figure 2 shows the convergence of the total VMC energy during the optimization of the 49 CSF parameters in a wave function composed of a CAS(8,7) determinantal part multiplied by a previously optimized Jastrow factor, using the linear, perturbative, and Newton [with the UF Hessian of Eq. (50)] methods. The linear method converges in one iteration, as it must, and does not require any stabilization. When stabilization is used, a_{diag} remains as low as 10^{-6}–10^{-8} during the whole optimization. The Newton and perturbative methods converge in two or three iterations and are not as intrinsically stable, a_{diag} being a few orders of magnitude larger for the Newton method and several orders of magnitude larger for the perturbative method. The energy denominators for the CSF parameters in the perturbative method, $\Delta\mathcal{E}_{c_I}$, calculated according to Eq. (44), span only one order of magnitude.

C. Optimization of the orbitals

Figure 3 shows the convergence of the total VMC energy during the optimization of all the 44 orbital parameters in a wave function composed of a single Slater determinant multiplied by a previously optimized Jastrow factor, using the linear, perturbative, and Newton [with the UF Hessian of Eq. (50)] methods. The three methods display very similar convergence rates, the energy being converged within the statistical error in one iteration using any of the three meth-

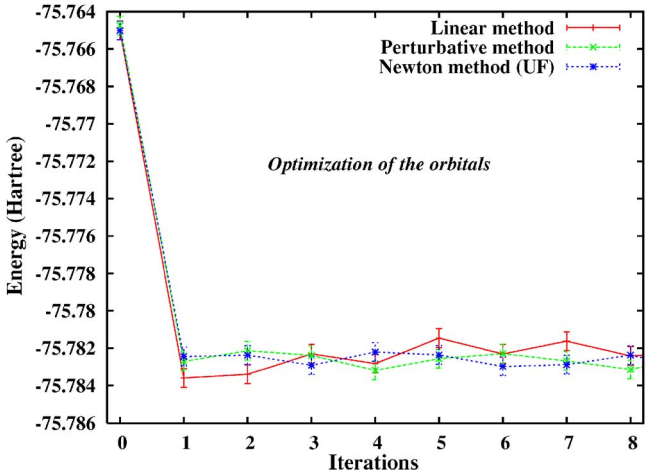

FIG. 3. Convergence of the VMC total energy E_{VMC} of the all-electron C_2 molecule during the optimization of the 44 orbital parameters in a wave function composed of a single Slater determinant multiplied by a previously optimized Jastrow factor. The linear, perturbative, and Newton [with the UF Hessian of Eq. (50)] energy minimization methods are compared. The statistical error on the energy at each iteration is 0.5 mhartree.

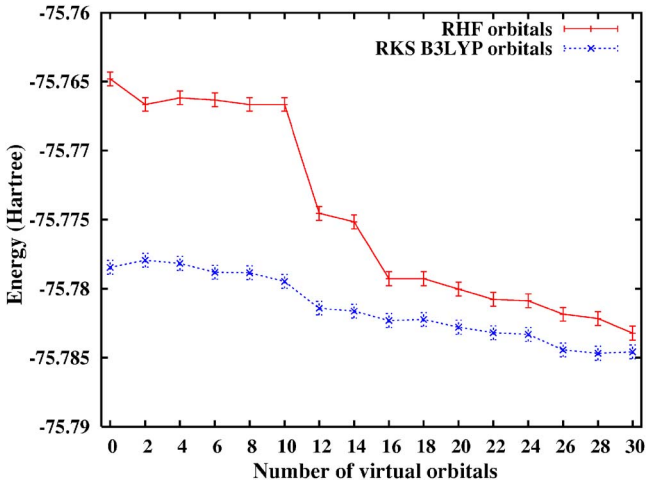

FIG. 4. Total VMC energy E_{VMC} of the all-electron C_2 molecule with respect to the number of virtual orbitals included in the optimization of the orbital parameters in a wave function composed of a single Slater determinant multiplied by a previously optimized Jastrow factor, using RHF and RKS B3LYP starting orbitals. The orbitals are ordered according to their energies. The statistical error on the energy is 0.5 mhartree.

ods. In this example, the linear and perturbative methods converged even without stabilization, whereas the Newton method required stabilization. The energy denominators for the orbital parameters in the perturbative method, $\Delta\mathcal{E}_{kl}$, calculated according to Eq. (44), typically span two orders of magnitude from 1 to 100.

In the previous orbital optimization, we have considered a full optimization of all the orbital parameters, i.e., all the allowed excitations from the 6 closed occupied orbitals to the 30 virtual orbitals were included in the calculation. One may also consider a partial orbital optimization by restricting the excitations to the lowest several virtual orbitals, as also proposed within the EFP or perturbative EFP approaches.[56] This allows one to reduce the computational effort and also to decrease the statistical noise in the calculation since it is the excitations to the highest-lying virtual orbitals that modify the most nodal structure of the wave function, leading to large fluctuations of the ratio $\Psi_i(\mathbf{R})/\Psi_0(\mathbf{R})$. Figure 4 shows the total VMC energy with respect to the number of virtual orbitals included in the optimization for a wave function composed of a single Slater determinant multiplied by a previously optimized Jastrow factor. Two sets of starting orbitals are compared: orbitals obtained from a RHF calculation and orbitals obtained from a restricted Kohn-Sham (RKS) calculation with the hybrid exchange-correlation functional B3LYP,[57,58] using the ordering given by the orbital energies. In both cases, as expected, the energy decreases monotonically within the statistical error as the number of virtual orbitals included in the optimization increases. However, the slope of the energy does not change monotonically and it is necessary to include almost all the orbitals to get close to the optimal energy. From Fig. 4 we see that for the C_2 molecule the B3LYP orbitals provide a better starting point than the RHF orbitals. In our experience, this is often but not always the case. It is possible that the selection of the virtual orbitals adopted here, based on the orbital energy ordering, may not be the best choice and other selections based on symmetry or chemical intuition could lead to a more rapid convergence.

Note that Fig. 4 was obtained by just optimizing the orbital parameters for a fixed, previously optimized Jastrow factor. If instead the Jastrow and orbital parameters are optimized simultaneously a significantly lower energy is obtained, e.g., including all 30 virtual orbitals gives an energy of −75.8069(5) hartree (see Table I) as opposed to −75.7845(5) hartree in Fig. 4.

To summarize, the Newton and the linear methods converge very rapidly when optimizing any kind of parameter, though the linear method is more stable for the optimization of the determinantal part of the wave function. The perturba-

TABLE I. Total VMC and DMC energies, E_{VMC} and E_{DMC}, and VMC standard deviation of the local energy σ_{VMC} of the C_2 molecule for different trial wave functions and different levels of optimization. The kind and number of optimized parameters are indicated. When not optimized in VMC, the CSF and orbital coefficients have been fixed at their RHF values for the single-determinant case and at their CAS MCSCF values for the multiconfiguration cases. For the large Jastrow×RAS-SDTQ(8,26) wave function, the VMC and DMC values are obtained by an extrapolation procedure (see Sec. VI E and Fig. 8). For the energies, the numbers in parentheses are estimates of the statistical error on the last digit. All units are hartree.

Wave function form	Parameters optimized in VMC	E_{VMC}	E_{DMC}	σ_{VMC}
Jastrow×determinant	Jastrow (24)	−75.7648(5)	−75.8570(5)	1.4
	Jastrow (24)+orbitals (44)	−75.8069(5)	−75.8682(5)	1.1
Jastrow×CAS(8,5)	Jastrow (24)	−75.8045(3)	−75.8750(5)	1.3
	Jastrow (24)+CSFs (6)	−75.8094(5)	−75.8807(5)	1.3
	Jastrow (24)+CSFs (6)+orbitals (52)	−75.8374(5)	−75.8882(5)	1.0
Jastrow×CAS(8,7)	Jastrow (24)	−75.8469(5)	−75.8973(5)	1.2
	Jastrow (24)+CSFs (49)	−75.8546(5)	−75.9032(5)	1.2
	Jastrow (24)+CSFs (49)+orbitals (64)	−75.8769(5)	−75.9092(5)	0.9
Jastrow×CAS(8,8)	Jastrow (24)	−75.8462(5)	−75.8999(6)	1.1
	Jastrow (24)+CSFs (164)	−75.8562(5)	−75.9050(6)	1.1
	Jastrow (24)+CSFs (164)+orbitals (70)	−75.8801(6)	−75.9099(5)	0.9
Jastrow×RAS-SDTQ(8,26)	Jastrow+CSFs+orbitals (extrapolation)	−75.9016(5)	−75.9191(5)	···
Exact			−75.9265(8)	

tive method is a good, less expensive alternative for the optimization of the orbital parameters and, to a lesser extent, for the optimization of the CSF parameters, but is very slowly convergent for the Jastrow parameters.

It is clear from Eq. (43) that the perturbative method can be viewed as a Newton method with an approximate Hessian. The poor behavior of the perturbative method for the Jastrow parameters means that this Hessian is a bad approximation to the exact Hessian, whose eigenvalues span more than ten orders of magnitude for these parameters. In fact, any method based on an approximate Hessian that is not able to reproduce all these orders of magnitude, such as the steepest descent method, is bound to converge very slowly. On the other hand, the eigenvalues of the Hessian for the CSF and orbital parameters span only a couple of orders of magnitude and the approximate Hessian of the perturbative method is sufficient to allow rapid convergence.

D. Optimization of all the parameters: Simultaneous or alternated optimization?

After having studied the behavior of the energy minimization methods for the optimization of each kind of parameter, we now move on to the more practical problem of how to optimize all the parameters.

The most obvious possibility is to optimize *simultaneously* the Jastrow, CSF, and orbital parameters using the linear method, the method having the best overall efficiency for all these parameters. In practice, we proceed as follows. We start an optimization run with a short Monte Carlo simulation with a large statistical error (e.g., 0.02 hartree for the C_2 molecule), and we decrease progressively the statistical error at each iteration until the energy is converged to 10^{-3} hartree for three consecutive iterations. We choose the optimal parameters to be those from the iteration with the smallest value of E_{VMC} plus three times the statistical error of E_{VMC}, which is often but not always the last iteration. A typical example of the convergence of the total VMC energy and of the standard deviation σ_{VMC} is shown in Fig. 5 for the simultaneous optimization of the Jastrow, CSF, and orbital parameters in a wave function composed of a CAS(8,7) determinantal part multiplied by a Jastrow factor. In this case, the energy converges in four or five iterations. The standard deviation typically converges a little slower than the energy since we are optimizing just the energy here. A faster convergence, to a somewhat smaller value of the standard deviation, can be achieved by optimizing a linear combination of the energy and variance as in Ref. 23.

Another possibility is to *alternate* between the optimization of the different kinds of parameters until global convergence. This has the advantage of allowing one to use different optimization methods for the various parameters, e.g., optimization of the Jastrow factor and the CSF coefficients with the Newton or linear method and optimization of the orbitals with the less expensive but still very efficient perturbative method. Figure 6 shows the convergence of the total VMC energy and of the standard deviation during the alternated optimization of the Jastrow parameters and of the orbital parameters in a wave function composed of a single Slater determinant multiplied by a Jastrow factor for the all-electron C_2 molecule. The convergence of the energy is surprisingly very slow; the convergence of the standard deviation is even worse. This is an indication of the presence of a strong coupling between some Jastrow and orbital parameters. This situation is in sharp contrast with the case where a pseudopotential is used to remove the core electrons. Figure 7 shows the convergence of the total VMC energy during the alternated optimization of the Jastrow parameters and of the orbital parameters in a wave function composed of a single Slater determinant multiplied by a Jastrow factor for the C_2 molecule with a Hartree-Fock pseudopotential[59] and an adequate Gaussian one-electron basis set. The convergence is very fast, the energy being essentially converged within the statistical error in one macroiteration. This favorable behavior has already been observed in other systems with pseudopotentials,[56] but we have also found pseudopotential systems for which the convergence is not as fast.

For the all-electron case, it thus seems that simultaneous optimization of the parameters is much preferable. The coupling between the different parameters seems to be too

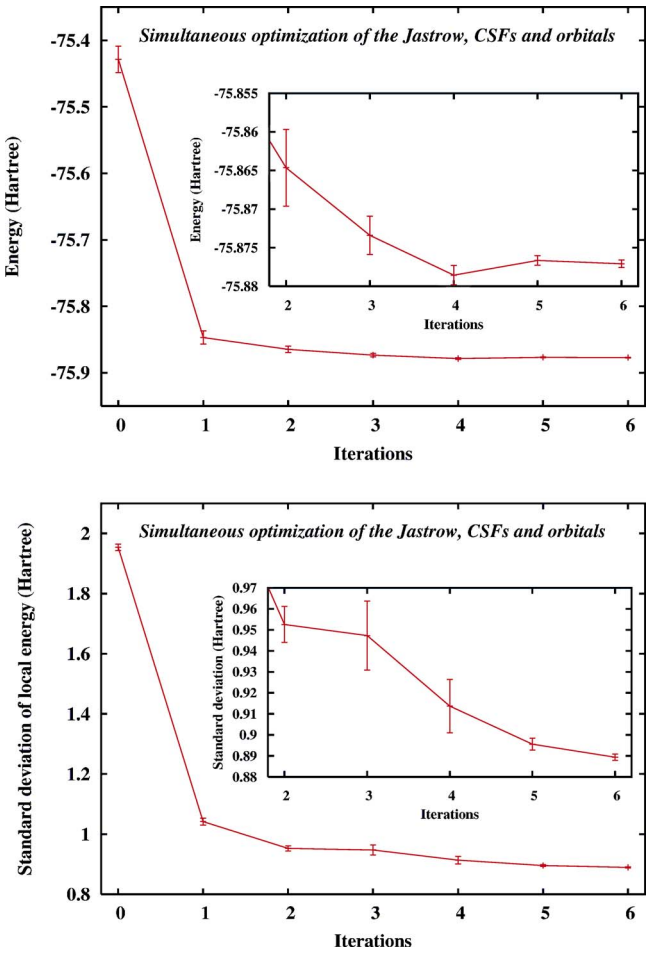

FIG. 5. Convergence of the VMC total energy E_{VMC} (upper plot) and of the VMC standard deviation of the local energy σ_{VMC} (lower plot) of the all-electron C_2 molecule during the *simultaneous* optimization of the 24 Jastrow parameters, 49 CSF parameters, and 64 orbital parameters in a wave function composed of the CAS(8,7) determinantal part multiplied by a Jastrow factor, using the linear energy minimization method. The statistical error on the energy is initially of 0.2 hartree and is decreased by a factor of 2 at each iteration until 0.5 mhartree. The insets are enlargements of the last five iterations.

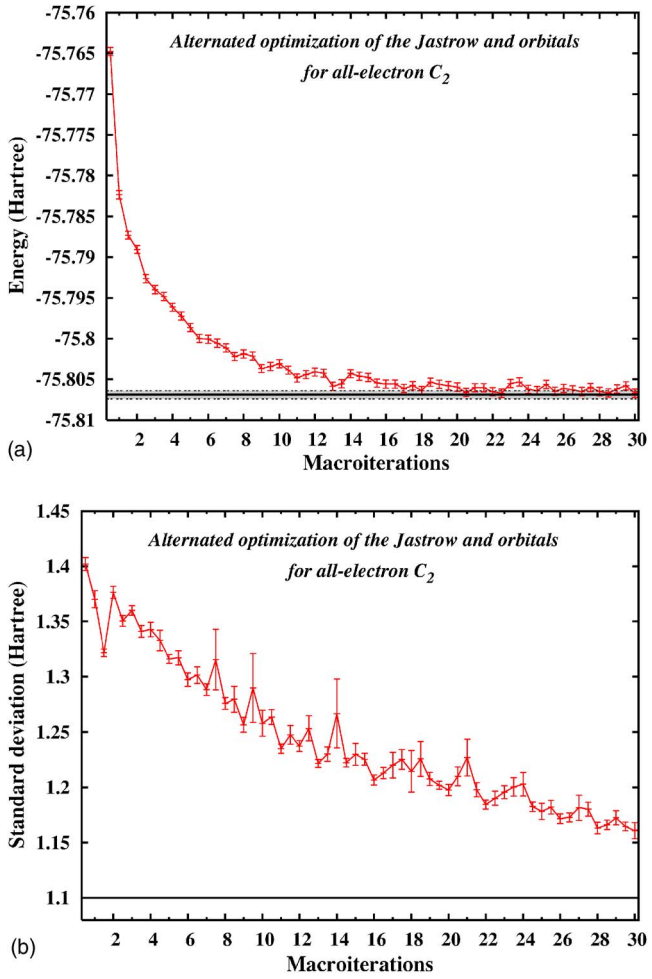

FIG. 6. Demonstration of the slow convergence of the VMC total energy E_{VMC} (upper plot) and of the VMC standard deviation of the local energy σ_{VMC} (lower plot) of the all-electron C_2 molecule during the *alternated* optimization of the 24 Jastrow parameters and 44 orbital parameters in a wave function composed of a single Slater determinant multiplied by a Jastrow factor. The half-integer macroiteration numbers correspond to the optimization of the Jastrow factor and the integer macroiteration numbers correspond to the optimization of the orbitals. The statistical error on the energy is always 0.5 mhartree. The simultaneous optimization of the Jastrow and orbital parameters gives an energy of −75.8069(5) hartree and a standard deviation of 1.1 hartree, indicated on the plots by horizontal lines.

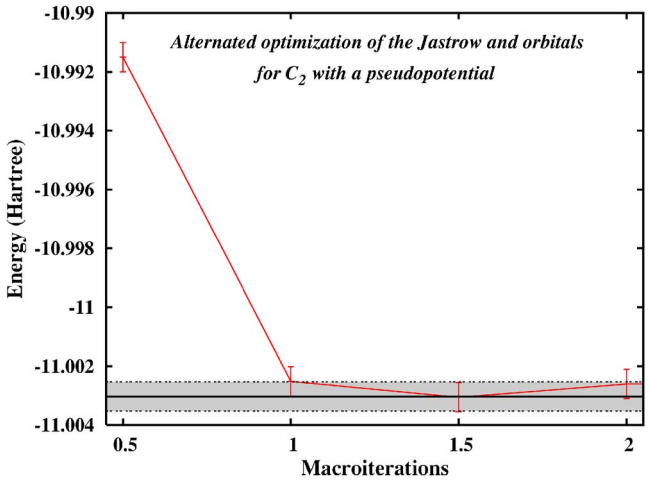

FIG. 7. Convergence of the VMC total energy E_{VMC} of the C_2 molecule with a pseudopotential removing the core electrons during the *alternated* optimization of the 24 Jastrow parameters and 42 orbital parameters in a wave function composed of a single Slater determinant multiplied by a Jastrow factor. The half-integer macroiteration numbers correspond to the optimization of the Jastrow factor and the integer macroiteration numbers correspond to the optimization of the orbitals. The statistical error on the energy is always 0.5 mhartree. The *simultaneous* optimization of the Jastrow and orbital parameters gives an energy of −11.0030(5) hartree, indicated on the plot by horizontal lines.

strong to allow an efficient alternated optimization. For large systems most of the wave function parameters are orbital and CSF parameters for which the perturbative method works well. It seems then promising to simultaneously optimize all the parameters with the Newton or the linear methods, using for the part of the Hessian or the Hamiltonian matrices involving the CSF and orbital coefficient rough approximations inspired by the perturbative method.[60]

E. Systematic improvement by wave function optimization

Table I reports the total VMC and DMC energies, E_{VMC} and E_{DMC}, and the VMC standard deviation of the local energy $\sigma_{\mathrm{VMC}}=\sqrt{\langle E_L(\mathbf{R})^2\rangle-\langle E_L(\mathbf{R})\rangle^2}$ for the different trial wave functions considered. For the single-determinant, CAS(8,5), CAS(8,7), and CAS(8,8), wave functions, we present the results for three levels of optimization. At the first level, only the Jastrow factor is optimized. At the second level, the Jastrow factor and the CSF coefficients are optimized together. At the third level, the Jastrow factor, the CSF coefficients, and the orbitals are all optimized together. Going from one level to the next one improves the accuracy of the wave function but also increases the computational cost of the optimization. We note that it is important to reoptimize the determinantal (CSF and orbital) parameters, along with the Jastrow parameters, rather than keeping them fixed at the values obtained from the MCSCF wave functions. For each wave function, the effect of reoptimizing the determinantal part is to lower the VMC energy by about 0.03–0.04 hartree, and the standard deviation of the energy by about 0.2–0.3 hartree. More remarkably, even though the optimization is performed at the VMC level, the DMC energy also goes down by about 0.01 hartree implying that the nodal hypersurface of the trial wave function also improves. In addition, one observes a systematic improvement of the VMC and DMC energies when the size of the CAS increases, provided that at least the CSF coefficients are reoptimized with the Jastrow factor.

Including all the 110 481 CSFs of the RAS-SDTQ(8,26) wave function is too costly in quantum Monte Carlo, but one can use a series of truncated wave functions obtained by retaining only small numbers of CSFs with coefficients larger in absolute value than a variable cutoff and then estimate the energy by extrapolation to the limit that all the CSFs are kept. Figure 8 shows the VMC and DMC energies obtained with these truncated, fully reoptimized multideterminantal wave functions with respect to the sum of the squares of the MCSCF CSF coefficients retained, $\Sigma_{i=1}^{N_{\mathrm{CSF}}}(c_i^{\mathrm{MCSCF}})^2$. Since the RAS-SDTQ(8,26) wave function is normalized, the latter quantity is equal to 1 in the limit where all the CSFs are kept in the wave function. Experience

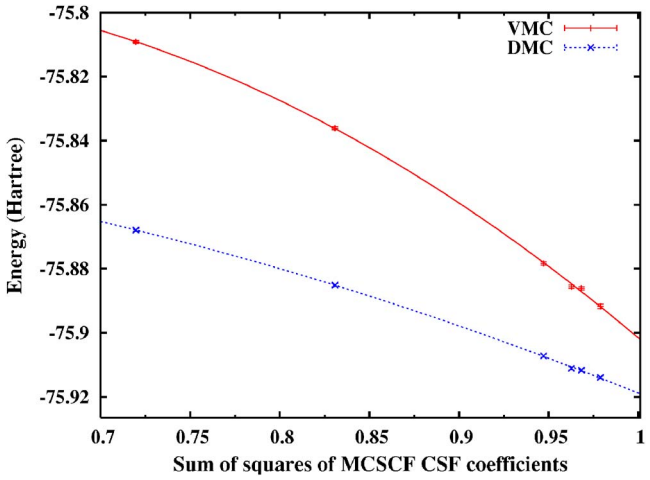

FIG. 8. VMC and DMC energies obtained with the truncated, fully reoptimized Jastrow-Slater RAS-SDTQ(8,26) wave functions with respect to the sum of the squares of the MCSCF CSF coefficients retained, $\sum_{i=1}^{N_{\mathrm{CSF}}}(c_i^{\mathrm{MCSCF}})^2$. The latter quantity is equal to 1 in the limit where all the CSFs of the RAS-SDTQ(8,26) calculation are kept in the wave function which is extrapolated by quadratic fits.

shows that the energies are well extrapolated by quadratic fits. The extrapolated DMC energy is −75.9191(5) which amounts for 98.6% of the correlation energy (using the HF energy of −75.406 20 hartree calculated in Ref. 40).

On the other hand, to calculate accurate well depths (dissociation energy+zero-point energy) it is often sufficient to rely on some partial cancellation of error between the atom and the molecule by employing atomic and molecular wave functions that are consistent with each other. For example, using the DMC energy of the C_2 molecule given by the Jastrow-Slater full-valence CAS(8,8) wave function and the DMC energy of the C atom given by the consistent Jastrow-Slater full-valence CAS(4,4) wave function with the same one-electron basis leads to a well depth of 6.46(1) eV, in perfect agreement within the uncertainty with the exact, nonrelativistic well depth estimated at 6.44(2) eV.[41,61] In contrast, the well depth calculated from MCSCF with the molecular CAS(8,8) and atomic CAS(4,4) wave functions (without Jastrow factor) is 5.62 eV, in poor agreement with the exact value.

VII. CONCLUSIONS

We have studied three wave function optimization methods based on energy minimization in a VMC context: the Newton, linear, and perturbative methods. These general methods have been applied here to the optimization of wave functions consisting of a multiconfiguration expansion multiplied by a Jastrow factor for the all-electron C_2 molecule. The Newton and linear methods are both very efficient for the optimization of the Jastrow, CSF, and orbital parameters, the linear method being generally more stable. The less computationally expensive perturbative method is efficient only for the CSF and orbital parameters. We have used the linear method to simultaneously optimize the Jastrow, CSF, and orbital parameters, a much more efficient procedure than alternating between optimizing the different kinds of parameters. The linear method is capable of yielding not only ground-state energies but excited state energies as well.[11]

Although the optimization is performed at the VMC level, we have observed for the C_2 molecule studied here, as well as for other systems not discussed in the present paper, that as more parameters are optimized the DMC energies decrease monotonically, implying that the nodal hypersurface also improves monotonically. In fact, a sequence of trial wave functions consisting of multiconfiguration expansions of increasing sizes multiplied by a Jastrow factor, with all the Jastrow, CSF, and orbital parameters optimized together, allows one to systematically reduce the fixed-node error of DMC calculations for the systems studied.

Future directions for this work include optimization of the exponents of the one-electron basis functions (either Slater or Gaussian functions), direct optimization of the DMC energy, and optimization of the geometry.

ACKNOWLEDGMENTS

The authors thank Claudia Filippi, Peter Nightingale, Sandro Sorella, Richard Hennig, Roland Assaraf, Andreas Savin, Anthony Scemama,Wissam Al-Saidi, and Paola Gori-Giorgi for stimulating discussions and useful comments on the manuscript, and Eric Shirley for having provided them with the code for generating Hartree-Fock pseudopotentials. This work was supported in part by the National Science Foundation (DMR-0205328 and EAR-0530301), Sandia National Laboratory, a Marie Curie Outgoing International Fellowship (039750-QMC-DFT), and DOE-CMSN. The calculations were performed at the Cornell Nanoscale Facility and the Theory Center.

[1] B. L. Hammond, J. W. A. Lester, and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry* (World Scientific, Singapore, 1994).
[2] *Quantum Monte Carlo Methods in Physics and Chemistry*, NATO Advanced Studies Institute, Series C: Mathematical and Physical Sciences, edited by M. P. Nightingale and C. J. Umrigar (Kluwer, Dordrecht, 1999), Vol. 525.
[3] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[4] C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, Phys. Rev. Lett. **60**, 1719 (1998).
[5] 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, Berlin, 1988).
[6] C. J. Umrigar, Int. J. Quantum Chem. **23**, 217 (1989).
[7] C. Filippi and C. J. Umrigar, J. Chem. Phys. **105**, 213 (1996).
[8] C.-J. Huang, C. J. Umrigar, and M. P. Nightingale, J. Chem. Phys. **107**, 3007 (1997).
[9] M. Snajdr and S. M. Rothstein, J. Chem. Phys. **112**, 4935 (2000).
[10] F. J. Gálvez, E. Buendía, and A. Sarsa, J. Chem. Phys. **115**, 1166 (2001).
[11] M. P. Nightingale and V. Melik-Alaverdian, Phys. Rev. Lett. **87**, 043401 (2001).
[12] S. Fahy, in *Quantum Monte Carlo Methods in Physics and Chemistry*, NATO Advanced Studies Institute, Series C: Mathematical and Physical Sciences, edited by M. P. Nightingale and C. J. Umrigar (Kluwer, Dordrecht, 1999), Vol. 525, p. 101.
[13] C. Filippi and S. Fahy, J. Chem. Phys. **112**, 3523 (2000).
[14] D. Prendergast, D. Bevan, and S. Fahy, Phys. Rev. B **66**, 155104 (2002).
[15] F. Schautz and S. Fahy, J. Chem. Phys. **116**, 3533 (2002).
[16] F. Schautz and C. Filippi, J. Chem. Phys. **120**, 10931 (2004).
[17] A. Scemama and C. Filippi, Phys. Rev. B **73**, 241101 (2006).
[18] S. Sorella, Phys. Rev. B **64**, 024512 (2001).
[19] M. Casula and S. Sorella, J. Chem. Phys. **119**, 6500 (2003).
[20] M. Casula, C. Attaccalite, and S. Sorella, J. Chem. Phys. **121**, 7110

(2004).
[21] X. Lin, H. Zhang, and A. M. Rappe, J. Chem. Phys. **112**, 2650 (2000).
[22] M. W. Lee, M. Mella, and A. M. Rappe, J. Chem. Phys. **112**, 244103 (2005).
[23] C. J. Umrigar and C. Filippi, Phys. Rev. Lett. **94**, 150201 (2005).
[24] S. Sorella, Phys. Rev. B **71**, 241103 (2005).
[25] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. Hennig, e-print cond-mat/0611094.
[26] C.-J. Huang, C. Filippi, and C. J. Umrigar, J. Chem. Phys. **108**, 8838 (1998).
[27] C. J. Umrigar (unpublished).
[28] A. D. Güçlü, G. S. Jeon, C. J. Umrigar, and J. K. Jain, Phys. Rev. B **72**, 205327 (2005).
[29] T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957).
[30] M. W. Schmidt and M. S. Gordon, Annu. Rev. Phys. Chem. **49**, 233 (1998).
[31] H. J. A. Jensen, in *Relativistic and Electron Correlation Effects in Molecules and Solids*, edited by G. L. Malli (Plenum, New York, 1994), p. 179.
[32] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Chichester, 2002).
[33] E. Dalgaard, Chem. Phys. Lett. **65**, 559 (1979).
[34] E. Dalgaard and P. Jørgensen, J. Chem. Phys. **69**, 3833 (1978).
[35] B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, Chem. Phys. **48**, 157 (1980).
[36] F. Grein and T. C. Chang, Chem. Phys. Lett. **12**, 44 (1971).
[37] D. M. Ceperley and B. Bernu, J. Chem. Phys. **89**, 6316 (1988).
[38] E. Artacho and L. M. del Bosch, Phys. Rev. A **43**, 5770 (1991).
[39] D. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B **16**, 3081 (1977).
[40] P. E. Cade and A. C. Wahl, At. Data Nucl. Data Tables **13**, 340 (1974).
[41] L. Bytautas and K. Ruedenberg, J. Chem. Phys. **122**, 154110 (2005).
[42] M. W. Schmidt, K. K. Baldridge, J. A. Boatz *et al.*, J. Comput. Chem. **14**, 1347 (1993).
[43] E. Clementi and C. Roetti, At. Data Nucl. Data Tables **14**, 177 (1974).
[44] T. Koga, H. Tatewaki, and A. J. Thakkar, Phys. Rev. A **47**, 4510 (1993).
[45] W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).
[46] R. F. Stewart, J. Chem. Phys. **52**, 431 (1970).
[47] CHAMP, a quantum Monte Carlo program written by C. J. Umrigar and C. Filippi with contributions by co-workers (URL http://www.tc.cornell.edu/~cyrus/champ.html).
[48] C. J. Umrigar, Phys. Rev. Lett. **71**, 408 (1993).
[49] C. J. Umrigar, in *Quantum Monte Carlo Methods in Physics and Chemistry*, NATO Advanced Studies Institute, Series C: Mathematical and Physical Sciences, edited by M. P. Nightingale and C. J. Umrigar (Kluwer, Dordrecht, 1999), Vol. 525, p. 129.
[50] J. B. Anderson, J. Chem. Phys. **63**, 1499 (1975).
[51] J. B. Anderson, J. Chem. Phys. **65**, 4121 (1976).
[52] P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, J. Chem. Phys. **77**, 5593 (1982).
[53] J. W. Moskowitz, K. E. Schmidt, M. A. Lee, and M. H. Kalos, J. Chem. Phys. **77**, 349 (1982).
[54] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, J. Chem. Phys. **99**, 2865 (1993).
[55] The convergence in two iterations, mentioned near the end of Ref. 23, was obtained using the correlated sampling adjustment of a_{diag} and the TU Hessian.
[56] C. Filippi (private communication).
[57] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[58] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, J. Phys. Chem. **98**, 11623 (1994).
[59] We used the code of Shirley to generate norm-conserving Hartree-Fock pseudopotential according to the construction of D. Vanderbilt, Phys. Rev. B **32**, 8412 (1985).
[60] C. Filippi, J. Toulouse, and C. J. Umrigar (unpublished).
[61] Note that this estimate of the exact well depth differs from the one used in Ref. 25 where we used instead the scalar-relativistic, valence-corrected estimate of Ref. 41, since calculations were performed with a relativistic pseudopotential.