

Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo

Claudia Filippi,^{1, a)} Roland Assaraf,^{2, b)} and S. Moroni^{3, c)}

¹⁾*MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*

²⁾*Laboratoire de Chimie Théorique, CNRS, UMR 7616, Université Pierre et Marie Curie Paris VI, Case 137, 4 Place Jussieu, 75252 Paris Cedex 05, France*

³⁾*CNR-IOM DEMOCRITOS, Istituto Officina dei Materiali, and SISSA Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, I-34136 Trieste, Italy*

We present a simple and general formalism to compute efficiently the derivatives of a multi-determinant Jastrow-Slater wave function, the local energy, the interatomic forces, and similar quantities needed in quantum Monte Carlo. Through a straightforward manipulation of matrices evaluated on the occupied and virtual orbitals, we obtain an efficiency equivalent to algorithmic differentiation in the computation of the interatomic forces and the optimization of the orbital parameters. Furthermore, for a large multi-determinant expansion, the significant computational gain afforded by a recently introduced table method is here extended to the local value of any one-body operator and to its derivatives, in both all-electron and pseudopotential calculations.

I. INTRODUCTION

In the application of quantum Monte Carlo (QMC) methods to electronic systems in real space^{1,2}, one computes expectation values of random variables depending on $\psi(\mathbf{R})$ where ψ is a variational ansatz of the exact wave function, and $\mathbf{R} = (\mathbf{r}_1 \dots \mathbf{r}_N)$ are the coordinates of the N electrons. The total energy is for instance estimated as the expectation value of the local energy $E_L = \hat{H}\psi/\psi$, where \hat{H} is the Hamiltonian of the system. Other examples are the derivatives of ψ and E_L with respect to the atomic coordinates or the variational parameters, which are needed to evaluate the interatomic forces or to optimize the wave function ψ , respectively. It is very important to compute these quantities efficiently because of their large number, typically $O(N)$ or $O(N^2)$, and also because they must be calculated for the many steps of the sampling process needed to collect significant statistics on the quantity of interest.

Here, we propose a general, simple and efficient method to compute these properties for the most common ansatz of ψ found in the literature for electronic structure calculations, namely, a sum of $N_e + 1$ Slater determinants times a Jastrow correlation factor $J(\mathbf{R})$,

$$\psi(\mathbf{R}) = J(\mathbf{R})(D_0 + \sum_{i=1}^{N_e} c_i D_i), \quad (1)$$

where D_0 is a reference determinant (the Hartree-Fock solution, for example) of spin-orbitals (one-body functions depending on the position and the spin) and D_i are excited determinants.³ The formulation we propose here relies on the fact that any one-body operator \hat{O} , as well as any logarithmic derivative, can be written using the same compact expression, namely the trace of the product of two matrices, when acting on a one-determinant Jastrow-Slater wave function. Consequently, derivatives of ψ , local values $O_L = \hat{O}\psi/\psi$ and derivatives of O_L are easy to obtain for the reference $J D_0$ and can be very simply and efficiently computed when D_0 is replaced by an excited determinant D_i . In practice, the method requires only the calculation of molecular orbitals and their derivatives with respect to some parameter λ related to the quantity being computed (e.g. $dE_L/d\lambda$) for all positions \mathbf{r}_i of the electrons. This information is stored in rectangular matrices of size $N \times N_{\text{orb}}$ where N_{orb} is the total number of orbitals (occupied in the reference plus virtual in case of a multi-determinant expansion). Such one-body quantities are very simple to code and are already available in many QMC codes. One has then to apply the few formulas we develop that involve inverses and products of selected square submatrices. Because these formulas are simple and common to all the properties introduced above, this method requires minimal programming effort.

Our theoretical framework is very efficient in the regime of small and large N_e . In case of a single-determinant wave function, so often used in QMC calculations and for small N_e , our formulas can for instance be used to achieve

^{a)}Electronic mail: c.filippi@utwente.nl

^{b)}Electronic mail: assaraf@lct.jussieu.fr

^{c)}Electronic mail: moroni@democritos.it

great computational savings in the evaluation of first-order derivatives such as the $3N_{\text{atoms}}$ internuclear forces using zero-variance estimators⁴ and the space-warp transformation^{5,6}. We recover in fact the same scaling of $O(N^3)$ which was obtained in Ref. 7 with the use of algorithmic differentiation (AD). A similar favorable scaling is also obtained in the computation of the derivatives needed to optimize all orbital parameters in the determinantal component of the wave function. Here, however, we do not need to employ AD since we have at our disposal a very simple and transparent formula for the derivative of the energy. We note that AD tends to be memory intensive and requires some care when applied to complicated programs.

For a wavefunction expanded on a large number N_e of Slater determinants, we provide a very compact formula for O_L and its derivatives evaluated for an excited determinant, D_i , by exploiting that D_i differs from D_0 by a few orbital excitations. In the calculation of ψ , we know from the work by Clark *et al.*⁸ that, once D_0 has been computed, D_i can be updated using matrices of order k where k is the order of the excitation. Here, because in our formulation all the properties introduced above are treated on an equal footing, the favorable asymptotic scaling of $O(k^3 N_e)$, obtained⁸ in the calculation of ψ , $\nabla_i \psi$ and $\Delta_i \psi$ applies to all properties including E_L and its derivatives. Importantly, it holds in all-electron and **pseudopotential** calculations alike. We stress that the formulas we propose here are general and the resulting computational gain comes in addition to other savings which may result from specific features of the wave function ψ (e.g. the possible equality of some determinants of a given spin in the expansion⁹).

The remainder of this paper is organized as follows: In Section II, we introduce the formulas for computing first- and second-order derivatives of a single-determinantal wave function. We also show that, for a Slater determinant, one-body operators can be written as first-order derivatives. The extension to the multi-determinantal case is done in Section III. In Section IV, we present further details on how the expressions are modified in the presence of the Jastrow factor or non-local pseudopotentials, and give numerical demonstration of our formulation in Section V. The formulas for the second-order derivative of an excited determinant are given in Section VI.

II. DERIVATIVES AND ONE-BODY OPERATORS

A. First-order derivatives

We begin with a single (Hartree-Fock) Slater determinant with occupied orbitals, $\phi_1 \dots \phi_N$, and denote it as

$$D(\mathbf{R}) = |\phi_1 \phi_2 \dots \phi_N| = \sum_P (-1)^P \phi_1(\mathbf{r}_{P(1)}) \phi_2(\mathbf{r}_{P(2)}) \dots \phi_N(\mathbf{r}_{P(N)}) = \det(A), \quad (2)$$

with the Slater matrix A ,

$$A_{ij} = \phi_j(\mathbf{r}_i). \quad (3)$$

The orbitals and the electrons correspond respectively to the columns and the rows.

Many important quantities like the drift, the local energy, or the internuclei forces involve the derivatives of a Slater determinant with respect to some parameter λ . The following identity will be the basis of all subsequent developments

$$\frac{d \ln D}{d\lambda} = \text{tr}(A^{-1} B) \quad \text{where} \quad B = \frac{dA}{d\lambda}, \quad (4)$$

where the dependence of A with respect to the derivation parameter λ is implicit. For a proof, one can for example resort to simple chain rule and differentiate with respect to the elements of A

$$\frac{d \ln D}{d\lambda} = \sum_{ij} \frac{d \ln D}{dA_{ij}} \frac{dA_{ij}}{d\lambda} = \text{tr}(A^{-1} \frac{dA}{d\lambda}). \quad (5)$$

The second equality comes from the expansion of the determinant in minors.

Note. $\frac{d \ln D}{dA_{ij}} = \frac{d \ln D}{dD} \cdot \frac{dD}{dA_{ij}} = \frac{1}{D} \cdot \frac{dD}{dA_{ij}}$. We can expand D in the second fraction to $\sum_j A_{ij} C_{ij}$, where C_{ij} is the cofactor matrix. By product rule, we have $\frac{\partial D}{\partial A_{ij}} = \sum_k \frac{\partial A_{ik}}{\partial A_{ij}} C_{ik} + \sum_k A_{ik} \frac{\partial C_{ik}}{\partial A_{ij}}$. Cofactor matrix does not contain any elements on the same row, so $\frac{\partial C_{ik}}{\partial A_{ij}} = 0$. We also have $\frac{\partial A_{ik}}{\partial A_{ij}} = \delta_{kj}$. Therefore, $\frac{\partial D}{\partial A_{ij}} = C_{ij}$ and $\frac{\partial \ln D}{\partial A_{ij}} = \frac{C_{ij}}{D} = (A^{-1})_{ji}$. Finally, $\frac{d \ln D}{d\lambda} = \sum_{ij} \frac{d \ln D}{dA_{ij}} \frac{dA_{ij}}{d\lambda} = \sum_{ij} (A^{-1})_{ji} \frac{dA_{ij}}{d\lambda} = \sum_j ((A^{-1})^T \frac{dA}{d\lambda})_{jj} = \text{tr}(A^{-1} \frac{dA}{d\lambda})$.

If λ is the first coordinate of the first electron x_1 , one obtains the corresponding component of the drift velocity

$$\frac{1}{D} \frac{\partial D}{\partial x_1} = \text{tr}(A^{-1}B) \quad \text{with} \quad B = \frac{\partial A}{\partial x_1}. \quad (6)$$

where the matrix B is zero with the exception of the first row. If one is interested in computing the interatomic forces, one needs to evaluate the derivative of ψ with respect to the atomic coordinates and can employ the same formula (5) with $B = \partial A / \partial R_a$ where R_a is an atomic coordinate. Derivatives of a Slater determinant with respect to variational parameters, useful for optimization purposes, also fit in the same scheme.

Importantly, the application of a one-body operator to a Slater determinant can also be written as a first-order derivative for an appropriate choice of the matrix B . This property will be the basis of an important simplification to compute E_L and its derivatives, but also at the core of highly improved efficiency for large-multideterminant wavefunctions. To show this, we consider the one-body operator

$$\hat{O} = O(\mathbf{r}_1) + \dots O(\mathbf{r}_N), \quad (7)$$

where $O(\mathbf{r}_i)$ is an operator which acts only on a function of \mathbf{r}_i . Applying the operator to the determinant as

$$\hat{O}D = \sum_P (-1)^P (O(\mathbf{r}_1) + \dots O(\mathbf{r}_N)) \phi_1(\mathbf{r}_{P(1)}) \phi_2(\mathbf{r}_{P(2)}) \dots \phi_N(\mathbf{r}_{P(N)}), \quad (8)$$

and expanding the product inside the sum, we have

$$\hat{O}D = |(O\phi_1)\phi_2 \dots \phi_N| + |\phi_1(O\phi_2) \dots \phi_N| + \dots |\phi_1\phi_2 \dots (O\phi_N)|, \quad (9)$$

which is the sum of all mono-excitations obtained by replacing in turn each orbital ϕ_i with $O\phi_i$. It is easy to check that

$$\frac{\hat{O}D}{D} = \frac{d}{d\lambda} \ln \det(A + \lambda B) = \text{tr}(A^{-1}B), \quad (10)$$

where the derivative is taken at $\lambda = 0$ and

$$B_{ij} = (O\phi_j)(\mathbf{r}_i). \quad (11)$$

To prove the first identity in Eq. (10), we just have to perform the derivative of

$$\det(A + \lambda B) = |(\phi_1 + \lambda O\phi_1)(\phi_2 + \lambda O\phi_2) \dots| \quad (12)$$

with respect to λ and use the multi-linearity of the determinant.

Note. The expansion of $\det(A + \lambda B)$ is $|\phi_1\phi_2 \dots| + \lambda \cdot \hat{O}D + O(\lambda^2)$. Take derivative with respect to λ and evaluates at $\lambda = 0$ gives $\hat{O}D$.

The second identity follows from Eq. (4). A very important example is the kinetic energy operator $\hat{O} = \hat{T}$ with

$$O(\mathbf{r}_i) = -\frac{1}{2}\Delta_i \equiv -\frac{1}{2} \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \right). \quad (13)$$

Eq. (10) then holds with the definition

$$B_{ij} = -\frac{1}{2}\Delta A_{ij} = -\frac{1}{2}\Delta\phi_j(\mathbf{r}_i). \quad (14)$$

In other words the Laplacian, which is a sum of second-order derivatives with respect to the electron coordinates, can be written as a first-order derivative with respect to an auxiliary parameter λ when applied to a Slater determinant. Expression (10) can be generalized to wave functions including a Jastrow factor and to other operators such as a non-local pseudopotential. The corresponding B matrices are easy to write and will be given in Section IV.

B. Second-order derivatives

The compact trace expression of a local value O_L (Eq. 10) offers the advantage that its derivative with respect to a parameter μ can be straightforwardly written as

$$\frac{\partial}{\partial \mu} \frac{\hat{O}D}{D} = \text{tr}(A^{-1}\partial_\mu B - X\partial_\mu A), \quad (15)$$

where $\partial_\mu A$ and $\partial_\mu B$ are the matrices of the derivatives of the elements of A and B , respectively, and the matrix X is defined as

$$X = A^{-1}BA^{-1}. \quad (16)$$

This can easily be shown by using $d(A^{-1}) = -A^{-1}dA A^{-1}$ and the cyclic property of the trace.

Note. $\text{tr}(\partial_\mu(A^{-1})B) = -\text{tr}(A^{-1}\partial_\mu A A^{-1}B) = -\text{tr}(X\partial_\mu A)$

Therefore, to compute the derivatives of O_L with respect to many parameters, one evaluates and stores the matrix X at a cost proportional to N^3 , and then completes the derivative for each parameter, Eq. (15), at a cost of at most N^2 . For instance, this procedure allows the efficient calculation of the derivatives of E_L with respect to the nuclear coordinates¹⁰ with a cost per Monte Carlo step proportional to the one of the energy: since the number of atoms N_{atoms} scales linearly with the number of electrons, the cost of computing the forces at each Monte Carlo step is $N^3 + N_{\text{atoms}} \times N^2 \sim N^3$.

The same expression (Eq. 15) can also be used in an orbital optimization run to efficiently compute the derivatives of E_L with respect to the orbital variations. If we consider $\phi_i \rightarrow \phi_i + \mu \phi_j$ where ϕ_i is occupied in the original determinant and ϕ_j unoccupied¹¹, we have

$$\frac{\partial E_L}{\partial \mu} = \frac{\partial}{\partial \mu} \frac{\hat{H}D}{D} = \tilde{M}_{ij}, \quad (17)$$

where the derivative is taken at $\mu = 0$ and the rectangular matrix

$$\tilde{M} \equiv A^{-1}\tilde{B} - X\tilde{A} = A^{-1}\tilde{B} - A^{-1}BA^{-1}\tilde{A}, \quad (18)$$

is computed from the \tilde{A} and \tilde{B} rectangular extensions of A and B to the N_{virt} virtual orbitals (defined here as the orbitals which are unoccupied in the reference determinant).

Note. *Extend the excited states to the right of the ground state functions: $\tilde{A} = |\phi_1\phi_2 \cdots \phi_{e_1}\phi_{e_2} \cdots|$, where ϕ_{e_i} are the excited state functions.*

To show this, we simply note that all elements of the matrices $\partial_\mu A$ and $\partial_\mu B$ are zero with the exception of the i^{th} row which contains ϕ_j and $\hat{H}\psi_j$, respectively. Then, if the number of orbital variations is N_s (which equals the number of determinants \tilde{D} created via single excitations $D \rightarrow D + \mu\tilde{D}$), the cost of evaluating the matrix \tilde{M} is only $2N_s \times N$. In a standard implementation, one would instead compute the full inverse matrix of the corresponding mono-excitation to obtain for example the derivatives with respect to the electron positions in the kinetic energy with a cost proportional to $N_s \times N^2$. Therefore, also in the optimization of the determinantal component as in the case of the interatomic forces, our formulation leads to total cost of estimating the needed quantities proportional to the cost of computing the energy, N^3 , since N_s grows at most like $N \times N_{\text{virt}} \sim N^2$.

Thanks to Eq. (4), the same formula (15) also applies to the second derivative $\partial_\mu \partial_\lambda \ln D$, where the matrix B is then equal to $\partial_\lambda A$. The expression can then be cast in a form where λ and μ enter symmetrically,

$$\frac{\partial^2 \ln D}{\partial \mu \partial \lambda} = \text{tr}(A^{-1}\partial_\mu \partial_\lambda A - (A^{-1}\partial_\lambda A)(A^{-1}\partial_\mu A)), \quad (19)$$

Note. *Here again uses $d(A^{-1}) = -A^{-1}dA A^{-1}$ and the cyclic property of the trace.*

Also in this case, however, if $\{\lambda\}$ and $\{\mu\}$ denote two sets of variables and one set, for instance $\{\lambda\}$, is significantly smaller than the other, it is computationally convenient to group the matrices differently and precompute the matrices $X_\lambda = A^{-1}\partial_\lambda A A^{-1}$, followed by the evaluation of the trace with the more numerous $\partial_\mu A$.

III. MULTIPLE EXCITATIONS

We now consider multiple excitations of the original Slater determinant and deduce all subsequent formulas from Eq. (10) where the derivative is taken at $\lambda = 0$.

A. Determinant of a multiple excitation

If k columns of A are modified (in an excitation of order k), the new Slater determinant is

$$\bar{D} = \det(\bar{A}). \quad (20)$$

We introduce the $N \times k$ matrix P which selects (when applied on the right) the k columns which have been modified. In other words, $\bar{A}P$ is the $N \times k$ matrix made of the k columns of \bar{A} which differ from those of A . For example, if only the first and third column of \bar{A} and A are different,

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix}. \quad (21)$$

More generally, if $(i_1, i_2 \dots i_k)$ are the (ordered) indices of the columns which have been modified, $P_{i_1 1} = P_{i_2 2} \dots = P_{i_k k} = 1$ and the other coefficients of P zero. It is easy to check that PP^T is an $N \times N$ diagonal matrix with elements $(PP^T)_{ii} = 1$ if i is the index of a column which has been modified, and equal to zero otherwise. Consequently, the identity $\bar{A} - A = (\bar{A} - A)PP^T$ holds and, applying the determinant lemma, we have

$$\begin{aligned} \det(\bar{A}) &= \det(A + (\bar{A} - A)PP^T) \\ &= \det(A)\det(1 + P^T A^{-1}(\bar{A} - A)P). \end{aligned} \quad (22)$$

where the last determinant in the second equation is computed for a $k \times k$ matrix. Using that $P^T P$ is the $k \times k$ identity matrix, we obtain

$$\det(\bar{A}) = \det(A)\det(P^T A^{-1} \bar{A} P). \quad (23)$$

This formula was also obtained in Ref. 8.

Note that since $\bar{A}P$ are also columns of \tilde{A} , we can define the $(N + N_{\text{virt}}) \times k$ matrix Q such that $\bar{A}P = \tilde{A}Q$.

Note. A is the original matrix, \bar{A} is the modified matrix, and \tilde{A} is the extension of the original matrix with the addition of excited state functions to the right of A .

For example, if 4 and 6 are the indices of the excited orbitals (e.g. $1 \rightarrow 4$ and $3 \rightarrow 6$ are the list of excitations), the matrix Q is

$$Q = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \end{pmatrix}. \quad (24)$$

With this notation, equation (23) can also be written as

$$\det(\bar{A}) = \det(A)\det(P^T A^{-1} \tilde{A} Q). \quad (25)$$

In practice, once we have computed A^{-1} and $\det(A)$, we can evaluate and store the matrix $A^{-1} \tilde{A}$ where \tilde{A} is the rectangular extension of A to the unoccupied orbitals. For a k^{th} -order excitation, we only have to compute the

determinant of $P^T A^{-1} \tilde{A} Q$ which is a simple submatrix of $A^{-1} \tilde{A}$ with dimension $k \times k$. This submatrix is built by selecting the coefficients $(A^{-1} \tilde{A})_{ij}$ such that i is the index of a substituted orbital and j is the index of an excited orbital. For example, if $1 \rightarrow 11$ and $3 \rightarrow 15$ are the list of excitations, the matrix is

$$P^T A^{-1} \tilde{A} Q = \begin{bmatrix} (A^{-1} \tilde{A})_{1,11} & (A^{-1} \tilde{A})_{1,15} \\ (A^{-1} \tilde{A})_{3,11} & (A^{-1} \tilde{A})_{3,15} \end{bmatrix}. \quad (26)$$

Note that the first block composed of the N first columns of $A^{-1} \tilde{A}$ corresponds to occupied orbitals and is the identity matrix. It is never used and does not have to be stored. In practice, one needs to compute only the $N \times N_{\text{virt}}$ submatrix where N_{virt} is the number of virtual orbitals present in the multi-determinant expansion.

B. One-body operator applied to a k^{th} -excited determinant

Applying formula (10) to a new determinant \bar{D} , we have

$$\frac{\hat{O} \bar{D}}{\bar{D}} = \text{tr}(\bar{A}^{-1} \bar{B}) = \frac{d}{d\lambda} \ln \det(\bar{A} + \lambda \bar{B}), \quad (27)$$

where \bar{A} is built with different orbitals and \bar{B} is the corresponding new matrix B . Using equation (25),

$$\det(\bar{A} + \lambda \bar{B}) = \det(A + \lambda B) \det(P^T (A + \lambda B)^{-1} (\tilde{A} + \lambda \tilde{B}) Q),$$

and computing the logarithmic derivative at $\lambda = 0$, we have

$$\text{tr}(\bar{A}^{-1} \bar{B}) = \text{tr}(A^{-1} B) + \text{tr}((P^T A^{-1} \tilde{A} Q)^{-1} P^T \tilde{M} Q), \quad (28)$$

where we used again Eq. (5) and introduced the same matrix as in Eq. (18)

$$\tilde{M} = A^{-1} \tilde{B} - A^{-1} B A^{-1} \tilde{A}.$$

Note. From the equations above, we have $\text{tr}(\bar{A}^{-1} \bar{B}) = \frac{d}{d\lambda} \ln \det(A + \lambda B) + \frac{d}{d\lambda} \ln \det(P^T (A + \lambda B)^{-1} (\tilde{A} + \lambda \tilde{B}) Q)$, where the derivatives are evaluated at $\lambda = 0$. The second one $\frac{d}{d\lambda} \ln \det(P^T (A + \lambda B)^{-1} (\tilde{A} + \lambda \tilde{B}) Q) = \text{tr}((P^T A^{-1} \tilde{A} Q)^{-1} \partial_\lambda (P^T (A + \lambda B)^{-1} (\tilde{A} + \lambda \tilde{B}) Q))$.

Note that the expression (28) applies to any logarithmic derivative $\partial_\lambda (\ln \bar{D})$ with $B = \partial_\lambda A$ and $\tilde{B} = \partial_\lambda \tilde{A}$.

In practice, we need to compute the rectangular matrix \tilde{M} and, for a k^{th} -order excitation, construct the $k \times k$ submatrix $P^T \tilde{M} Q$, which is built in the same way as $P^T A^{-1} \tilde{A} Q$ is built from $A^{-1} \tilde{A}$. Then, one has to perform the trace of the inverse of the $k \times k$ matrix $P^T A^{-1} \tilde{A} Q$ times the matrix $P^T \tilde{M} Q$. The cost of this calculation is of order k^3 due to the computation of the inverse. For example, if $1 \rightarrow 11$ and $3 \rightarrow 15$ are the list of excitations, Eq. (28) becomes

$$\frac{\hat{O} \bar{D}}{\bar{D}} = \frac{\hat{O} D}{D} + \text{tr} \left(\begin{bmatrix} (A^{-1} \tilde{A})_{1,11} & (A^{-1} \tilde{A})_{1,15} \\ (A^{-1} \tilde{A})_{3,11} & (A^{-1} \tilde{A})_{3,15} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{M}_{1,11} & \tilde{M}_{1,15} \\ \tilde{M}_{3,11} & \tilde{M}_{3,15} \end{bmatrix} \right). \quad (29)$$

For a mono-excitation $i \rightarrow j$, P^T and Q select a one dimensional space. $(P^T A^{-1} \tilde{A} Q)^{-1}$ is therefore a scalar, which is equal to D/\bar{D} thanks to Eq. (25), $P^T \tilde{M} Q$ is the matrix element \tilde{M}_{ij} , and $\bar{D} = \partial_\mu D$ where μ is a mono-excitation parameter introduced in equation (17). Using Eq. (28), we recover that for a mono-excitation

$$\tilde{M}_{ij} = \left(\frac{\hat{O} \bar{D}}{\bar{D}} - \frac{\hat{O} D}{D} \right) \frac{\bar{D}}{D} = \frac{d}{d\mu} \frac{\hat{O} D}{D}. \quad (30)$$

We note that the first N columns of \tilde{M} are identically zero and, as in the case of the matrix $A^{-1} \tilde{A}$, do not have to be stored. The matrix elements of \tilde{M} should only be computed for the rows and columns corresponding to the N_s active single excitations, which are in general fewer than the product $N_{\text{act}} N_{\text{virt}}$ of the occupied active orbitals in the reference determinant and the virtual orbitals. The cost is $O(N^2 N_{\text{act}}) + O(N_s N)$ if this product is evaluated from the left to the right, while it is $O(N^2 N_{\text{virt}}) + O(N_s N)$ if one starts from the right. If the matrices B and the corresponding \tilde{B} are sparse, the cost is smaller. In particular, for the drift velocity, these matrices have only one non-zero row, and the cost of evaluating the \tilde{M} needed to compute the derivative with respect to the coordinates of one electron is $O(N_s)$.

In Appendix B, we also provide a more lengthy derivation of expression (28) using the Sherman-Morrison-Woodbury formula instead of performing the derivative of (25).

IV. JASTROW FACTOR, PSEUDOPOTENTIALS, AND OTHER EXPRESSIONS FOR B

When a Jastrow factor is included,

$$\psi(\mathbf{R}) = J(\mathbf{R})D(\mathbf{R}) = J(\mathbf{R})\det(A(\mathbf{R})), \quad (31)$$

and the expression of the matrix B in O_L must be modified to account for the presence of the Jastrow factor.

We begin with the local kinetic energy,

$$\frac{\hat{T}\psi}{\psi} = -\frac{1}{2} \sum_i \frac{\Delta_i \psi}{\psi} = \frac{1}{\det(A)} \sum_i \left[-\frac{1}{2} \left(\Delta_i + 2 \frac{\nabla_i J}{J} \cdot \nabla_i + \frac{\Delta_i J}{J} \right) \right] \det(A). \quad (32)$$

Following the derivations in Section II, we identify a generalization of the operator $O(\mathbf{r}_i)$ as the operator within the square brackets and obtain

$$B_{ij}^{\text{kin}} = -\frac{1}{2} \left[\Delta \phi_j(\mathbf{r}_i) + 2 \frac{\nabla_i J}{J} \cdot \nabla \phi_j(\mathbf{r}_i) + \frac{\Delta_i J}{J} \phi_j(\mathbf{r}_i) \right]. \quad (33)$$

The local kinetic energy can then be written as

$$\frac{\hat{T}\psi}{\psi} = \text{tr}(A^{-1}B^{\text{kin}}). \quad (34)$$

It is possible to cast the contribution of the potential to E_L in a similar form, starting from the more complicated non-local component

$$\frac{\hat{V}_{\text{NL}}\psi}{\psi} = \sum_i \sum_a \sum_l v_l^a(r_{ia}) \int_{|\mathbf{r}'_{ia}|=r_{ia}} d\Omega' P_l(\cos \theta') \frac{\psi(\mathbf{r}_1, \dots, \mathbf{r}'_i, \dots, \mathbf{r}_N)}{\psi(\mathbf{r}_1, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N)}, \quad (35)$$

where the summations over i , a , and l run over the electrons, the nuclei, and the angular components of the non-local pseudopotential, respectively. For each electron coordinate, the integral is over a sphere centered on a nucleus with radius given by the electron-nucleus distance $r_{ia} = |\mathbf{r}_{ia}| = |\mathbf{r}_i - \mathbf{R}_a|$ and the angle θ' is between the vectors \mathbf{r}'_{ia} and \mathbf{r}_{ia} . In QMC, the integral is computed as a sum over quadrature points characterized by weights w_q and unit directions $\hat{\mathbf{u}}_q$,

$$\frac{\hat{V}_{\text{NL}}\psi}{\psi} = \frac{1}{\det(A)} \sum_i \left[\sum_a \sum_l v_l^a(r_{ia}) \sum_q w_q P_l(\cos \theta_q^a) \frac{J(\dots, \mathbf{q}_i^a, \dots)}{J(\dots, \mathbf{r}_i, \dots)} \right] \det(A(\dots, \mathbf{q}_i^a, \dots)) \quad (36)$$

where $\mathbf{q}_i^a = \mathbf{R}_a + r_{ia} \hat{\mathbf{u}}_q$ and θ_q^a is the angle between $\hat{\mathbf{u}}_q$ and \mathbf{r}_{ia} . In general, a different number of angular components and quadrature points can be used for the different atom types. We then identify the matrix B^{NL} as

$$B_{ij}^{\text{NL}} = \sum_a \sum_l v_l^a(r_{ia}) \sum_q w_q P_l(\cos \theta_q^a) \frac{J(\dots, \mathbf{q}_i^a, \dots)}{J(\dots, \mathbf{r}_i, \dots)} \phi_j(\mathbf{q}_i^a), \quad (37)$$

so that

$$\frac{\hat{V}_{\text{NL}}\psi}{\psi} = \text{tr}(A^{-1}B^{\text{NL}}). \quad (38)$$

In analogy to the treatment of the Laplacian of the Jastrow in the local kinetic energy (Eq. 32), we can rewrite the contribution of the local potential as

$$V_{\text{loc}} = \frac{1}{\det(A)} \sum_i \left[\sum_a v_{\text{loc}}^a(r_{ia}) + \sum_{j < i} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right] \det(A) \quad (39)$$

and define

$$B_{ij}^{\text{loc}} = \left[\sum_a v_{\text{loc}}^a(r_{ia}) + \sum_{j < i} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \right] \phi_j(\mathbf{r}_i). \quad (40)$$

The complete matrix B in the trace expression of E_L is the sum of the kinetic (Eq. 32) and potential (Eqs. 37 and 40) contributions. From the rectangular extensions \tilde{A} and \tilde{B} of the A and B matrices to the unoccupied orbitals, one can compute the local energy of any multiple excitation at low cost and, therefore, of any wave function given by a CI expansion times a Jastrow factor,

$$\psi_{\text{CI}} = J \left[\det(A) + \sum_I c_I \det(\bar{A}_I) \right], \quad (41)$$

where A is the reference Slater matrix computed from the occupied orbitals and \bar{A}_I is a k_I^{th} -order excited Slater matrix. Once we have computed B (Eq. 33), A^{-1} , $A^{-1}\tilde{A}$, and \tilde{M} , the local energy reads

$$\frac{\hat{H}\psi_{\text{CI}}}{\psi_{\text{CI}}} = \text{tr}(A^{-1}B) + \frac{\sum_I c_I \text{tr}(\alpha_I^{-1}M_I) \det(\alpha_I)}{\sum_I c_I \det(\alpha_I)} \quad (42)$$

where the matrices α_I and M_I are $k_I \times k_I$ submatrices of $A^{-1}\tilde{A}$ and \tilde{M} , respectively. As discussed in Appendix D, the overall cost of using these formulas to build the sampling process in a typical QMC run (i.e. to compute the local energy and perform a sweep over the N electrons) is $O(N^3) + O(N^2 N_{\text{virt}}) + O(N N_s) + O(N N_e)$ where $N_s \leq N_{\text{act}} N_{\text{virt}}$. If $N_{\text{virt}} \leq O(N)$, this simplifies to $O(N^3) + O(N N_e)$ since $N_s \leq N_e$.

In Appendix A, we give the expressions of the derivatives of the matrices A and B with respect to the atomic coordinates (in particular, the formula for the derivatives of B^{NL}) needed in the computation of the interatomic forces. We also discuss how to efficiently evaluate the additional terms in the force estimator introduced by the use of the space-warp transformation on the electron^{5,6}, which also require the derivatives of A and B with respect to the electronic coordinates. From the extension of these matrices to the unoccupied orbitals, one can easily compute the forces for a general multi-determinant Jastrow-Slater wave function.

V. NUMERICAL EXAMPLES

We demonstrate the formulas above on the C_nH_{n+2} molecular series with $n=4-60$. We employ the CHAMP code¹² with scalar-relativistic energy-consistent Hartree-Fock pseudopotentials and the corresponding cc-pVDZ basis set^{13,14}. The Jastrow factor is limited to a simple two-body electron-electron term and the single determinant is built from Hartree-Fock orbitals.

The low computational cost of the **expression of the derivative** (Eq. 15) in the VMC calculation of the interatomic forces for a one-determinant Jastrow-Slater wave function is demonstrated in Fig. 1: For the largest system considered here which includes 122 atoms, computing all interatomic gradients costs less than 4 times a VMC simulation where one only evaluates the total energy. A similar factor has been reported in Ref. 7 where the forces were however evaluated with the aid of algorithmic differentiation (AD). Here, we demonstrate that a simple algebraic manipulation of the quantities needed to compute the forces leads to transparent, simple formulas to implement and an equivalent computational gain to the use of AD. We note that the ratio of the CPU time of evaluating E_L and the interatomic forces to the time of computing the energy alone should asymptotically be constant: The very weak linear dependence on the number of atoms (electrons) observed here is due to the N^2 term in the computational cost being more important in the energy than in the force calculation, **at least for these system sizes**.

We illustrate the gain achieved in the application of the same **expression of the derivative** to the orbital optimization of a one-determinant Jastrow-Slater wave function (Eq. 17) in Fig. 2. For each system, we consider all possible orbital variations $\phi_i \rightarrow \phi_i + \mu\phi_j$ and compute E_L together with the quantities $\partial_\mu\psi$ and $\hat{H}\partial_\mu\psi$ needed in the linear optimization method¹⁵. The ratio of the cost of such a VMC simulation to the cost of only evaluating E_L should not grow with system size: $\hat{H}\partial_\mu\psi$ can be straightforwardly obtained from $\partial_\mu(\hat{H}\psi/\psi)$ and the cost of calculating $\partial_\mu(\hat{H}\psi/\psi)$ for all possible orbital variations is proportional to N^3 as discussed in Section II B. We find that this ratio remains well below 4 for system sizes leading to as many as 4.5×10^4 orbital variations.

Finally, we demonstrate the speedup in a VMC simulation performed using expression (Eq. 28) to compute O_L for a multi-determinant wave function. We focus on the local energy, which is evaluated after all the electrons have been moved once, and employ the same formula also in the computation of the gradient with respect to the coordinates of the electron being moved during the sweep over all the electrons. For C_4H_6 , C_8H_{10} , and $\text{C}_{16}\text{H}_{18}$, we generate a set of N_e doubly excited determinants in either the up- or the down-spin component, treat all up- and down-spin determinants as distinct, and excite also from the core. Additional computational saving can therefore be achieved by exploiting that different excitations may share the same spin component⁹ or by limiting the number N_{act} of active orbitals. For $\text{C}_{16}\text{H}_{18}$, we also investigate the use of triple excitations in either the up- or the down-spin determinants (for the purpose of testing the efficiency, we construct in this case wavefunctions with only triple excitations).

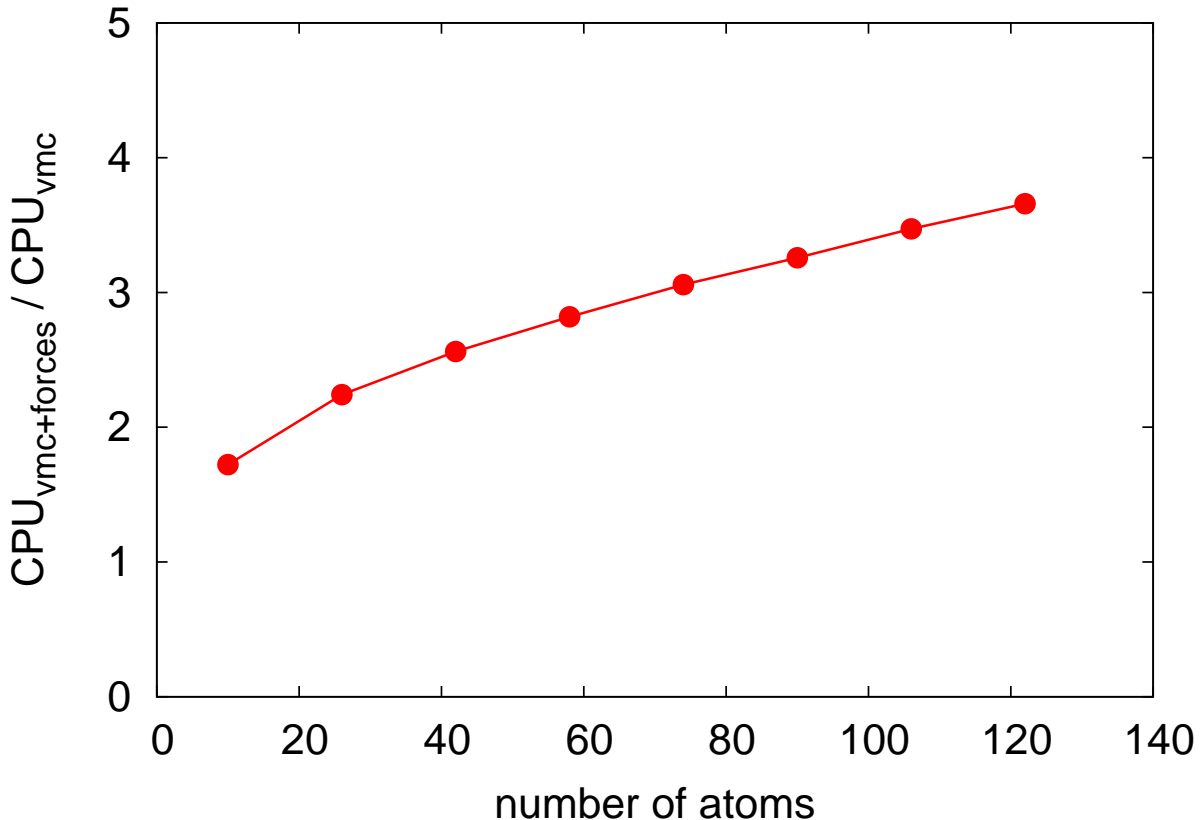


FIG. 1. Ratio of the CPU time for a VMC calculation of the forces to the CPU time for the same simulation of the energy alone. The number of atoms refers to the sequence of molecules C_nH_{n+2} with n between 4 and 60. The forces are calculated after moving all the electrons once.

In Fig. 3, we present three different measures of speedup. In the left panel, we compare with the results presented in Ref. 8 (see their Fig. 3 and green curve) and only estimate the cost of computing the wave function, the drift, and the relevant matrix updates (not the orbitals) with respect to the standard method of computing and updating the inverse matrices of all determinants. In agreement with Ref. 8, we find that the ratio between the two computational costs increases quickly with the number N_{det} of determinants before settling to a value, which we however find to be greatly dependent on the machine and compiler used. A more direct comparison would require further knowledge of the separate performance of the algorithms employed in the standard and improved calculation. The speedup for more than approximately 100 determinants ranges between about 10 and 100 for the systems studied here.

The central panel represents a more realistic assessment of the formulas presented, showing the ratio of the time of a complete VMC computation of the energy with the standard and the new algorithm. The speedup measured in this way is rather comparable to that reported in the left panel, indicating that other parts of the code either scale similarly or do not affect the overall ratio. As expected, the gain is a bit smaller for triple than for double excitations due to the larger dimension of the matrices needed to evaluate the local energy of a higher excitation (Eq. 28). Finally, the right panel allows a comparison with the way the speedup is measured in Ref. 9, where the cost of the improved calculation is compared to N_{det} times the cost of a run with only one determinant, namely, $N_{\text{det}} \times \text{time}(1)/\text{time}(N_{\text{det}})$. The higher values obtained with this measure, however, are not the speedups one gains in reality in comparison with the standard method **for which the computational cost is lower than** $N_{\text{det}} \times \text{time}(1)$.

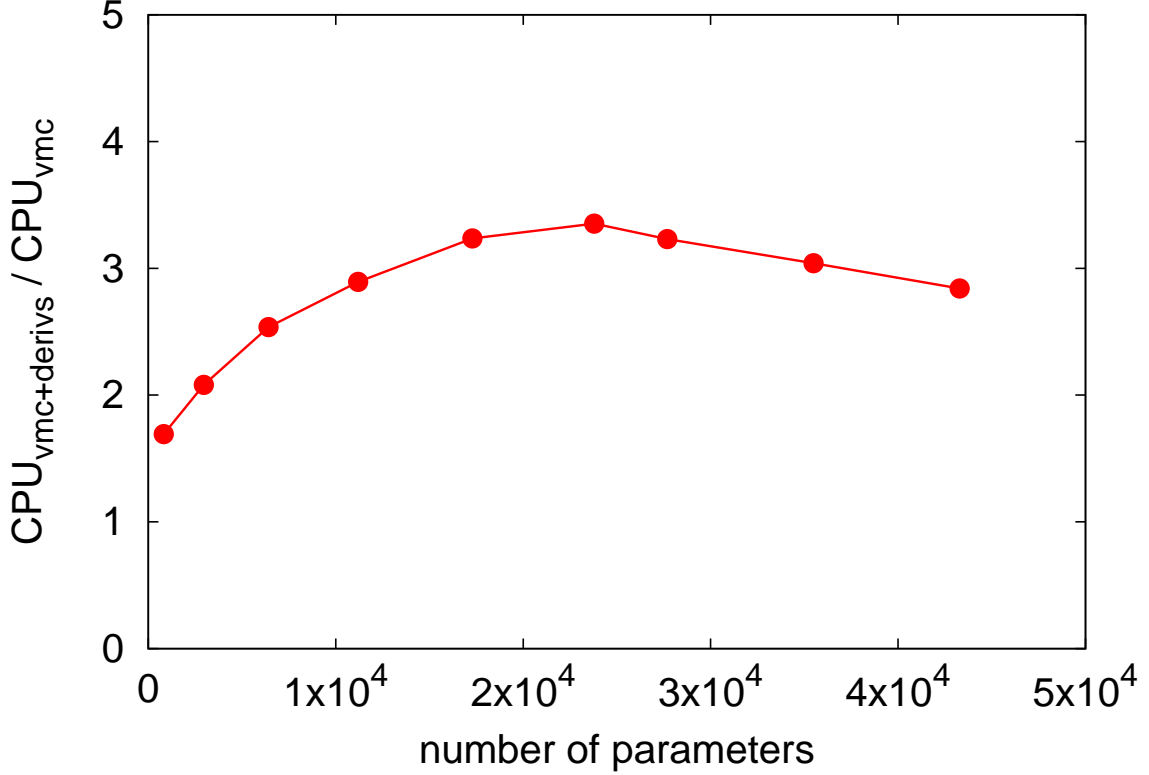


FIG. 2. Ratio of the CPU time of a VMC simulation where all derivatives required for orbital optimization are computed, and of the same simulation with the energy alone. The number of variational parameters refers to the sequence of molecules C_nH_{n+2} with n between 4 and 44. The derivatives are calculated after moving all the electrons once.

VI. SECOND-DERIVATIVE OF A k^{th} -EXCITED DETERMINANT

We consider here the derivative of a local quantity O_L computed for an excited determinant \bar{D} . This is the equivalent of formula (15) applied to \bar{D} , and is obtained by differentiating expression (28) with respect to μ ,

$$\frac{\partial}{\partial \mu} \frac{\hat{O}\bar{D}}{\bar{D}} = \frac{\partial}{\partial \mu} \frac{\hat{O}D}{D} + \text{tr} \left((\partial_\mu \alpha^{-1}) P^T \tilde{M} Q + \alpha^{-1} P^T \partial_\mu \tilde{M} Q \right), \quad (43)$$

where $\alpha = P^T A^{-1} \tilde{A} Q$. With the use of chain rule, it is straightforward to show that

$$\partial_\mu \tilde{M} = [A^{-1} \partial_\mu \tilde{B} - X \partial_\mu \tilde{A}] - [A^{-1} \partial_\mu B - X \partial_\mu A] (A^{-1} \tilde{A}) - (A^{-1} \partial_\mu A) \tilde{M} \quad (44)$$

and

$$\partial_\mu (A^{-1} \tilde{A}) = A^{-1} \partial_\mu \tilde{A} - (A^{-1} \partial_\mu A) (A^{-1} \tilde{A}), \quad (45)$$

so that $\partial_\mu \alpha^{-1} = -\alpha^{-1} P^T \partial_\mu (A^{-1} \tilde{A}) Q \alpha^{-1}$. To evaluate these rectangular matrices, we need to extend the computation of $\partial_\mu A$ and $\partial_\mu B$ (Eq. 15) to the virtual orbitals while other relevant matrices like $A^{-1} \tilde{A}$ and \tilde{M} are already available from the computation of the excited determinant and the corresponding local quantity. The matrices $P^T \partial_\mu \tilde{M} Q$ and $P^T \partial_\mu (A^{-1} \tilde{A}) Q$ are **still simple $k \times k$ submatrices** constructed from the elements whose row and column indices correspond to the substituted occupied and the excited orbitals, respectively. We note that $\partial_\mu \tilde{M}$ is the matrix of second derivatives of the local quantity with respect to μ and the mono-excitation parameter (Eq. 17).

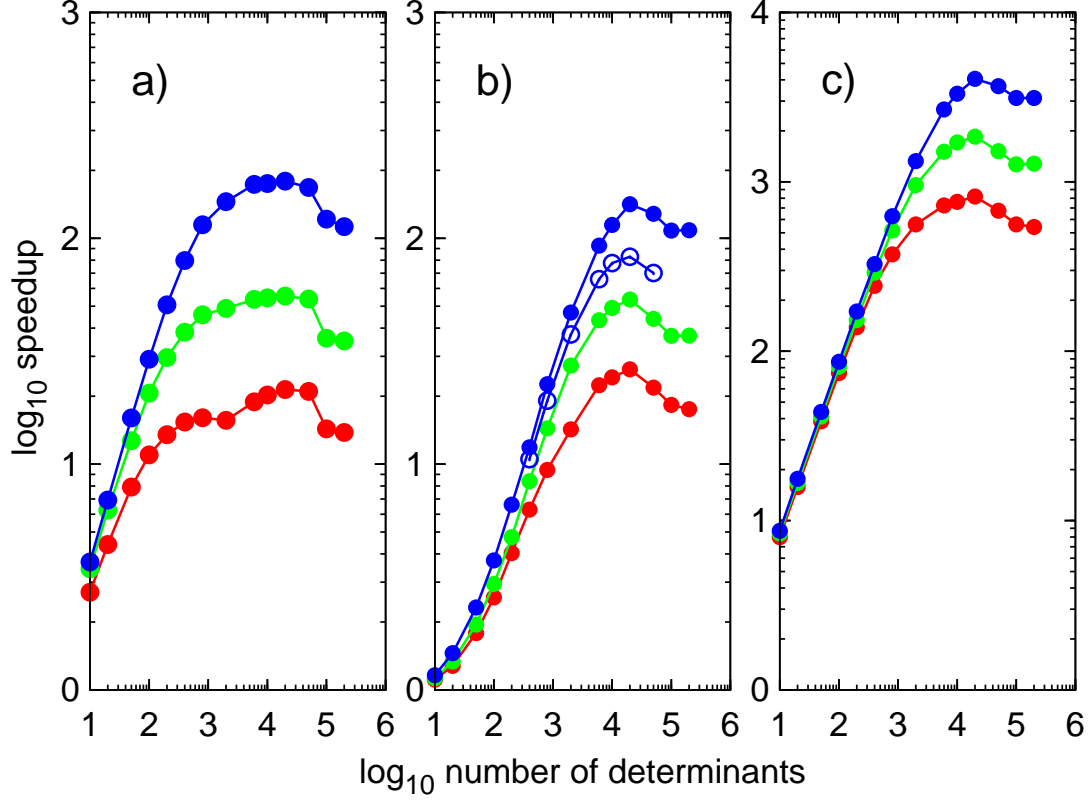


FIG. 3. Speedup of the improved algorithm for a multi-determinantal wave function measured as a) gain over a standard algorithm in the calculation of the wave function, the drift, and the relevant matrix updates, b) gain over a standard algorithm in the complete VMC computation of the energy, and c) gain over N_{det} times the cost of a VMC run with one determinant. The local energy is computed after a sweep over all electrons. Data are shown for C_4H_6 (red), C_8H_{10} (green), and $\text{C}_{16}\text{H}_{18}$ (blue) as a function of the number of determinants. Filled (empty) circles refer to double (triple) excited determinants. All determinants in the expansion are treated as distinct and core excitations are included.

It should be apparent by now that, in practice, one needs to calculate the product of A^{-1} with other matrices as in $A^{-1}B$, $A^{-1}\tilde{B}$, $A^{-1}\partial_\mu A$, $A^{-1}\partial_\mu \tilde{A}$ etc. and that these matrix products constitute the building blocks of the second derivatives and all other quantities derived so far. Additionally, in the computation of the second derivatives, it might be computationally advantageous to evaluate the products $X\tilde{A}$, $X\partial_\mu A$, and $X\partial_\mu \tilde{A}$ as detailed above and in Section II B. The formulas more explicitly written in terms of these products and therefore closer to the actual implementation are given for clarity in Appendix C.

Since the computational cost of building $\partial_\mu \tilde{M}$ is of order $N^2 \times N_{\text{virt}}$, the total cost of evaluating the last term in expression (43) for a multi-determinant wave function typically becomes $O(N^2 N_{\text{virt}}) + O(k^3 N_e)$ where N_e is the number of excited determinants. The same scaling will also characterize higher-order derivatives. If μ represents the coordinate of one atom, the final cost to construct the gradient with respect to the $3N_{\text{atoms}}$ nuclear coordinates is

$$O(N^3) + O(N^2 N_{\text{virt}} N_{\text{atoms}}) + O(N_e N_{\text{atoms}}),$$

If the number of active occupied orbitals $N_{\text{act}} < N_{\text{virt}}$, the computation of the three terms in Eq. 44 can be carried out in the same way as discussed before¹⁶ for the matrix \tilde{M} . For a single-determinant wave function or a small expansion with N_{act} and/or N_{virt} small, we recover the cost described in Section II B.

Finally, we stress that the formula above describes not only the derivative of a local quantity but also the second derivative $\partial_\mu \partial_\lambda \ln \tilde{D}$. The matrices B and \tilde{B} are then equal to $\partial_\lambda A$ and $\partial_\lambda \tilde{A}$, respectively. An expression equivalent to Eq. 43 but where λ and μ are treated on an equal footing is given in Appendix C.

VII. CONCLUDING REMARKS

We have presented general and simple formulas to efficiently compute derivatives of wave functions, local quantities, and their derivatives needed in QMC simulations, when the wave function ψ is written as a Jastrow factor times an expansion of $N_e + 1$ Slater determinants.

The simplicity of the formulas stems from the fact that a derivative of one-determinant wave function and a local quantity such as the local energy are treated on an equal footing and expressed as a trace of matrices which only require the computation of one-body functions (molecular orbitals) and their derivatives. The extension of these formulas to excited determinants is straightforward: One evaluates the matrix elements also for the virtual orbitals (unoccupied in the reference) and computes products involving the resulting rectangular matrices, sub-matrices and their inverses in the spirit of what Clark *et al.*⁸ had developed for the calculation of the multi-determinant wave function. Furthermore, our formulation allows an easy generalization to higher derivatives. Regarding the efficiency, it leads to significant gains for large N_e when one computes many local properties and/or derivatives in a practical simulation.

Appendix A: Interatomic forces

1. Derivatives with respect to nuclei positions

To calculate the derivative of the local energy and wave function with respect to the nuclear coordinates, we need to evaluate the corresponding matrices $\partial_\mu A$ and $\partial_\mu B$ in addition to the logarithmic derivatives of the Jastrow factor. The matrix elements of $\partial_\mu A$ are simply the derivative of the single-particle orbitals, which we expand on an atomic basis $\{\chi\}$ as

$$A_{ij} = \phi_j(\mathbf{r}_i) = \sum_a \sum_{l_a}^{L_a} b_{jl_a} \chi_{l_a}(\mathbf{r}_i - \mathbf{R}_a), \quad (\text{A1})$$

where L_a is the number of basis functions on atom a . Then, we obtain

$$\nabla_a A_{ij} = \nabla_a \phi_j(\mathbf{r}_i) = - \sum_{l_a}^{L_a} b_{kl_a} \nabla \chi_{l_a}(\mathbf{r}_i - \mathbf{R}_a) \quad (\text{A2})$$

Consequently, since the gradients of the basis functions are also needed to calculate $\nabla_i \phi_j(\mathbf{r}_i)$, the computation of $\partial_\mu A$ only requires quantities which are normally evaluated in sampling the local energy. Similarly, the computation of $\partial_\mu B^{\text{kin}}$ (Eq. 33) requires derivatives of the basis functions, most of which have already been evaluated for the local energy, with the exception of the off-diagonal components of the Hessian and the gradient of the Laplacian of χ . We do not report here the relatively simple expression of $\partial_\mu B^{\text{kin}}$ but focus on the somewhat more complicated $\partial_\mu B^{\text{NL}}$ (Eq. 37).

In taking the derivative of B^{NL} with respect to the nuclear coordinates, we need to consider its explicit dependence on the nuclear coordinates \mathbf{R}_a as for instance in $\phi_j(\mathbf{r})$ (Eq. A1), as well as the implicit dependence through the quadrature points. Therefore, we have

$$\begin{aligned} \nabla_a B_{ij}^{\text{NL}} = & - \sum_q w_q \phi_j(\mathbf{q}_i^a) \frac{J(\mathbf{q}_i^a)}{J(\mathbf{r}_i)} \times \\ & \times \sum_l \left\{ \frac{dv_l^a(r)}{dr} \Big|_{r=r_{ia}} \frac{\mathbf{r}_{ia}}{r_{ia}} P_l(\cos \theta_q^a) + v_l^a(r_{ia}) \frac{dP_l(\cos \theta)}{d \cos \theta} \Big|_{\cos \theta_q^a} \left(\cos \theta_q^a \frac{\mathbf{r}_{ia}}{r_{ia}} - \hat{\mathbf{u}}_q \right) \frac{1}{r_{ia}} \right\} + \\ & + \sum_l v_l^a(r_{ia}) \sum_q w_q P_l(\cos \theta_q^a) \left\{ \nabla'_i [\phi_j(\mathbf{r}'_i) J(\mathbf{r}'_i)]|_{\mathbf{r}'_i=\mathbf{q}_i^a} - \nabla'_i [\phi_j(\mathbf{r}'_i) J(\mathbf{r}'_i)]|_{\mathbf{r}'_i=\mathbf{q}_i^a} \cdot \hat{\mathbf{u}}_q \frac{\mathbf{r}_{ia}}{r_{ia}} \right\} \frac{1}{J(\mathbf{r}_i)} \\ & + \sum_b \sum_l v_l^b(r_{ib}) \sum_q w_q P_l(\cos \theta_q^b) \nabla_a \left[\phi_j(\mathbf{r}'_i) \frac{J(\mathbf{r}'_i)}{J(\mathbf{r}_i)} \right]_{\mathbf{r}'_i=\mathbf{q}_i^b}, \end{aligned} \quad (\text{A3})$$

where we simplified the notation as $J(\mathbf{r}'_i)/J(\mathbf{r}_i) = J(\dots, \mathbf{r}'_i, \dots)/J(\dots, \mathbf{r}_i, \dots)$. Therefore, differentiating the term in the local energy due to the non-local potential results in a very compact formula (instead of the multiple expressions presented in Ref. 17). This requires only the gradients of the orbitals and Jastrow factor with respect to the electronic and nuclear positions (Eq. A1) computed at the quadrature points, and simple quantities such as some geometrical terms and the derivatives of the radial components of the non-local potentials.

When the determinantal component of ψ is a sum of determinants, we just need to compute the rectangular extensions $\nabla_a \tilde{A}$ and $\nabla_a \tilde{B}$ of the matrices $\nabla_a A$ and $\nabla_a B$. Calculations of $\nabla_a(\ln \psi)$, and $\nabla_a E_L$ are then straightforward using expressions (28) and (43).

2. Warped coordinates

An improved estimator of forces (and also other observables) is obtained through the use of warped coordinates^{5,6} which, as detailed in Refs. 4 and 7, introduces additional terms in the force estimator: For any component of the force, one also needs to compute $\mathbf{v} \cdot \nabla(\ln \psi)$ and $\mathbf{v} \cdot \nabla E_L$, where the gradient is taken with respect to the $3N$ electron coordinates. The vector field \mathbf{v} depends on the electron and nuclear positions, and is different for the force components of the different atoms. These two terms can be written as first derivatives of $\ln \psi$ and E_L

$$\mathbf{v} \cdot \nabla \ln \psi(\mathbf{R}) = \left. \frac{d}{d\mu} \ln \psi(\mathbf{R} + \mu \mathbf{v}) \right|_{\mu=0} \quad (\text{A4})$$

$$\mathbf{v} \cdot \nabla E_L(\mathbf{R}) = \left. \frac{d}{d\mu} E_L(\mathbf{R} + \mu \mathbf{v}) \right|_{\mu=0}. \quad (\text{A5})$$

When ψ is a single determinant times a Jastrow factor $\psi = J \det(A)$, the first term (A4) is

$$\mathbf{v} \cdot \frac{\nabla(J \det(A))}{J \det(A)} = \text{tr}(A^{-1} \mathbf{v} \cdot \nabla A) + \mathbf{v} \cdot \frac{\nabla J}{J} \quad (\text{A6})$$

where the coefficients of the matrix $\mathbf{v} \cdot \nabla A$ are

$$(\mathbf{v} \cdot \nabla A)_{ij} \equiv \mathbf{v} \cdot \nabla A_{ij} = \mathbf{v}_i \cdot \nabla \phi_j(\mathbf{r}_i), \quad (\text{A7})$$

where, in the last term, only the 3 components corresponding to the i -th electron survive. The second expression (A5) is the derivative of the local energy $E_L = \text{tr}(A^{-1} B)$, so $\mathbf{v} \cdot \nabla E_L$ is given by the expression (15) with

$$\partial_\mu B = \mathbf{v} \cdot \nabla B \quad \text{where } (\mathbf{v} \cdot \nabla B)_{ij} \equiv \mathbf{v} \cdot \nabla B_{ij}. \quad (\text{A8})$$

We recall that, when pseudopotentials are employed, $B \equiv B^{\text{kin}} + B^{\text{NL}}$ with B^{kin} and B^{NL} given in Eqs. (33) and (37). The expression of ∇B^{NL} includes a subset of the terms required to evaluate Eq. A3.

When the determinantal component of ψ is a sum of determinants, we simply need to compute the rectangular extensions $\mathbf{v} \cdot \nabla \tilde{A}$ and $\mathbf{v} \cdot \nabla \tilde{B}$ of the matrices $\mathbf{v} \cdot \nabla A$ and $\mathbf{v} \cdot \nabla B$. Again, calculations of (A4) and (A5) are then straightforward using expressions (28) and (43). Note that, if we want to keep the calculation of $\mathbf{v} \cdot \nabla \tilde{A}$ and $\mathbf{v} \cdot \nabla \tilde{B}$ of order $O(N^2)$, the vector field \mathbf{v} should be localized around the atom whose force we are evaluating, i.e. $\mathbf{v}_i(\mathbf{r}_i) = 0$ when the distance between the electron \mathbf{r}_i and the atom we are considering is larger than a given threshold. This is in fact how the space-warp transformation was introduced in Refs. 5 and 6.

Appendix B: First derivative using the Sherman-Morrison-Woodbury formula

We can also obtain (28) using the Sherman-Morrison-Woodbury formula instead of performing the derivative of (25). This is the strategy exploited in Ref. 8. The calculation is much less straightforward, but allows us to understand the relationship with this reference and Ref. 9.

We need to update the trace expression (4) when the Slater matrix A and its derivative B are replaced by an excited Slater matrix \bar{A} and its derivative \bar{B} ,

$$\text{tr}(\bar{A}^{-1} \bar{B}) = \text{tr}(\bar{A}^{-1} B + \bar{A}^{-1}(\bar{B} - B)). \quad (\text{B1})$$

Writing $\bar{A} = A + (\bar{A} - A)PP^T$ and applying the Sherman-Morrison-Woodbury formula, we have

$$\begin{aligned}\bar{A}^{-1} &= A^{-1} - A^{-1}(\bar{A} - A)P(1 + P^T A^{-1}(\bar{A} - A)P)^{-1} P^T A^{-1} \\ &= A^{-1} - A^{-1}(\bar{A} - A)P(P^T A^{-1} \bar{A} P)^{-1} P^T A^{-1}\end{aligned}\quad (\text{B2})$$

We will also make use of the following identity

$$P^T \bar{A}^{-1} = (P^T A^{-1} \bar{A} P)^{-1} P^T A^{-1} \quad (\text{B3})$$

which is easily obtained by multiplying both sides of equation (B2) by P^T . Exploiting that $(\bar{B} - B) = (\bar{B} - B)PP^T$ and Eq. B2, Eq. B1 becomes

$$\begin{aligned}\text{tr}(\bar{A}^{-1} \bar{B}) &= \text{tr}(A^{-1} B) - \text{tr}((P^T A^{-1} \bar{A} P)^{-1} P^T A^{-1} B A^{-1} (\bar{A} - A) P) \\ &\quad + \text{tr}((P^T A^{-1} \bar{A} P)^{-1} P^T A^{-1} (\bar{B} - B) P)\end{aligned}$$

To obtain this expression we have used the cyclic property of the trace, and Eq. B3. Using that $A^{-1} A = 1$ the last expression simplifies

$$\text{tr}(\bar{A}^{-1} \bar{B}) = \text{tr}(A^{-1} B) + \text{tr}((P^T A^{-1} \bar{A} P)^{-1} P^T (A^{-1} \bar{B} - A^{-1} B A^{-1} \bar{A}) P), \quad (\text{B4})$$

Reminding that we can write $\bar{A}P = \tilde{A}Q$ and $\bar{B}P = \tilde{B}Q$ we recover the expression (28). Note that, in the method proposed by Scemama *et al.*⁹, the inverses of all excited determinants are updated with the Sherman-Morrison formula, which amounts to the restriction of (B2) to one-column updates¹⁸. The left-hand side of (B1) is then computed straightforwardly. The computational scaling of these updates is $O(N^2)$ per excited determinant, leading to an overall scaling of $O(N^3) + O(N^2 N_e)$ for a large number N_e of excited determinants distinct in both spin components ($N_e \gg N$). The formula (B4) avoids these updates for excited determinants and the scaling is reduced to $O(N^3) + O(N_e)$.

Appendix C: Symmetric and more compact formulas, higher-order derivatives

We rewrite the formulas (28) and (43) using a slightly different notation which emphasizes how the building blocks in our formulation are products of A^{-1} with other matrices, and which is more convenient when handling higher-order derivatives. As we had done in Eq. 19, we express the formulas symmetrically in λ and μ when second derivatives with respect to λ and μ are considered. Note that a straightforward implementation of these formulas is not necessarily the most efficient. In particular, one should use the matrix X as in (15), when there are a large number of parameters μ as compared to λ in a small determinantal expansion.

We note $\tilde{a} \equiv A^{-1} \tilde{A}$, $\tilde{a}_\lambda \equiv A^{-1} \partial_\lambda \tilde{A}$, $\tilde{a}_{\lambda\mu} \equiv A^{-1} \partial_{\lambda\mu} \tilde{A}$. This notation extends naturally to derivatives of any order.

$$\tilde{a}_{\alpha_1 \dots \alpha_n} \equiv A^{-1} \partial_{\alpha_1 \dots \alpha_n} \tilde{a} \quad (\text{C1})$$

These matrices are the basic quantities from which any usual quantity can be simply expressed. The columns of these matrices which corresponds to occupied orbitals in the reference slater matrix A (usually the first N columns of \tilde{A}) are $a_\lambda \equiv A^{-1} \partial_\lambda A$, $a_\mu \equiv A^{-1} \partial_\mu A$, $a_{\lambda\mu} \equiv A^{-1} \partial_{\lambda\mu} A$ and so on. For any order $n \geq 0$, the following algebraic identity holds

$$\partial_\alpha \tilde{a}_{\alpha_1 \dots \alpha_n} = \tilde{a}_{\alpha \alpha_1 \dots \alpha_n} - a_\alpha \tilde{a}_{\alpha_1 \dots \alpha_n}. \quad (\text{C2})$$

With this formula, the derivative of the rectangular extension of the slater matrix becomes

$$\partial_\lambda \tilde{a} = \tilde{a}_\lambda - a_\lambda \tilde{a}, \quad (\text{C3})$$

which is the matrix \tilde{M} given in Eq. 18. The second-order derivative is

$$\begin{aligned}\partial_{\lambda\mu} \tilde{a} &= \tilde{a}_{\lambda\mu} - a_\mu \tilde{a}_\lambda \\ &\quad - a_\lambda (\tilde{a}_\mu - a_\mu \tilde{a}) \\ &\quad - (a_{\lambda\mu} - a_\mu a_\lambda) \tilde{a} \\ &= \tilde{a}_{\lambda\mu} + (-a_{\lambda\mu} + a_\mu a_\lambda + a_\lambda a_\mu) \tilde{a} - a_\lambda \tilde{a}_\mu - a_\mu \tilde{a}_\lambda.\end{aligned}\quad (\text{C4})$$

It is a symmetric expression in the parameters λ and μ of the matrix $\partial_\mu \tilde{M}$ given in Eq. 44. Every quantity introduced in this paper depends on the logarithmic derivatives of

$$\det(\bar{A}) = \det(A) \det(P^T \tilde{a} Q). \quad (\text{C5})$$

The first-order derivative is

$$\partial_\lambda \ln(\det(\bar{A})) = \text{tr}(a_\lambda) + \text{tr}((P^T \tilde{a} Q)^{-1} P^T \partial_\lambda \tilde{a} Q). \quad (\text{C6})$$

In analogy to Eq. C1, we introduce the notation

$$p_l \equiv (P^T \tilde{a} Q)^{-1} P^T (\partial_l \tilde{a}) Q. \quad (\text{C7})$$

where l is a list of n derivation parameters: $l \equiv (\alpha_1 \dots \alpha_n)$. The same algebraic identity as (C2) holds

$$\partial_\alpha p_l = p_{\alpha l} - p_\alpha p_l. \quad (\text{C8})$$

With this notation, the first derivative is

$$\partial_\lambda \ln(\det(\bar{A})) = \text{tr}(a_\lambda) + \text{tr}(p_\lambda), \quad (\text{C9})$$

and the second derivative is

$$\partial_{\mu\lambda} \ln(\det(\bar{A})) = \text{tr}(a_\lambda - a_\mu a_\lambda) + \text{tr}(p_{\lambda\mu} - p_\mu p_\lambda). \quad (\text{C10})$$

We can also write the third-order derivative

$$\begin{aligned} \partial_{\alpha\mu\lambda} \ln(\det(\bar{A})) = & \text{tr}(a_{\alpha\lambda\mu} + (-a_{\lambda\mu} + a_\mu a_\lambda + a_\lambda a_\mu) a_\alpha - a_{\lambda\alpha} a_\mu - a_{\mu\alpha} a_\lambda) \\ & + \text{tr}(p_{\alpha\lambda\mu} + (-p_{\lambda\mu} + p_\mu p_\lambda + p_\lambda p_\mu) p_\alpha - p_{\lambda\alpha} p_\mu - p_{\mu\alpha} p_\lambda). \end{aligned} \quad (\text{C11})$$

In all these expressions, the first trace is the logarithmic derivative of the original determinant (occupied orbitals) and the second trace is the corrective term for a k^{th} -order excitation. Both terms have exactly the same algebraic structure.

The $k \times k$ matrices p_l are easy to compute: One uses expression (C7), the matrix $P^T \partial_l \tilde{a} Q$ being a simple square submatrix of $\partial_l \tilde{a}$, which is itself given by the recursion formula (C2) or expressions (C3) and (C4). All these quantities depend only on $\tilde{a} = A^{-1} \tilde{A}$ and the related transformed derivatives $\tilde{a}_l = A^{-1} \partial_l \tilde{A}$. Writing expressions for higher-order derivatives is straightforward.

Appendix D: Update of A^{-1} and $A^{-1} \tilde{A}$ for one-electron moves and scaling

We discuss here the computational cost of a typical QMC run using the formulas we developed. We assume that we compute the energy in either variational or diffusion Monte Carlo before performing a full sweep over the electrons in an all-electron move or N one-electron moves. The cost of such a QMC run is summarized in Table I where the main operations are the computation of the local energy before the sweep and the update of the wave function and the drift velocity of the electron being moved during the sweep. We note that the cost of updating $A^{-1} \tilde{A}$ using a straightforward matrix multiplication is $O(N^2 N_s)$ but that the dependence on N_s can be made N times smaller by updating A^{-1} and $A^{-1} \tilde{A}$ as described below. The overall cost to build the sampling process is then $O(N^3) + O(N^2 N_{\text{virt}}) + O(N N_s) + O(N N_e)$ where $N_s \leq N_{\text{act}} N_{\text{virt}}$. With the exception of the $O(N N_s)$ term **instead of** $O(N^2 N_s)$, this is the same scaling reported in Ref. 8, where only the computation of the kinetic component in the local energy was however discussed.

To understand how to update the matrix A^{-1} and $A^{-1} \tilde{A}$ during the sweep, let us assume that the i^{th} electron has been moved. The new Slater matrix A_e differs from A only in one row. The same remark holds for the rectangular extension \tilde{A}_e to virtual orbitals, so only the i -th row is non-zero in $\tilde{A}_e - \tilde{A}$. Using the Sherman-Morrison formula,

$$A_e^{-1} = A^{-1} - \frac{A^{-1}(A_e - A)A^{-1}}{(A_e A^{-1})_{ii}}, \quad (\text{D1})$$

the inverse matrix can be updated at a cost $O(N^2)$. Writing $A_e^{-1} \tilde{A}_e = A_e^{-1}(\tilde{A} + (\tilde{A}_e - \tilde{A}))$, using expression (D1) and expanding the product, we obtain

$$\begin{aligned} A_e^{-1} \tilde{A}_e = & A^{-1} \tilde{A} + A^{-1}(\tilde{A}_e - \tilde{A}) \\ & - \frac{A^{-1}(A_e - A)A^{-1} \tilde{A}}{(A_e A^{-1})_{ii}} \\ & - \frac{A^{-1}(A_e - A)A^{-1}(\tilde{A}_e - \tilde{A})}{(A_e A^{-1})_{ii}} \end{aligned} \quad (\text{D2})$$

Since $A^{-1} \tilde{A}$ is known and all the products needed to compute the last three terms always involve a matrix with only one non-zero row, the overall scaling for updating A^{-1} and $A^{-1} \tilde{A}$ in a one-electron move is $O(N^2) + O(N N_{\text{virt}})$. For a sweep (N one-electron moves), the cost for these updates is then $O(N^3) + O(N^2 N_{\text{virt}})$.

| Step | Operation | Cost |
|----------------------|---|-------------------------------|
| Before the sweep | Compute A , A^{-1} , and B for E_L | $O(N^3)$ |
| | Compute \tilde{A} and \tilde{B} for E_L | $O(N^2 N_{\text{virt}})$ |
| | Compute $A^{-1} \tilde{A}$ | $O(N^2 N_{\text{virt}})$ |
| | Compute \tilde{M} for E_L | $O(N^2 N_{\text{virt}})^{16}$ |
| | Compute N_e excited determinants and corresponding E_L | $O(N_e)$ |
| Sweep over electrons | Compute \tilde{B} for drift of i -th electron | $O(N N_{\text{virt}})$ |
| | Compute \tilde{M} for drift of i -th electron | $O(N N_{\text{virt}} + N_s)$ |
| | Update A , A^{-1} , and B for drift of i -th electron | $O(N^2)$ |
| | Update \tilde{A} and \tilde{B} for drift of i -th electron | $O(N N_{\text{virt}})$ |
| | Update $A^{-1} \tilde{A}$ | $O(N N_{\text{virt}})$ |
| | Compute \tilde{M} for i -th drift velocity | $O(N N_{\text{virt}} + N_s)$ |
| | Compute N_e excited determinants and corresponding i -th drift velocity | $O(N_e)$ |

TABLE I. Computational cost of a typical QMC run. We compute the local energy before the sweep over the electrons, and update the wave function and the drift of the electron being moved during the sweep.

ACKNOWLEDGMENTS

We thank Bryan Clark, Miguel Morales and Cyrus Umrigar for useful discussions. C.F. acknowledges support from an ECHO grant (712.011.005) of the Netherlands Organisation for Scientific Research (NWO). This work was supported in part by the Italian MIUR through PRIN 2011.

¹W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).

²J. Kolorenč and L. Mitas, Rep. Prog. Phys. **74**, 026502 (2011).

³Generalization to more general cases such as determinants with (short range) backflow, geminals or pfaffians will be presented elsewhere.

⁴R. Assaraf and M. Caffarel, J. Chem. Phys. **119**, 10536 (2003).

⁵C. J. Umrigar, Int. J. Quantum Chem. **23**, 217 (1989).

⁶C. Filippi and C. J. Umrigar, Phys. Rev. B **61**, R16291 (2000).

⁷S. Sorella and L. Capriotti, J. Comp. Phys. **133**, 234111 (2010).

⁸B. K. Clark, M. A. Morales, J. McMinis, J. Kim, and G. E. Scuseria, J. Chem. Phys. **135**, 244105 (2011).

⁹A. Scemama, T. Applencourt, E. Giner, and M. Caffarel, <http://arxiv.org/abs/1510.00730>.

¹⁰The local energy contains one-body operators (the kinetic energy and the possibly non-local electron-nucleus potential), and the two-body electron-electron potential; the latter is diagonal in R and the calculation of both its value and its derivatives is trivial.

¹¹Given a set of M_b basis functions, one can construct M_b orbitals (e.g. the occupied and virtual Hartree-Fock orbitals) which span the same space. Performing orbital variations by mixing orbitals recovers the same variational freedom as varying individual orbital coefficients and offers several advantages. One can for instance easily preserve the orbital symmetry or omit redundant rotations.

¹²CHAMP is a quantum Monte Carlo program package written by C. J. Umrigar, C. Filippi, S. Moroni and collaborators.

¹³M. Burkatzki, C. Filippi, and M. Dolg, J. Chem. Phys. **126**, 234105 (2007).

¹⁴For the hydrogen atom, we use a more accurate BFD pseudopotential and basis set. Dolg, M.; Filippi, C., private communication.

¹⁵C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, Phys. Rev. Lett. **98**, 110201 (2007).

¹⁶If the products are carried out from the left to the right, the factor N_{virt} appearing in the scaling must be substituted with N_{act} .

¹⁷A. Badinski and R. J. Needs, Phys. Rev. E **76**, 036707 (2007).

¹⁸In Ref. 9, a tree from the set of excited determinants is built such that a child Slater determinant differs from his parent in only one column.