JCTC Journal of Chemical Theory and Computation

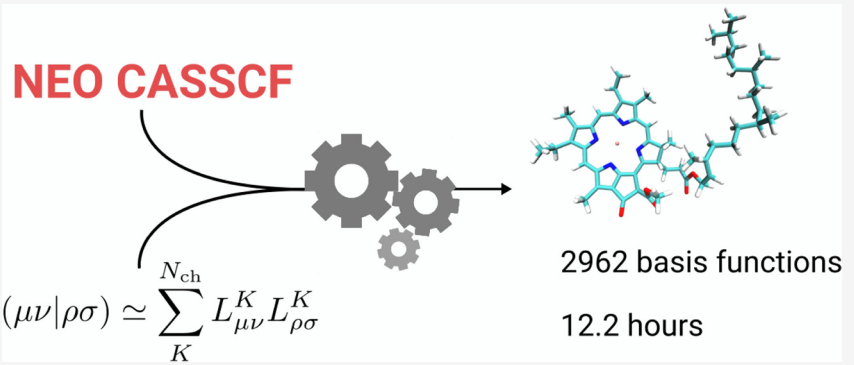
pubs.acs.org/JCTC Article

Second-Order CASSCF Algorithm with the Cholesky Decomposition of the Two-Electron Integrals

Tommaso Nottoli, Jürgen Gauss, and Filippo Lipparini*

Cite This: https://doi.org/10.1021/acs.jctc.1c00327 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: In this contribution, we present the implementation of a second-order complete active space−self-consistent field (CASSCF) algorithm in conjunction with the Cholesky decomposition of the two-electron repulsion integrals. The algorithm, called norm-extended optimization, guarantees convergence of the optimization, but it involves the full Hessian and is therefore computationally expensive. Coupling the second-order procedure with the Cholesky decomposition leads to a significant reduction in the computational cost, reduced memory requirements, and an improved parallel performance. As a result, CASSCF calculations of larger molecular systems become possible as a routine task. The performance of the new implementation is illustrated by means of benchmark calculations on molecules of increasing size, with up to about 3000 basis functions and 14 active orbitals.

1. INTRODUCTION

The complete active space−self-consistent field (CASSCF) method[1−3] is a powerful tool to achieve a qualitatively correct description of strongly correlated systems. Thanks to its intrinsic multireference nature, it can be used to compute the structure and molecular properties of a large manifold of interesting systems that are poorly described with standard single-reference methods. These include many open-shell systems, molecules with stretched bonds, and therefore reactivity, excited states, and others. It can also provide a starting point for subsequent high-level correlated treatments, such as internally contracted multireference-configuration interaction[4,5] (CI) and coupled cluster,[6−13] multireference perturbation theory such as CASPT2[14,15] and NEVPT2,[16−18] or even quantum Monte Carlo methods.[19,20] Unfortunately, the method suffers from three major complications that restrict its applicability. First, it is not a black box method, as it requires the user to select the active space for the calculation. While there are a few strategies to aid the selection,[21−23] achieving good results relies still on the user's chemical intuition and understanding of the system. Second, the CASSCF wave function's optimization problem is notoriously hard to converge. Third, the method is computationally very demanding.

The computational cost of a CASSCF calculation stems from two concurring factors. The most prominent one is that the method requires one to solve a full CI (FCI) problem in the active space. Due to the combinatorial scaling of FCI, the investigation of large active spaces is not possible using standard direct CI techniques. Approximations to the FCI wave function can be used to overcome this otherwise overwhelming barrier, the most common example being the use of a density-matrix renormalization group[24] (DMRG). However, many interesting systems can be successfully described with a relatively small active space (up to 12−14 electrons in as many orbitals). If a careful choice of the active space that allows one to capture the static correlation of the wave function with a limited number of active orbitals is possible, the cost of the CI part is either negligible (for active spaces with less than 10 orbitals) or manageable with

Received: April 3, 2021

traditional implementations. In such cases, the cost of the calculation is dominated by the operations involving the manipulation of the electron repulsion integrals (ERIs).

Convergence problems can be mitigated, if not completely solved, by using an optimization algorithm with guaranteed convergence to the closest local minimum. Methods based on a restricted step second-order optimization offer such a guarantee and are, therefore, a very attractive option. However, because they involve the evaluation of the energy Hessian with respect to the variational parameters, i.e., orbital rotations and CI expansion coefficients, they are in general more expensive than their first-order counterparts and require cumbersome and involved implementations. Nevertheless, second-order CASSCF implementations have been successfully achieved and are based on two main algorithms. The first algorithm, originally proposed by Werner and Meyer[25] and further refined by Werner, Knowles, and others,[26,27] is based on the definition of a model energy function which is infinite order in orbital rotations and that is optimized. The coupling between CI and orbital optimization is introduced up to second order, ensuring thus quadratic convergence. This algorithm shows excellent convergence properties and overall performances. A similar strategy has been followed by Sun et al.,[28] and the resulting algorithm, which is based on an integral-direct implementation and can use DMRG as a CASSCF solver, exhibits impressive performances. A second choice is to use a more traditional trust-region second-order method, such as the Levenberg−Marquardt method.[29] Augmented with an adaptive choice of the trust radius, as proposed by Fletcher (we refer to the global strategy as FLM), it is possible to prove that the FLM method always converges to the closest local minimum and that the rate of convergence is quadratic. A very efficient implementation of the FLM method, known as the norm-extended optimization (NEO) algorithm, has been proposed by Jensen and co-workers.[30,31] In this contribution, we follow the latter strategy, which we have previously implemented in the CFOUR[32,33] suite of programs.

A second-order CASSCF implementation requires one to work with ERIs transformed in the molecular orbital (MO) basis with at least two indices spanning the full rank of MOs. The transformation of the ERIs from the atomic orbitals (AOs) to the MO basis is expensive, requiring $O(MN_b^4)$ floating point operations, where M is the number of internal and active orbitals and N_b the number of basis functions. Furthermore, it is not easily implemented in an efficient way. This is due to the fact that the ERIs matrix is usually too large to fit in memory, especially in the MO basis, which implies that the transformation involves slow disk I/O. Furthermore, the AO ERIs are computed (and stored) in an order that depends on the shell structure of the basis set for the specific system. As a consequence, the integrals are read (or recomputed, in integral-direct implementations) in a system-dependent order, which makes the use of efficient BLAS routines[34,35] and, more in general, vectorization, particularly challenging.

To address the computational cost involved with the manipulation of the ERIs, it is possible to adopt a low-rank approximation of the ERIs, such as density fitting[36−44] (DF) or Cholesky decomposition[45−51] (CD). Both techniques have been successfully applied in many contexts of quantum chemistry,[52−61] including CASSCF.[62−64] The CD technique is particularly attractive, as it allows a rigorous, a priori control of the approximation error. Furthermore, it offers a compact representation of the ERIs that is well-suited for vectorized, efficient implementations, as the Cholesky-decomposed ERIs, can be often kept in memory with standard computer hardware and are easily manipulated using highly optimized level 3 BLAS routines. Furthermore, all of the ERIs manipulation can be written as the sum of independent operations on a given Cholesky vector and are therefore very easy to parallelize. Other approaches that aim at large-scale applications are present in literature; see for instance refs 65 and 66, where a first-order implementation on graphical processing units is shown, and the computational cost in the orbital part is mitigated by exploiting the sparsity of the two-electron integrals.

In this contribution, we present an implementation of NEO CASSCF in the CFOUR suite of programs[32,33] based on the CD of the ERIs. The implementation is tested on several molecular systems of increasing size, for active spaces that go from small (CAS(6,6)) to large (CAS(14,14)) and using up to about 3000 basis functions. This work is organized as follows. In Section 2, the derivation of the NEO CASSCF method is reviewed. The implementation of the algorithm is discussed in Section 3 with a special focus on the Cholesky implementation. In Section 4, benchmark calculations are presented for the purpose of showing the performance of the algorithm in the optimization of medium-to-large systems. Finally, concluding remarks and some perspectives on future developments are given in Section 5.

2. NORM-EXTENDED OPTIMIZATION CASSCF

In this section, we recapitulate the main aspects of NEO CASSCF. First, the parametrization of the wave function is discussed in Section 2.1. Then, the NEO algorithm is briefly summarized in Section 2.2. Further details regarding the optimization algorithm can be found in ref 30 or in a previous work by two of us.[67]

2.1. Parametrization of the CASSCF Wave Function.

The starting point for the following discussion is given by a set of molecular orbitals (MOs) $\{\varphi_p\}_{p=1}^{N_b}$, where N_b is the number of basis functions. In CASSCF, the MOs are subdivided into three classes according to their allowed occupation number in a Slater determinant—namely, *internal*, which are always doubly occupied; *active*, which are subjected to no restriction; and *external*, which are always empty. To distinguish an orbital among such classes, the following labels are used: i, j, k refer to inactive, u, v, x to active, a, b, c to external, and p, q, r to generic orbitals. Indices that run over the determinantal space are labeled with capital letters I, J.

A convenient parametrization for the wave function, first proposed by Jensen and Jørgensen,[30] is

$$|\Psi\rangle = e^{-\hat{\kappa}} \frac{|0\rangle + \hat{P}|\mathbf{c}\rangle}{\||0\rangle + \hat{P}|\mathbf{c}\rangle\|} \tag{1}$$

Here, $|0\rangle = \sum_I^{N_{det}} c_I^{(0)} |\Phi_I^{(0)}\rangle$ is the current approximation to the wave function, or current expansion point (CEP). $|\mathbf{c}\rangle$ is the correction vector that collects the CI variational parameters, c_I,

$$|\mathbf{c}\rangle = \sum_I^{N_{det}} c_I |\Phi_I\rangle \tag{2}$$

and $\hat{P} = 1 - |0\rangle\langle 0|$ is the operator that projects $|\mathbf{c}\rangle$ in the orthogonal complement of $|0\rangle$, thus keeping any redundant vector parallel to the CEP.

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Orbital variations are described through a unitary transformation, $\bar{\boldsymbol{\varphi}} = \boldsymbol{\varphi}\mathbf{U}$, that is conveniently parametrized by using an exponential map

$$U = e^{-\kappa}; \quad \hat{\kappa} = \sum_{p>q} \kappa_{pq}(\hat{E}_{pq} - \hat{E}_{qp}) = \sum_{p>q} \kappa_{pq}\hat{E}^{-}_{pq} \tag{3}$$

where $\hat{E}_{pq} = \sum_{\sigma} \hat{a}^{\dagger}_{p\sigma}\hat{a}_{q\sigma}$ is the spin-traced singlet excitation operator. The variational parameters are given by the elements of the antisymmetric matrix, $\boldsymbol{\kappa}$. Since only rotations between different orbitals classes produce a variation in the energy, the expression for $\hat{\kappa}$ can be simplified as follows:

$$\hat{\kappa} = \sum_{i}^{N_{int}}\sum_{u}^{N_{act}} \kappa_{iu}\hat{E}^{-}_{iu} + \sum_{i}^{N_{int}}\sum_{a}^{N_{ext}} \kappa_{ia}\hat{E}^{-}_{ia} + \sum_{a}^{N_{ext}}\sum_{u}^{N_{act}} \kappa_{au}\hat{E}^{-}_{au} \tag{4}$$

Hence, $\boldsymbol{\kappa}$ is considered as a vector whose dimension is given by all nonredundant orbitals rotations; i.e., $N_{rot} = N_{int}N_{act} + N_{int}N_{ext} + N_{act}N_{ext}$.

2.2. Optimization of the CASSCF Wave Function. Equation 1 is used to define a variational expression for the electronic energy that reads

$$\mathcal{E}(\boldsymbol{\kappa}, \mathbf{c}) = \frac{(\langle 0| + \langle \mathbf{c}|\hat{P})e^{\hat{\kappa}}\hat{\mathcal{H}}e^{-\hat{\kappa}}(|0\rangle + \hat{P}|\mathbf{c}\rangle)}{1 + \langle \mathbf{c}|\hat{P}|\mathbf{c}\rangle} \tag{5}$$

In eq 5, $\hat{\mathcal{H}}$ is the nonrelativistic Hamiltonian operator written in second quantization

$$\hat{\mathcal{H}} = \sum_{pq} h_{pq}\hat{E}_{pq} + \frac{1}{2}\sum_{pqrs} (pq|rs)\hat{e}_{pqrs} + E_{nuc} \tag{6}$$

where

$$\hat{e}_{pqrs} = \hat{E}_{pq}\hat{E}_{rs} - \delta_{rq}\hat{E}_{ps} \tag{7}$$

h_{pq} are one-electron integrals, $(pq|rs)$ are two-electron integrals written in Mulliken's notation, and E_{nuc} is the nuclear repulsion term. A second-order algorithm can be developed by defining a quadratic model for the energy; therefore, we expand eq 5 in power series up to second order. To this end, it is useful to define a generic parameter point, $\mathbf{x} = (\mathbf{c}, \boldsymbol{\kappa})$, and the reference one, $\mathbf{x}_0 = (\mathbf{c}^{(0)}, \mathbf{0})$ such that

$$\mathcal{E}(\mathbf{x}) \approx Q(\mathbf{x}) \doteq E_0 + \mathbf{g}^{\dagger}(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^{\dagger}\mathbf{G}(\mathbf{x} - \mathbf{x}_0) \tag{8}$$

In eq 8, E_0 is the reference energy, that is $\langle 0|\hat{\mathcal{H}}|0\rangle$, while $\mathbf{g}$ and $\mathbf{G}$ are respectively the electronic gradient and Hessian evaluated at the CEP. Analytical expression for such quantities can be obtained by direct differentiation of eq 5 and by exploiting the Baker−Campbell−Hausdorff (BCH) formula. The gradient is given as

$$g^{c}_{I} = \frac{\partial \mathcal{E}}{\partial c_I} = 2\langle \Phi_I|\hat{P}\hat{\mathcal{H}}|0\rangle \tag{9}$$

$$g^{o}_{pq} = \frac{\partial \mathcal{E}}{\partial \kappa_{pq}} = \langle 0|[\hat{E}^{-}_{pq}, \hat{\mathcal{H}}]|0\rangle \tag{10}$$

and the Hessian

$$G^{cc}_{I,J} = \frac{\partial^2 \mathcal{E}}{\partial c_I \partial c_J} = 2\langle \Phi_I|\hat{P}(\hat{\mathcal{H}} - E_0)\hat{P}|\Phi_J\rangle$$

$$G^{co}_{I,pq} = \frac{\partial^2 \mathcal{E}}{\partial c_I \partial \kappa_{pq}} = 2\langle \Phi_I|\hat{P}[\hat{E}^{-}_{pq}, \hat{\mathcal{H}}]|0\rangle$$

$$G^{oo}_{pq,rs} = \frac{\partial^2 \mathcal{E}}{\partial \kappa_{pq} \partial \kappa_{rs}} = \frac{1}{2}(1 + \hat{P}_{pq,rs})\langle 0|[\hat{E}^{-}_{pq}, [\hat{E}^{-}_{rs}, \hat{\mathcal{H}}]]|0\rangle \tag{11}$$

The minimization of the quadratic model directly leads to the Newton−Raphson (NR) equations. However, the radius of convergence of NR is small, and the Hessian can be non-positive-definite at the beginning of the optimization leading to incorrect search directions. To overcome this issue, a more robust strategy consists of using a trust-region optimization algorithm, e.g., the Levenberg−Marquardt (LM) method,[29] where the minimization is performed in a reduced domain such that the Hessian has the correct signature. The LM equations can be seen as diagonally shifted NR ones, where the shifting parameter controls the step length to be within a predefined trust radius, R_t. The norm-extended optimization algorithm[30,68] is an elegant way to recast the LM minimization problem into an eigenvalue−eigenvector one:

$$\mathbf{L}(\alpha)\mathbf{y} = \lambda\mathbf{y} \tag{12}$$

where $\mathbf{L}(\alpha)$ is the gradient-scaled augmented Hessian matrix,

$$\mathbf{L}(\alpha) = \mathbf{G} + \alpha(\mathbf{x}_0\mathbf{g}^{\dagger} + \mathbf{g}^{\dagger}\mathbf{x}_0) \tag{13}$$

It can be shown that for ground-state optimization the optimal direction is given by the first eigenvector of $\mathbf{L}(\alpha)$. Once $\mathbf{y}$ is given, the NEO step can be computed as

$$\mathbf{x}(\alpha) = \mathbf{x}_0 + \frac{1}{\alpha}\mathbf{P}\mathbf{y}(\alpha) \tag{14}$$

Here, $\boldsymbol{P}$ is the matrix representation of the projector operator $\hat{P}$. The step length is controlled by the parameter α and can be obtained by solving the equation

$$\|\mathbf{x}(\alpha) - \mathbf{x}_0\| = R_t \tag{15}$$

Eventually, the trust radius is changed adaptively during the optimization procedure according to Fletcher's algorithm.[29] If the energy increases, the step is discarded, and the trust radius is decreased. Otherwise, R_t is either increased or left untouched on the basis of the value of the ratio between the predicted variation of the energy and the actual one. The combined strategy—NEO plus Fletcher's update—leads to an algorithm that always converges to the closest local minimum for well-behaved wave functions.

3. IMPLEMENTATION

In this section, the implementation of the NEO algorithm within the CFOUR[32,33] suite of programs is discussed. In Section 3.1, we present working expressions for the gradient and for the linear transformations that describe the action of the augmented Hessian matrix on a trial vector. In Section 3.2, the Cholesky decomposition of the two-electron integrals is introduced. Details regarding a cost-effective implementation that exploits the Cholesky vectors are given for a specific example.

3.1. Direct NEO Equations. The NEO algorithm can be thought of as a two-level procedure. In the first level—the

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

macroiterations—the parameter hypersurface is scanned by updating the CEP and the MOs. In the second level—the microiterations—a specific NEO eigenvalue−eigenvector problem is iteratively solved with the intention of getting the optimal step direction. At each macroiteration an AOs to MO transformation is performed. Then, the orbital and CI gradient are assembled, and the electronic energy is calculated. The CASSCF energy can be written as

$$E_0 = \sum_{uv} \gamma_{uv} F^I_{uv} + \frac{1}{2}\sum_{uvxy} \Gamma_{uvxy}(uv|xy) + E_i + E_{\text{nuc}} \tag{16}$$

where γ_{uv} and Γ_{uvxy} are the one- and two-body reduced density matrices, respectively, which can be computed as the expectation value of the excitation operators

$$\gamma_{uv} = \langle 0|\hat{E}_{uv}|0\rangle, \quad \Gamma_{uvxy} = \langle 0|\hat{e}_{uvxy}|0\rangle \tag{17}$$

F^I_{pq} are the elements of the inactive Fock matrix

$$F^I_{pq} = h_{pq} + \sum_i [2(pq|ii) - (pi|qi)] \tag{18}$$

and E_i is the energy contribution that stems from the inactive electrons and is called inactive energy,

$$E_i = \sum_i (h_{ii} + F^I_{ii}) \tag{19}$$

Manipulation of eq 10 leads to an antisymmetric expression for the orbital gradient

$$g_{pq} = 2(F_{pq} - F_{qp}) \tag{20}$$

In eq 20 we have introduced the generalized Fock matrix, whose elements can be written in terms of the inactive Fock matrix, active Fock matrix, and **Q** matrix. The last two are defined as follows:

$$F^A_{pq} = \sum_{uv} \gamma_{uv}\left[(pq|uv) - \frac{1}{2}(pu|qv)\right] \tag{21}$$

$$Q_{up} = \sum_{vxy} \Gamma_{uvxy}(pv|xy) \tag{22}$$

As eq 9 states, the CI gradient can be evaluated as the action of the Hamiltonian operator on $|0\rangle$

$$g_I = 2\langle \Phi_I|\sum_{uv} F^I_{uv}\hat{E}_{uv} + \frac{1}{2}\sum_{uvxy}(uv|xy)\hat{e}_{uvxy}|0\rangle + c^{(0)}_I(E_0 - E_i) \tag{23}$$

where the last term is a vector parallel to the CEP that stems from the presence of the projector operator in the wave function definition.

The iterative solution of the NEO eigenvalue−eigenvector problem (microiterations) requires setting up expressions for the matrix−vector product between the augmented Hessian and a trial vector:

$$\begin{pmatrix} \mathbf{L}^{cc} & \mathbf{L}^{co} \\ \mathbf{L}^{oc} & \mathbf{L}^{oo} \end{pmatrix}\begin{pmatrix} \mathbf{v}^c \\ \mathbf{v}^o \end{pmatrix} = \begin{pmatrix} \mathbf{L}^{cc}\mathbf{v}^c \\ \mathbf{L}^{oc}\mathbf{v}^c \end{pmatrix} + \begin{pmatrix} \mathbf{L}^{co}\mathbf{v}^o \\ \mathbf{L}^{oo}\mathbf{v}^o \end{pmatrix} \tag{24}$$

The present implementation makes use of the split-Davidson algorithm,[68] where configurations-only, $\boldsymbol{v}_{\text{conf}} = (\boldsymbol{v}^c, \mathbf{0})$, and orbitals-only, $\boldsymbol{v}_{\text{orb}} = (\mathbf{0}, \boldsymbol{v}^o)$, vectors are added in the Krylov-like subspace. This procedure allows one to adaptively add to the subspace either $\boldsymbol{v}_{\text{conf}}$ or $\boldsymbol{v}_{\text{orb}}$, depending on the part that exhibits the largest residual. Here we report the expressions for the direct product

$$\sum_J L_{I,J} v_J = 2(\langle \Phi_I|\hat{\mathcal{H}}|\mathbf{v}_c\rangle - E_0 v_I) + (\alpha - 1) \times \left[c^{(0)}_I \sum_J v_J g_J + g_I \sum_J v_J c^{(0)}_J\right] \tag{25}$$

$$\sum_I L_{pq,I} v_I = 2g^T_{pq} + (\alpha - 2) g_{pq} \sum_I c^{(0)}_I v_J \tag{26}$$

$$\sum_{pq} L_{I,pq} v_{pq} = 2\langle 0|\tilde{\mathcal{H}}|\Phi_I\rangle + (\alpha - 2) c^{(0)}_I \sum_{pq} g_{pq} v_{rs} \tag{27}$$

$$\sum_{rs} L_{pq,rs} v_{rs} = \langle 0|[\hat{E}^-_{pq}, \tilde{\mathcal{H}}]|0\rangle + \frac{1}{2}\sum_s (g_{sp} v_{qs} - g_{sq} v_{ps}) \tag{28}$$

where $|\boldsymbol{v}_c\rangle = \sum_I v_I|\Phi_I\rangle$, and $\tilde{\mathcal{H}} = [\hat{\kappa}, \hat{\mathcal{H}}]$ is the one-index transformed Hamiltonian operator. In eq 26, we have introduced the transition gradient g^T_{pq}, that is, a gradient computed with symmetrized transition density matrices. The first term of eq 28 is a gradient-like contribution computed with one-index transformed one- and two-electron integrals. It can be effectively computed by means of the transformed inactive Fock matrix, active Fock matrix, and **Q** matrix whose expressions are given below

$$\tilde{F}^I_{pq} = \sum_r (F^I_{pr} v_{qr} - F^I_{rq} v_{pr}) + \sum_r \sum_i [4(pq|ir) - (pr|qi) - (pi|qr)] v_{ir} \tag{29}$$

$$\tilde{F}^A_{pq} = \sum_r (F^A_{pr} v_{qr} - F^A_{rq} v_{pr}) + \sum_r \sum_{uv} \gamma_{uv}\left[2(pq|ur) - \frac{1}{2}(pr|uq) - \frac{1}{2}(pu|qr)\right] v_{ur} \tag{30}$$

$$\tilde{Q}_{up} = \sum_r Q_{ur} v_{pr} + \sum_r \sum_{vxy} \{\Gamma_{uvxy}(pr|xy) + [\Gamma_{uxvy} + \Gamma_{uxyv}](px|ry)\} v_{vr} \tag{31}$$

Explicit expressions for eq 28 are given in the Supporting Information.

The transformed matrices have to be computed at each step of the microiterations, and together with the AO to MO transformation constitute the bottleneck of the algorithm when the chosen active space is small. A summary of the NEO algorithm is given in Figure 1.

3.2. NEO Equations with Cholesky Vectors. The ERI matrix is symmetric and positive-semidefinite; therefore, it can be decomposed according to the Cholesky decomposition:

$$(pq|rs) \simeq \sum_K^{N_{\text{ch}}} L^K_{pq} L^K_{rs} \tag{32}$$

We compute the CD of the integrals using the partial pivoting algorithm proposed by Koch et al.[47] (a much more efficient CD algorithm has been recently proposed; see for instance ref

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

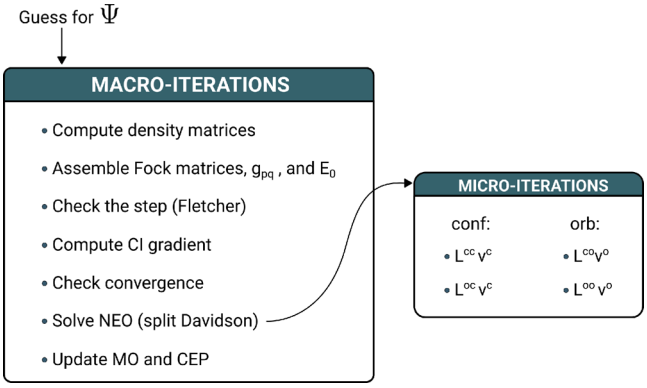


Figure 1. Macro-/microiterations scheme for the NEO algorithm.

51), which has been implemented inside the Mainz integral package[69] (MINT) in CFOUR.[32,33] The procedure stops whenever the residual of the diagonal is below a user defined threshold. Using the Cauchy−Schwarz inequality, it can be shown that the error on the reconstructed integrals is always lower than or equal to the threshold, so it can be controlled systematically. In eq 32, N_{ch} is the number of Cholesky vectors generated; the higher the decomposition threshold the lower the number of Cholesky vectors.

The Cholesky representation of the integrals has been substituted in all equations, namely, the Fock matrices, the transformed Fock matrices, and the active ERI matrix. In order to illustrate the implementation of the evaluation of the aforementioned quantities, we discuss in detail the calculation of the transformed **Q** matrix. Implemented expressions for the transformed Fock matrices can be found in the Supporting Information. Inserting eq 32 into eq 31, we get

$$\tilde{Q}_{up} = \sum_{r} Q_{ur} v_{pr} + \sum_{K}^{N_{ch}} \sum_{r} \sum_{vxy} \{\Gamma_{uvxy} L^{K}_{pr} L^{K}_{xy} + [\Gamma_{uxvy} + \Gamma_{uxyv}] L^{K}_{px} L^{K}_{ry}\} v_{vr} \tag{33}$$

The first term of eq 33 can be straightforwardly computed from the **Q** matrix. The second term is evaluated by first assembling, for each Cholesky vector, the intermediate quantities

$$T^{K}_{uv} = \sum_{xy} \Gamma_{uvxy} L^{K}_{xy} \tag{34}$$

and

$$S^{K}_{vp} = \sum_{r} v_{vr} L^{K}_{rp} \tag{35}$$

such that

$$\sum_{r} \sum_{vxy} \Gamma_{uvxy} L^{K}_{pr} L^{K}_{xy} v_{vr} = \sum_{v} T^{K}_{uv} S^{K}_{vp}. \tag{36}$$

Regarding the last term, we notice that $\sum_r v_{vr} L^{K}_{ry}$ is the fully active part of the intermediate S^K of eq 35. Hence, we define

$$V^{K}_{ux} = \sum_{vy} \Gamma_{uxvy} S^{K}_{vy} \tag{37}$$

and

$$\sum_{vy} \Gamma_{uxyv} S^{K}_{vy} = \sum_{vy} \Gamma_{xuvy} S^{K}_{vy} = (\mathbf{V}^{T})_{ux} \tag{38}$$

where in eq 38 we exploited the symmetry of the two-body reduced density matrix. Gathering together eqs 36, 37, and 38, we can rewrite eq 31 as

$$\tilde{Q}_{up} = \sum_{r} Q_{ur} v_{pr} + \sum_{K}^{N_{ch}} [(\mathbf{T}^{K}\mathbf{S}^{K})_{up} + (\mathbf{X}^{K}\mathbf{L}^{K})_{up}] \tag{39}$$

where $X^{K}_{ux} = V^{K}_{ux} + V^{K}_{xu}$. A detailed discussion of the computational cost and scaling of the present algorithm, together with a critical comparison with other existing CASSCF codes, can be found in Appendix A.

4. BENCHMARKS

In this section we present benchmark calculations to illustrate the performance of the CD-CASSCF implementation. In all of the calculations, convergence is achieved when the root-mean-square norm of both the orbital and CI gradient is below 10^{-7}. The threshold for the Cholesky decomposition has been set to 10^{-4}.

4.1. Importance of a Good Starting Guess. When using a traditional second-order method, such as the one described in the present work, a good starting guess for the molecular orbitals is beneficial for reducing the number of macro-iterations. Also, the quality of the orbitals can affect the number of microiterations required to solve the NEO eigenvalue−eigenvector problem (see eq 24), whose convergence can sometimes be slow. The poor convergence of standard second-order methods has been thoroughly investigated by Werner and co-workers,[25,26] and it is due to the combination of two facts. As the energy is parametrized in terms of orbital rotations, which are defined as the exponential of a skew-symmetric matrix, a second-order approximation cannot faithfully reproduce the periodicity of rotations. Furthermore, the active space is bound to contain almost doubly occupied and almost empty orbitals. If a HF guess is used, the rotations that mix these orbitals with internal and external ones, respectively, will be associated with very small gradients and small Hessian eigenvalues. This in turn translates into a poor convergence radius of a second-order expansion, which makes the use of a trust-radius algorithm paramount. This motivated Werner, Meyer, and Knowles to use a higher order expansion for their MCSCF algorithm, which allows them to achieve an extremely robust and fast convergence. To better illustrate the importance of a good starting guess when using a traditional second-order algorithm and to further describe the behavior of the optimization procedure when a poor one is used, we analyze the convergence of our implementation using the serotonin molecule as an example and Pople's 6-31G basis set.[70] For this example, we use the conventional implementation (without CD). We compare three different starting points. As a production reference, we employ the unrestricted-natural-orbitals (UNOs) criterion of Pulay and Hamilton.[71] This strategy has been recently cross-validated against other active space selection methods,[23] and it has been shown to produce almost the same active space as the AVAS-,[22] FOD-,[72] and DMRG-based strategies.[21,73] To compute the UNO, we first look for triplet instabilities in the RHF wave function by computing the lowest eigenvalues of the corresponding instability Hessian.[74] If one or more negative eigenvalues are present, we perturb the MOs along

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

the direction given by the associated eigenvectors, and we run an UHF calculation. Finally we compute the UNO from the averaged charge density matrix and choose as active orbitals the ones with occupation numbers between 0.01 and 1.99. According to this protocol, the active space for the serotonin consists in 8 electrons and 8 orbitals. A second starting point is obtained from a RHF reference by manually selecting the π/π^* orbitals. Lastly, we choose the four highest occupied and four lowest unoccupied canonical RHF orbitals. In Figure 2, we

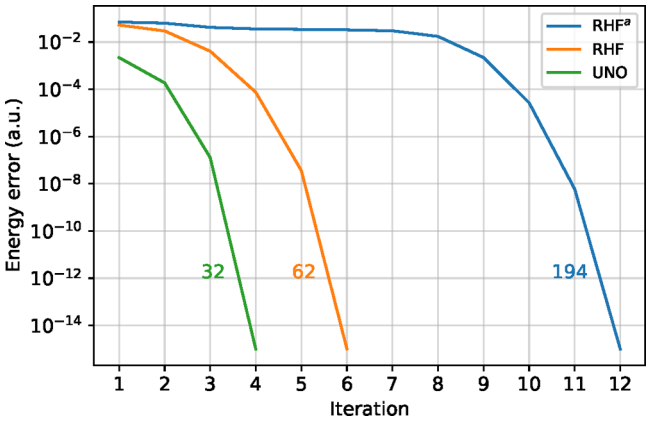


Figure 2. Convergence profile for the CASSCF optimization of the serotonin molecule with 8 electrons in an 8 orbitals active space and 6-31G basis set. In blue (RHF[a]), the active orbitals were selected as the four lowest and four highest canonical RHF; in orange (RHF), as the π/π^* canonical RHF; and in green (UNO) as the unrestricted natural orbitals. Next to the three curves, the total number of microiterations is reported. In all three cases the converged CASSCF energy is −569.279 597 357 hartrees.

report the convergence profile (energy difference with respect to the converged solution as a function of the number of macroiterations) for the three aforementioned starting points. We also report the overall number of microiterations (summed over the number of macroiterations).

All three calculations converged to the same state. The calculation using the RHF π/π^* orbitals converged relatively quickly, but the presence of small eigenvalues in the MO rotation at the beginning of the calculations requires two additional iterations with respect to the calculation starting from the UNO, during which small steps are taken and the gradient remains relatively unchanged. As expected, the last choice presents the worst convergence trend, with the first eight iterations being spent looking for the quadratic region. This can be easily understood. To achieve convergence, the orbitals have to be swapped, which corresponds to a large orbital rotation. Due to the small convergence radius of a second-order expansion, this requires in turn a large number of steps. We illustrate this fact in Figure 3, where we represented the lowest energy MO at various steps of the optimization (the corresponding iterate is reported to the left of the MO picture). We note that in the first steps the orbital changes its shape—the rotation with a π inactive orbital is gradually magnified during the optimization.

This example shows that a good starting guess and a careful selection of the active orbitals are paramount to achieve a smooth and fast convergence. Nevertheless, it also shows that the NEO algorithm is robust and is able to achieve convergence even when starting from a particularly bad reference.

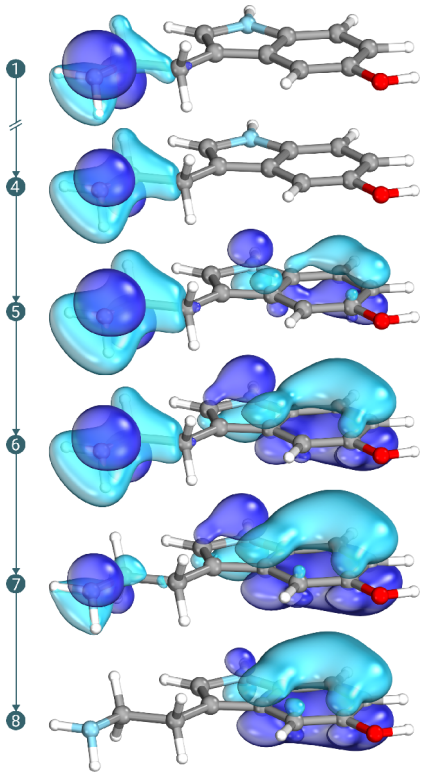


Figure 3. Snapshots of the lowest energy active orbital taken at various CASSCF macroiterations. Reported are only iterations where a qualitative change on the orbital's shape is visible. The starting guess for the orbitals was given by the canonical RHF ones, without any manual choice. The orbitals were visualized using the IboView software.[75,76]

4.2. Benchmark Calculations. The starting benchmark set used to test the CD-CASSCF code is composed of 21 aromatic molecules. The geometries were taken from ref 77; the set was used also by Kreplin et al. to test their MCSCF solver.[27,78] We augmented the benchmark set with 5 larger molecules whose geometries were optimized at the B3LYP/6-31G(d)[70,79] level of theory using the Gaussian 16 suite of programs.[80] All of the new structures can be found in the Supporting Information; a pictorial representation of the molecules is given in Figures 4 and 5. All calculations were done using Dunning's cc-pVTZ[81] basis set using spherical harmonics.

As a starting analysis, we compared the storage requirement of a CD-based CASSCF calculation with a conventional one. The exact CD of the ERI matrix would generate $N_b(N_b + 1)/2$ Cholesky vectors. We define a compression rate

$$f = \frac{N_b(N_b + 1)/2}{N_{ch}} \tag{40}$$

to measure the effectiveness of the truncated CD in reducing the dimension of the computational problem. We report in Table 1, for seven selected molecules from the benchmark set, the number of basis functions (N_b), the disk space needed to store the two-electron integrals, and the compression rate.

As expected, the storage requirements for the CD vectors are significantly lower than for the standard two-electron integrals. It is worth remarking that, even for the largest system of this reduced set, the Cholesky vectors can easily be kept in memory even on a standard desktop computer. This is one of the main

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

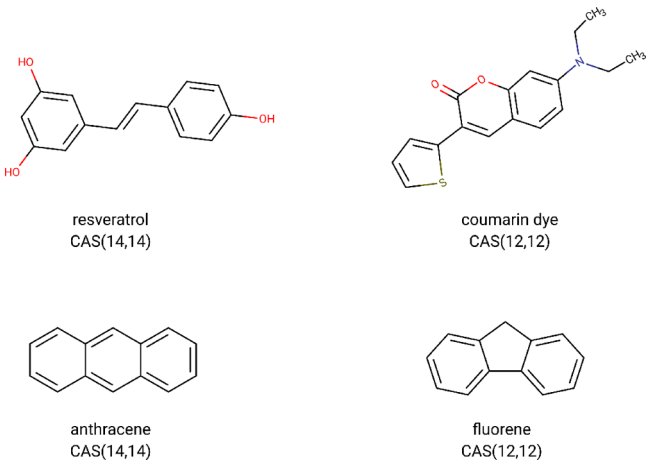


Figure 4. New aromatic molecules used to test the CD-CASSCF algorithm together with their active spaces.

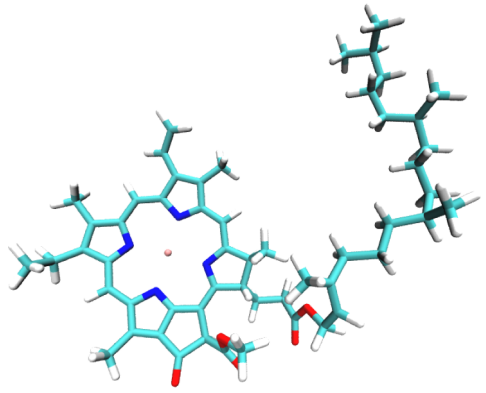


Figure 5. Structure of the chlorophyll molecule.

Table 1. Systems Used to Compare the Different Storage Requirement of the Standard and CD Implementations in CFOUR[a]

			size (GB)	
molecule	N_b	f	CD	STD
catechol	324	24.88	0.9	18.3
naphthalene	412	31.98	1.8	43.1
nicotine	556	43.46	4.4	139.7
tryptophan	618	47.75	6.1	170.6
pyridoxamine	528	41.00	3.8	108.8
2Me4HSdiox	446	34.10	2.3	61.8
indole	368	28.37	1.3	30.4

[a]For each molecule, we report the number of basis functions (N_b), the compression factor (f) defined in eq 40, and the size in gigabytes of the Cholesky vectors (CDs) and two-electron integrals (STD).

advantages of using a reduced order approximation of the ERIs, as it allows one to perform full in-core calculations, avoiding thus slow disk I/O operations. In Figure 6, we plotted the compression rate for the whole benchmark set with respect to the number of basis functions. According to our results, we deduce a linear scaling of f with respect to the system's size. Therefore, we can obtain an even more compact representation of the integrals and, thus, a greater efficiency of the CD, with larger systems.

The previously selected molecules were also used to compare the accuracy of the CD-CASSCF energy with respect

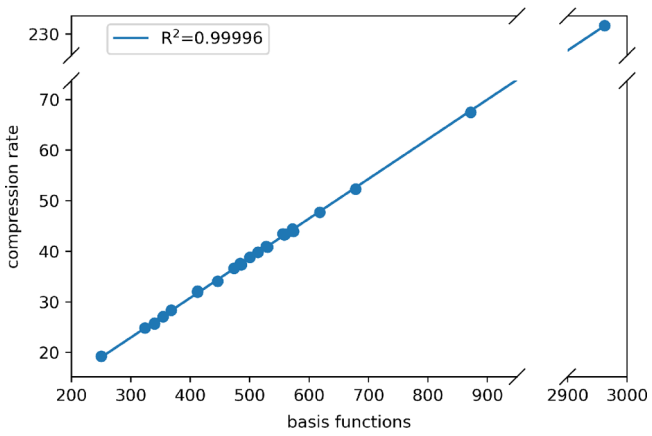


Figure 6. Compression rate trend with respect to the number of basis functions. The points are fitted using standard linear regression.

to the one obtained with the standard algorithm. The chosen decomposition threshold (10^{-4}) allows us to obtain a high compression rate while retaining an overall good accuracy. In Table 2, we report the active space (CAS) and the converged

Table 2. Comparison between the Converged CASSCF Energy of the Cholesky and Standard Implementation[a]

molecule	CAS	CD energy	STD energy
catechol	6,6	−380.624 556 05	−380.624 560 70
naphthalene	10,10	−383.592 108 03	−383.592 112 05
nicotine	6,6	−495.948 509 96	−495.948 509 88
tryptophan	8,8	−682.501 156 65	−682.501 137 25
pyridoxamine	6,6	−568.806 086 69	−568.806 101 53
2Me4HSdiox	6,6	−855.034 927 82	−855.034 975 73
indole	8,8	−361.673 841 72	−361.673 848 10

[a]Energy values are given in hartrees.

CASSCF energy obtained with both the Cholesky and standard implementations for the selected molecules. As it can be seen from the table, the two results are in agreement to at least the fourth decimal digit, with the largest deviation being about 50 μE_h. We note that it has been documented in the literature that the errors in CD energies are systematic such that CD benefits from error cancellation, thus further increasing the accuracy of energy differences.[62,82]

To test the performances of the CD-CASSCF algorithm, we run calculations on the whole benchmark set that involved active spaces up to CAS(14,14) and as many as 2962 basis functions. For each system, all of the orbitals, including the core ones, are fully variationally optimized. All of the calculations presented here were performed on a single cluster node equipped with 4 Intel Xeon Gold 6140 M CPUs, running at 2.30 GHz. The lower triangular part of the Cholesky vectors were kept in memory. Shared-memory parallelization is exploited in all of the calculations, sharing the work among 28 cores. We point out here that we do not expect the implementation to be fully scalable, the limiting factor being the full CI code. This is due to the fact that the sequential code is highly cache-optimized, which causes an overload of the cache, and consequent loss of efficiency, when more cores of the same processor share cache access. Nevertheless, even a straightforward OpenMP parallelization of the main loops is beneficial. In Table 3 we report, for each molecule, the active space (CAS), the number of basis functions, the number of

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

macroiterations required to converge, and the total CPU wall time in minutes.

Table 3. Benchmarks Set Results[a]

molecule	CAS	N_b	it.	time (min)
adrenaline	6,6	572	4	1.04
anthracene	14,14	560	5	8.27
azulene	10,10	412	5	0.57
biphenyl	12,12	500	4	0.93
catechol	6,6	324	4	0.11
dopamine	6,6	484	4	0.63
fluorene	12,12	530	4	0.95
indole	8,8	368	5	0.31
l-dopamine	6,6	574	4	1.08
naphthalene	10,10	412	4	0.34
niacin	6,6	340	4	0.14
niacinamide	6,6	354	4	0.17
nicotine	6,6	556	4	0.85
nor-adrenaline	6,6	514	4	0.74
picolinic acid	6,6	340	4	0.16
pyridine	6,6	250	4	0.05
pyridoxal	8,8	486	5	0.88
pyridoxamine	6,6	528	5	1.21
pyridoxin	6,6	514	4	0.86
resveratrol	14,14	678	5	11.43
serotonin	8,8	558	4	0.95
tryptophan	8,8	618	5	2.08
2Me2HSdiox	4,4	474	5	0.76
2Me4HSdiox	6,6	446	5	0.61
coumarin dye	12,12	872	5	4.55
chlorophyll	12,12	2962	12	735.01

[a]For each molecule, the active space (CAS), the number of basis functions (N_b), the number of macroiterations (it.), and the total CPU wall time (time) in minutes are presented. 2Me2HSdiox is the abbreviation for 5,7-dimethyl-2*H*,3*H*-thieno[3,4-*b*][1,4]dioxine.

For all of the systems, we used as MO guess the UNO. If more than one instability is found, we compute the UNO from the averaged one-body charge density matrix of the various unrestricted solutions, the only exception being the chlorophyll molecule where we compute the UNO following a single instability (out of three)—the one with the highest negative eigenvalue in absolute value. This choice is in principle suboptimal, but unfortunately, when we performed the computation with a smaller basis set (cc-pVDZ), we found that the optimal active space prescribed by the UNO strategy, a CAS(22,22), was out of the reach of a traditional full CI implementation. To obtain a feasible active space, we reduced the occupation thresholds used to select the active orbitals from 0.01−1.99 to 0.05−1.95, which resulted in a CAS(12,12). We verified that, with this choice, the cc-pVDZ calculation with the UNO obtained following a single instability converged to the same minimum and exhibited the same convergence pattern as the one using the UNO obtained following all of the instabilities.

All of the calculations for which the full active space built using the UNO procedure was computationally affordable, i.e., all but the one on chlorophyll, converged in at most 5 iterations, which demonstrates not only the robustness of the NEO algorithm and the overall efficiency of the implementation but also the remarkable quality of the UNO guess. Unfortunately, the active space suggested by the procedure for chlorophyll was out of reach of a traditional full CI solver. As a consequence, we had to reduce it to a more manageable CAS(12,12). The non-optimality of such a choice is reflected in the larger number of iterations required to converge the wave function. Nevertheless, the overall calculation could be completed in little more than 12 h.

To further illustrate the behavior of the algorithm, we can divide the work into three main tasks—the AO to MO transformation, the optimization of the MOs (MO opt.), and the optimization of the CI coefficients (CI opt.). The MO optimization includes the calculation of the orbital gradient (eq 20), which in turn requires one to assemble the various Fock matrices (eqs 18, 21, and 22), the calculation of the diagonal of the MO Hessian (which is used as the preconditioner in the Davidson diagonalization), and the evaluation of the direct eqs 27 and 28 for each microiterations. On the other hand, the CI optimization consists of computing the reduced density matrices, assembling the CI gradient (eq 23), and evaluating eqs 26 and 25 at each microiteration. Table 4 shows the percentage time to perform these three operations with respect to the total time of the last macroiteration, which is usually the one that requires the largest number of microiterations to converge the NEO problem (microit. in Table 4).

Table 4. Percentage Time of the Three Leading Operations with Respect to the Total Time of the Last Macroiteration (Time)[a]

molecule	AO to MO (%)	MO opt. (%)	CI opt. (%)	microit.	time (s)
adrenaline	8.76	90.63	0.04	20	29.0
biphenyl	5.06	71.07	22.0	12	21.4
naphthalene	7.83	82.41	7.67	13	6.6
tryptophan	5.75	93.54	0.14	24	42.2
anthracene	0.98	14.35	81.65	11	194.9

[a]AO to MO refers to the atomic orbitals to molecular orbitals transformation of the Cholesky vectors, MO opt. is the time spent in the MO optimization and includes operations such as calculation of the orbital gradient, evaluation of the NEO augmented Hessian−orbital trial vector products. CI opt. refers to the CI optimization and include the following operations: calculation of the CI gradient, calculation of the reduced density matrices, and evaluation of the NEO augmented Hessian-configuration trial vector products. microit. is the number of microiterations required to solve the NEO eigenvalue−eigenvector problem.

The CD extremely facilitates the integrals transformation, shifting the bottleneck to the MO optimization part. For the systems considered, in particular, most of the time is spent in computing the transformed Fock matrices, an operation that is required to assemble the NEO Hessian-orbital trial vector product (eqs 27, 28). We also note that, for larger active spaces, such as in biphenyl, the cost associated with the CI part starts to become non-negligible and rapidly becomes the bottleneck as shown for anthracene. Here, the most expensive operations are the direct-CI steps needed to compute the CI gradient and the CI part of the NEO augmented Hessian-configuration trial vector products, together with the assembling of the reduced density matrices. As it can be seen, these operations take about 80% of the total time.

5. CONCLUSIONS

We have presented the implementation of a second-order CASSCF optimization algorithm that exploits the Cholesky

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

decomposition of the two-electron integrals. The algorithm is based on a trust-region method, which requires one to solve diagonally shifted Newton−Raphson equations known as Levenberg−Marquard equations. Also, it adaptively modifies the trust radius during the optimization according to the value of the energy with the result that the overall algorithm always converges to the closest minimum for regular enough functions. The coupling between orbitals and CI coefficients is naturally included in the off-diagonal blocks of the Hessian matrix, making this algorithm naturally second order in all parameters. The implementation is based on the norm-extended optimization formalism, where the LM equations are recast into an eigenvalue problem, where the first eigenvector provides the optimal direction for ground-state minimization problems.

To reduce the computational cost associated with orbital optimization, which is dominating for not-too-large active spaces, we implemented the NEO algorithm using the Cholesky decomposition of the two-electron integrals matrix. The NEO equations were rewritten in terms of the Cholesky vectors, taking particular care in recasting all of the equations in a way that allowed us to implement them efficiently thanks to an extensive use of level 3 BLAS routines. The implementation exploits a fully direct algorithm where the Hessian matrix is never explicitly calculated. Furthermore, since the Cholesky vectors are independent among the others, the code has been parallelized with shared-memory OpenMP directives.

The resulting algorithm was tested on various aromatic systems. We used a triple-ζ basis set with up to 2962 functions and active spaces up to CAS(14,14). All calculations converged swiftly and required limited computer time. Thanks to the effective compression of the two-electron integrals matrix operated by the CD, fully in-core calculations are possible for most systems, eliminating thus the bottleneck of slow disk I/O. While several further improvements and optimizations are possible, for instance, to improve the convergence of the microiterations, the benchmark calculations reported in this contribution show that a rigorous second-order algorithm can be used in large-scale applications at a reasonable computational cost. Future work will focus on both algorithmic improvements and extensions of the methodology. In particular, a first-order procedure such as super CI[83,84] could be used in the preliminary phase of a calculation to achieve an initial intermediate convergence goal, thus providing a very good starting point for the quadratically convergent optimization. We also plan to extend the second-order procedure to the simultaneous optimization of several electronic states and to the calculation of analytical gradients, by implementing differentiated Cholesky vectors.[63,85]

APPENDIX A: COMPUTATIONAL COST AND SCALING OF THE CD-NEO STRATEGY AND COMPARISON WITH DIFFERENT IMPLEMENTATIONS

In this appendix, we discuss in detail the computational cost of the various operations required to perform a NEO-CASSCF calculation using the CD of the two-electron integrals (CD-NEO) and briefly compare it to other CASSCF optimization strategies. We focus the discussion on the cost of orbital-related operations and therefore neglect the cost of the various direct CI steps. The latter become dominant for large active spaces and are of little importance for small ones.

Second-order methods require two-electron integrals in the MO basis with up to two virtual indices. This requires, in general, to perform $O((N_{int} + N_{act})N_b^4)$ floating point operations and is the main bottleneck in CASSCF calculations not dominated by the CI part. In our implementation, the MO transformed integrals are obtained by fully transforming each Cholesky vector to the MO basis, at a cost of $O(N_{ch}N_b^3) \approx O(N_b^4)$. This is formally still the rate-determining step in our algorithm; however, it can be performed very efficiently using level 3 BLAS routines and is trivially parallelized. Note that we never assemble the two-electron integrals explicitly, except for the ones with four active indices that are needed for the CI problem. The evaluation of the orbital gradients requires one to assemble the inactive and active Fock matrices and the $\mathbf{Q}$ matrix. Reporting only the leading terms, this requires one to perform $O(N_b^2N_{int}N_{ch})$, $O(N_b^2N_{act}N_{ch})$, and $O(N_{act}^4N_{ch})$ floating point operations, respectively. Again, all of these operations can be performed using level 2 and level 3 BLAS routines and are easily parallelized. All of the operations described so far need to be performed at each macroiteration. Assembling the NEO step requires the solution to the NEO eigenvalue (eq 12). This is done using an iterative procedure and, in particular, Davidson diagonalization. The calculation of the required matrix-vector products requires one to assemble one-index transformed Fock and $\mathbf{Q}$ matrices, as described in section 3. The leading terms in computational cost for such operations are $O(N_b^2N_{act}N_{ch})$ and $O(N_b^2N_{int}N_{ch})$, and therefore the computation of a matrix−vector product is cheaper than the transformation of the two-electron integrals. Once again, the various operations can be performed using efficient BLAS routines and are easily parallelized.

The memory requirements of the CD-NEO strategy are in principle very limited, with a few $O(N_b^2)$ arrays being required to assemble various intermediates—it should be noted that such intermediates are duplicated for parallel execution, therefore increasing the memory requirements of the run proportionally to the number of cores used for the calculation. However, the overall algorithm achieves significant speed ups if the Cholesky vectors can be held in core, which requires $N_{ch}N_b(N_b + 1)/2$ words of memory. If such an amount of memory is not available, the Cholesky vectors can be read in batches and then distributed among the available cores for computation. This, however, introduces a significant I/O bottleneck that can not only slow the calculation but also reduce the parallelization efficiency, even though modern hardware equipped with solid-state disks mitigates the loss of performance. Nevertheless, calculations on medium- to large-sized molecules can be performed usually by holding the Cholesky vectors in core. The biggest calculation reported in section 4, where about 3000 basis functions were used, required about 600GB of memory, which is a large, yet manageable, amount of RAM. Calculations for smaller systems, with up to 1000 basis functions, can be easily performed in core even on modest hardware.

It is interesting to compare here the computational cost of the CD-NEO strategy with a different second-order algorithm, the Werner−Meyer−Knowles (WMK) algorithm,[25,26] as it is one of the most well-established strategies to solve the

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

CASSCF optimization problem. In particular, a very efficient implementation of the WMK exists that relies on density fitting as a compression technique to handle the two-electron integrals.

CD and DF are intimately related from a theoretical point of view, to the point that the Cholesky decomposition can be considered a specific case of DF, where the auxiliary basis, here called the Cholesky basis, is automatically generated by the decomposition algorithm. In DF, the AO product densities are expanded in terms of optimized basis functions; therefore, the two-electron integrals are approximated as

$$(\mu\nu|\rho\sigma) \simeq \sum_{KJ} (\mu\nu|K)(K|J)^{-1}(J|\rho\sigma) \tag{41}$$

where K and J are elements of the fitting basis and $(K|J)$ is the overlap matrix of such functions, i.e., the metric. Different metrics can be used, where one considers either the straightforward overlap of the two functions or their Coulomb interaction, the latter being the preferred choice as it can be proven that it affords an approximation of the integrals that is correct up to first order.[86] According to eq 41, the Cholesky vectors can be expressed as

$$L_{\mu\nu}^{K} = \sum_{J} (J|\mu\nu)B_{KJ}^{-1} \tag{42}$$

with B_{KJ} being the Cholesky factor of the matrix $(K|J)$. When using an orthonormal Cholesky basis ($B_{KJ} = \delta_{KJ}$), the Cholesky vectors are exactly the fitting coefficients. Therefore, the CD is a DF procedure with a particular auxiliary basis.

From a computational point of view, there are, however, several differences. DF uses optimized, precomputed basis sets and requires thus only the computation of the inverse metric. The latter operation is in principle expensive if direct linear algebra techniques are used, but the inversion can be obtained at a reduced cost by using local approximations.[44,87] On the other hand, Cholesky vectors have to be determined for each system and each geometry. This is in practice not a major source of overhead, as very efficient algorithms have been recently proposed.[51] It is also important to remark that CD allows for easy and efficient parallelization of the code, as all of the operations involving different Cholesky vectors can be performed independently. As a consequence, parallelization is easily achieved by distributing the Cholesky vectors among the available processors, with a final reduction to be performed on the computed quantities. In our implementation, we exploit this feature by implementing all of the various operations with the loop over the Cholesky vector as the most external one. Even a simple-minded shared-memory parallelization with OpenMP directives[88] is already quite effective. Due to the theoretical similarities between CD and DF, the scaling of the various steps needed to perform a DF-WMK calculation are equivalent to the ones observed in a CD-NEO one. In particular, the leading computational step in DF-WMK is again the transformation of the ERI into the MO basis, for which the authors[78] report a cost of $O(MN_b^3)$, M being the number of auxiliary basis functions, which is comparable with the $O(N_{ch}N_b^3)$ one for CD-NEO. The same applies to the calculation of the MO gradient and to the optimization of the WMK model function (i.e., the WMK microiterations), which require one-index transformed quantities similar to the ones reported here. While a thorough comparison of the timings associated with the two implementations would be beyond the scope of this work, preliminary tests performed on the system used as a benchmark by Kreplin et al.[78] show very similar performances. Comparing the results reported in the literature for systems of similar size as the ones reported in section 4, we also observed overall similar timings. We stress here that this is only a qualitative comparison and that a more thorough comparison is needed to precisely assess the relative performances of the two implementations.

As a general remark, because the Cholesky vectors are precomputed, we expect a CD-based strategy to be particularly attractive in a pre-asymptotic regime, i.e., when the sparsity of the two-electron integrals matrix is not too pronounced. This includes calculations on compact molecules but also calculations that use large basis sets, especially if diffuse functions are included. For even larger systems, there are two factors that make a CD-based approach less competitive. First, the Cholesky vectors may not fit in memory, requiring thus slow disk I/O. Second, the sparsity of the two-electron integrals is poorly exploited with CD. On the contrary, DF-based approaches have the advantage of being operable using the conventional integral-direct techniques, which allow one not only to compute efficiently the required fitted integrals, but also to achieve reduced (up to linear) scaling regimes when assembling the Fock and Fock-like matrices thanks to the use of advanced integral evaluation methods, such as the J-engine,[89] continuous fast multipole method,[90] and various screening techniques[44,87,91] that efficiently exploit the sparsity in the integrals. Therefore, for very large systems, DF techniques are expected to achieve more significant speed ups than the CD. It should be stressed, however, that computational efficiency is not the only criterion of interest when comparing two different approaches. While the accuracy of the DF approach depends on the reliability of the fitting basis, which needs to be optimized for both the quantum chemistry model and the basis set, the accuracy of the CD is solely controlled by the decomposition threshold, making this approach virtually ab initio and model independent—we can continuously move from an approximate to a near exact description of the integrals. This feature is, in our opinion, particularly attractive, especially for applications for which the numerical accuracy of the results is fundamental.

It is also interesting to compare second-order strategies to first-order ones, where the MOs and CI coefficients are optimized separately in an alternating gradient spirit. First-order methods require MO integrals with up to one virtual index, and therefore the integral transformation costs $O(N_{act}N_b^4)$, which can be reduced to $O(N_{act}N_b^2N_{ch})$ using CD or, analogously, DF. In the pre-asymptotic regime, this is the leading contribution to the overall cost of such methods for what concerns orbital optimization, making such methods particularly efficient and well-suited for large-scale calculations. Furthermore, first-order methods are well-suited for being implemented in an integral-direct fashion, assembling the various Fock matrices needed to compute the MO gradient in the AO basis. This allows one to fully exploit the sparsity of the ERI matrix in the asymptotic regime, achieving an overall scaling that is between $O(N_{act}N_b)$ and $O(N_{act}N_b^2)$, as reported by Hohenstein et al.[65] in a recent paper. Furthermore, in such an implementation the AO integrals and their contraction with density matrices are performed using GPUs, achieving impressive performances. For very large molecules, especially when compact basis sets are used, this strategy is extremely

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

efficient and bound to overperform any second-order implementation. While a thorough comparison of such an implementation with our CD-NEO strategy is out of the scope of this work, preliminary tests show comparable performances for medium−large systems, that is, as long as the Cholesky vectors fit in memory and the sparsity of the ERI matrix is not too prominent. On the other hand, first-order methods have no guarantee of convergence and can require a very large number of iterations. The robustness of second-order procedures is, in our opinion, a very important feature, especially when one considers the notorious difficulty of converging a CASSCF calculation. Note that, in the pre-asymptotic regime, first-order methods can greatly benefit from the use of CD.[62] In conclusion, the CD-NEO strategy represents a very good compromise between robustness and efficiency for not-too-large systems, as it was demonstrated in section 4.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00327.

Explicit equations for the direct product with the MO Hessian, implementable expressions for the transformed Fock matrices, and optimized molecular geometries for the new set of aromatic molecules used as benchmarks (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Filippo Lipparini − *Dipartimento di Chimica e Chimica Industriale, Università di Pisa, I-56124 Pisa, Italy;* orcid.org/0000-0002-4947-3912; Email: filippo.lipparini@unipi.it

Authors

Tommaso Nottoli − *Dipartimento di Chimica e Chimica Industriale, Università di Pisa, I-56124 Pisa, Italy*

Jürgen Gauss − *Department Chemie, Johannes Gutenberg-Universität Mainz, D-55128 Mainz, Germany;* orcid.org/0000-0002-6432-9345

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00327

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

T.N. acknowledges the traineeship funds from the Erasmus+ program.

■ REFERENCES

(1) Werner, H.-J. Matrix-Formulated Direct Multiconfiguration Self-Consistent Field and Multiconfiguration Reference Configuration-Interaction Methods. *Adv. Chem. Phys.* **2007**, *69*, 1−62.

(2) Shepard, R. The Multiconfiguration Self-Consistent Filed Method. *Adv. Chem. Phys.* **2007**, *69*, 63−200.

(3) Roos, B. O. The Complete Active Space Self-Consistent Field Method and its Applications in Electronic Structure Calculations. *Adv. Chem. Phys.* **2007**, *69*, 399−445.

(4) Knowles, P. J.; Werner, H.-J. An efficient method for the evaluation of coupling coefficients in configuration interaction calculations. *Chem. Phys. Lett.* **1988**, *145*, 514−522.

(5) Werner, H.-J.; Knowles, P. J. An efficient internally contracted multiconfiguration-reference configuration interaction method. *J. Chem. Phys.* **1988**, *89*, 5803−5814.

(6) Jeziorski, B.; Monkhorst, H. J. Coupled-cluster method for multideterminantal reference states. *Phys. Rev. A: At., Mol., Opt. Phys.* **1981**, *24*, 1668−1681.

(7) Banerjee, A.; Simons, J. The coupled-cluster method with a multiconfiguration reference state. *Int. J. Quantum Chem.* **1981**, *19*, 207−216.

(8) Aoto, Y. A.; Köhn, A. Internally contracted multireference coupled-cluster theory in a multistate framework. *J. Chem. Phys.* **2016**, *144*, 074103.

(9) Hanauer, M.; Köhn, A. Pilot applications of internally contracted multireference coupled cluster theory, and how to choose the cluster operator properly. *J. Chem. Phys.* **2011**, *134*, 204111.

(10) Köhn, A.; Hanauer, M.; Mück, L. A.; Jagau, T.-C.; Gauss, J. State-specific multireference coupled-cluster theory. *WIREs Comput. Mol. Sci.* **2013**, *3*, 176−197.

(11) Evangelista, F. A.; Gauss, J. An orbital-invariant internally contracted multireference coupled cluster approach. *J. Chem. Phys.* **2011**, *134*, 114102.

(12) Evangelista, F. A.; Hanauer, M.; Köhn, A.; Gauss, J. A sequential transformation approach to the internally contracted multireference coupled cluster method. *J. Chem. Phys.* **2012**, *136*, 204108.

(13) Lipparini, F.; Kirsch, T.; Köhn, A.; Gauss, J. Internally Contracted Multireference Coupled Cluster Calculations with a Spin-Free Dirac−Coulomb Hamiltonian: Application to the Monoxides of Titanium, Zirconium, and Hafnium. *J. Chem. Theory Comput.* **2017**, *13*, 3171−3184.

(14) Andersson, K.; Malmqvist, P. Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-Order Perturbation Theory with a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94*, 5483−5488.

(15) Andersson, K.; Malmqvist, P. Å.; Roos, B. O. Second-order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **1992**, *96*, 1218−1226.

(16) Angeli, C.; Cimiraglia, R.; Evangelisti, S.; Leininger, T.; Malrieu, J. P. Introduction of n-electron valence states for multi-reference perturbation theory. *J. Chem. Phys.* **2001**, *114*, 10252.

(17) Angeli, C.; Cimiraglia, R.; Malrieu, J. P. N-electron valence state perturbation theory: A fast implementation of the strongly contracted variant. *Chem. Phys. Lett.* **2001**, *350*, 297−305.

(18) Angeli, C.; Cimiraglia, R.; Malrieu, J. P. n-electron valence state perturbation theory: A spinless formulation and an efficient implementation of the strongly contracted and of the partially contracted variants. *J. Chem. Phys.* **2002**, *117*, 9138−9153.

(19) Austin, B. M.; Zubarev, D. Y.; Lester, W. A. Quantum Monte Carlo and Related Approaches. *Chem. Rev.* **2012**, *112*, 263−288.

(20) Fracchia, F.; Filippi, C.; Amovilli, C. Size-Extensive Wave Functions for Quantum Monte Carlo: A Linear Scaling Generalized Valence Bond Approach. *J. Chem. Theory Comput.* **2012**, *8*, 1943−1951.

(21) Stein, C. J.; Reiher, M. Automated Selection of Active Orbital Spaces. *J. Chem. Theory Comput.* **2016**, *12*, 1760−1771.

(22) Sayfutyarova, E. R.; Sun, Q.; Chan, G. K. L.; Knizia, G. Automated Construction of Molecular Active Spaces from Atomic Valence Orbitals. *J. Chem. Theory Comput.* **2017**, *13*, 4063−4078.

(23) Tóth, Z.; Pulay, P. Comparison of Methods for Active Orbital Selection in Multiconfigurational Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 7328−7341.

(24) Chan, G. K.-L.; Sharma, S. The density matrix renormalization group in quantum chemistry. *Annu. Rev. Phys. Chem.* **2011**, *62*, 465−481.

(25) Werner, H.-J.; Meyer, W. A quadratically convergent multi-configuration-self-consistent field method with simultaneous optimization of orbitals and Cl coefficients. *J. Chem. Phys.* **1980**, *73*, 2342−2356.

(26) Werner, H.-J.; Knowles, P. J. A second order multiconfiguration SCF procedure with optimum convergence. *J. Chem. Phys.* **1985**, *82*, 5053−5063.
(27) Kreplin, D. A.; Knowles, P. J.; Werner, H.-J. Second-order MCSCF optimization revisited. I. Improved algorithms for fast and robust second-order CASSCF convergence. *J. Chem. Phys.* **2019**, *150*, 194106.
(28) Sun, Q.; Yang, J.; Chan, G. K.-L. A general second order complete active space self-consistent-field solver for large-scale systems. *Chem. Phys. Lett.* **2017**, *683*, 291−299.
(29) Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; Wiley: New York, 1999; Chapter 5.2, pp 100−107.
(30) Jensen, H. J. Aa.; Jørgensen, P. A direct approach to second-order MCSCF calculations using a norm extended optimization scheme. *J. Chem. Phys.* **1984**, *80*, 1204−1214.
(31) Jensen, H. J. Aa.; Ågren, H. A. Direct, restricted-step, second-order MC SCF program for large scale ab initio calculations. *Chem. Phys.* **1986**, *104*, 229−250.
(32) Stanton, J. F.; Gauss, J.; Cheng, L.; Harding, M. E.; Matthews, D. A.; Szalay, P. G. *CFOUR, Coupled-Cluster techniques for Computational Chemistry, a quantum-chemical program package*. With contributions from Auer, A. A.; Bartlett, R. J.; Benedikt, U.; Berger, C.; Bernholdt, D. E.; Blaschke, S.; Bomble, Y. J.; Burger, S.; Christiansen, O.; Datta, D.; Engel, F.; Faber, R.; Greiner, J.; Heckert, M.; Heun, O.; Hilgenberg, M.; Huber, C.; Jagau, T.-C.; Jonsson, D.; Jusélius, J.; Kirsch, T.; Klein, K.; Kopper, G. M.; Lauderdale, W. J.; Lipparini, F.; Metzroth, T.; Mück, L. A.; O'Neill, D. P.; Nottoli, T.; Price, D. R.; Prochnow, E.; Puzzarini, C.; Ruud, K.; Schiffmann, F.; Schwalbach, W.; Simmons, C.; Stopkowicz, S.; Tajti, A.; Vázquez, J.; Wang, F.; Watts, J. D. and the integral packages MOLECULE (Almlöf, J.; Taylor, P. R.), PROPS (Taylor, P. R.), ABACUS (Helgaker, T.; Jensen, H. J. Aa.; Jørgensen, P.; Olsen, J., and ECP routines (Mitin, A. V.; van Wüllen, C.. For the current version, see http://www.cfour.de.
(33) Matthews, D. A.; Cheng, L.; Harding, M. E.; Lipparini, F.; Stopkowicz, S.; Jagau, T.-C.; Szalay, P. G.; Gauss, J.; Stanton, J. F. Coupled-cluster techniques for computational chemistry: The CFOUR program package. *J. Chem. Phys.* **2020**, *152*, 214108.
(34) Lawson, C. L.; Hanson, R. J.; Kincaid, D. R.; Krogh, F. T. Basic Linear Algebra Subprograms for Fortran Usage. *ACM Trans. Math. Software* **1979**, *5*, 308−323.
(35) Dongarra, J. J.; Du Croz, J.; Hammarling, S.; Duff, I. S. A Set of Level 3 Basic Linear Algebra Subprograms. *ACM Trans. Math. Software* **1990**, *16*, 1−17.
(36) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496−4501.
(37) Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. On some approximations in applications of X_α theory. *J. Chem. Phys.* **1979**, *71*, 3396−3402.
(38) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. Integral approximations for LCAO-SCF calculations. *Chem. Phys. Lett.* **1993**, *213*, 514−518.
(39) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359−363.
(40) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.* **1995**, *240*, 283−290.
(41) Weigend, F.; Häser, M. RI-MP2: First derivatives and global consistency. *Theor. Chem. Acc.* **1997**, *97*, 331−340.
(42) Weigend, F. A fully direct RI-HF algorithm: Implementation, optimized auxiliary basis sets, demonstration of accuracy and efficiency. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285−4291.
(43) Sierka, M.; Hogekamp, A.; Ahlrichs, R. Fast evaluation of the Coulomb potential for electron densities using multipole accelerated resolution of identity approximation. *J. Chem. Phys.* **2003**, *118*, 9136−9148.
(44) Sodt, A.; Subotnik, J. E.; Head-Gordon, M. Linear scaling density fitting. *J. Chem. Phys.* **2006**, *125*, 194109.
(45) Beebe, N. H. F.; Linderberg, J. Simplifications in the Generation and Transformation of Two-Electron Integrals in Molecular Calculations. *Int. J. Quantum Chem.* **1977**, *12*, 683−705.
(46) Røeggen, I.; Wisløff-Nilssen, E. On the Beebe-Linderberg two-electron integral approximation. *Chem. Phys. Lett.* **1986**, *132*, 154−160.
(47) Koch, H.; Sánchez de Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481−9484.
(48) Røeggen, I.; Johansen, T. Cholesky decomposition of the two-electron integral matrix in electronic structure calculations. *J. Chem. Phys.* **2008**, *128*, 194107.
(49) Aquilante, F.; Boman, L.; Boström, J.; Koch, H.; Lindh, R.; de Merás, A. S.; Pedersen, T. B. In *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Dordrecht, The Netherlands, 2011; pp 301−343. DOI: 10.1007/978-90-481-2853-2.
(50) Weigend, F.; Kattannek, M.; Ahlrichs, R. Approximated electron repulsion integrals: Cholesky decomposition versus resolution of the identity methods. *J. Chem. Phys.* **2009**, *130*, 164106.
(51) Folkestad, S. D.; Kjønstad, E. F.; Koch, H. An efficient algorithm for Cholesky decomposition of electron repulsion integrals. *J. Chem. Phys.* **2019**, *150*, 194112.
(52) Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J. Chem. Phys.* **2003**, *118*, 8149−8160.
(53) Manby, F. R. Density fitting in second-order linear-r12 Møller-Plesset perturbation theory. *J. Chem. Phys.* **2003**, *119*, 4607−4613.
(54) Polly, R.; Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast Hartree-Fock theory using local density fitting approximations. *Mol. Phys.* **2004**, *102*, 2311−2321.
(55) Boström, J.; Aquilante, F.; Pedersen, T. B.; Lindh, R. Analytical gradients of hartree-fock exchange with density fitting approximations. *J. Chem. Theory Comput.* **2013**, *9*, 204−212.
(56) Bozkaya, U. Derivation of general analytic gradient expressions for density-fitted post-Hartree-Fock methods: An efficient implementation for the density-fitted second-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2014**, *141*, 124108.
(57) Bozkaya, U. Orbital-optimized second-order perturbation theory with density-fitting and cholesky decomposition approximations: An efficient implementation. *J. Chem. Theory Comput.* **2014**, *10*, 2371−2378.
(58) Bozkaya, U.; Sherrill, C. D. Analytic energy gradients for the coupled-cluster singles and doubles method with the density-fitting approximation. *J. Chem. Phys.* **2016**, *144*, 174103.
(59) Bozkaya, U. Orbital-optimized linearized coupled-cluster doubles with density-fitting and Cholesky decomposition approximations: An efficient implementation. *Phys. Chem. Chem. Phys.* **2016**, *18*, 11362−11373.
(60) Bozkaya, U.; Sherrill, C. D. Analytic energy gradients for the coupled-cluster singles and doubles with perturbative triples method with the density-fitting approximation. *J. Chem. Phys.* **2017**, *147*, 044104.
(61) Burger, S.; Lipparini, F.; Gauss, J.; Stopkowicz, S. NMR chemical shift computations at second-order Møller-Plesset perturbation theory using gauge-including atomic orbitals and Cholesky-decomposed two-electron integrals. *J. Chem. Phys.* **2021**, *155*, 074105.
(62) Aquilante, F.; Pedersen, T. B.; Lindh, R.; Roos, B. O.; Sánchez de Merás, A.; Koch, H. Accurate ab initio density fitting for multiconfigurational self-consistent field methods. *J. Chem. Phys.* **2008**, *129*, 024113.
(63) Delcey, M. G.; Freitag, L.; Pedersen, T. B.; Aquilante, F.; Lindh, R.; González, L. Analytical gradients of complete active space self-consistent field energies using Cholesky decomposition: Geometry optimization and spin-state energetics of a ruthenium nitrosyl complex. *J. Chem. Phys.* **2014**, *140*, 174103.

(64) Reynolds, R. D.; Yanai, T.; Shiozaki, T. Large-scale relativistic complete active space self-consistent field with robust convergence. *J. Chem. Phys.* **2018**, *149*, 014106.
(65) Hohenstein, E. G.; Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. An atomic orbital-based formulation of the complete active space self-consistent field method on graphical processing units. *J. Chem. Phys.* **2015**, *142*, 224103.
(66) Snyder, J. W.; Hohenstein, E. G.; Luehr, N.; Martínez, T. J. An atomic orbital-based formulation of analytical gradients and non-adiabatic coupling vector elements for the state-averaged complete active space self-consistent field method on graphical processing units. *J. Chem. Phys.* **2015**, *143*, 154107.
(67) Lipparini, F.; Gauss, J. Cost-effective treatment of scalar relativistic effects for multireference systems: A CASSCF implementation based on the spin-free Dirac-Coulomb Hamiltonian. *J. Chem. Theory Comput.* **2016**, *12*, 4284−4295.
(68) Jensen, H. J. Aa.; Jørgensen, P.; Ågren, H. Efficient optimization of large scale MCSCF wave functions with a restricted step algorithm. *J. Chem. Phys.* **1987**, *87*, 451−466.
(69) Gauss, J.; Lipparini, F.; Burger, S.; Blaschke, S.; Kitsaras, M.-P.; Stopkowicz, S. *Unpublished work*; Johannes Gutenberg-Universität Mainz.
(70) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self — Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian — Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.
(71) Pulay, P.; Hamilton, T. P. UHF natural orbitals for defining and starting MC-SCF calculations. *J. Chem. Phys.* **1988**, *88*, 4926−4933.
(72) Grimme, S.; Hansen, A. A Practicable Real-Space Measure and Visualization of Static Electron-Correlation Effects. *Angew. Chem., Int. Ed.* **2015**, *54*, 12308−12313.
(73) Stein, C. J.; Reiher, M. Automated identification of relevant frontier orbitals for chemical compounds and processes. *Chimia* **2017**, *71*, 170−176.
(74) Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XVIII. Constraints and stability in Hartree-Fock theory. *J. Chem. Phys.* **1977**, *66*, 3045−3050.
(75) Knizia, G. Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts. *J. Chem. Theory Comput.* **2013**, *9*, 4834−4843.
(76) Knizia, G.; Klein, J. E. Electron flow in reaction mechanisms—revealed from first principles. *Angew. Chem., Int. Ed.* **2015**, *54*, 5518−5522.
(77) Menezes, F.; Kats, D.; Werner, H.-J. Local complete active space second-order perturbation theory using pair natural orbitals (PNO-CASPT2). *J. Chem. Phys.* **2016**, *145*, 124115.
(78) Kreplin, D. A.; Knowles, P. J.; Werner, H.-J. MCSCF optimization revisited. II. Combined first- and second-order orbital optimization for large molecules. *J. Chem. Phys.* **2020**, *152*, 074102.
(79) Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(80) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, Revision A.03; Gaussian: Wallingford, CT, USA, 2016.
(81) Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(82) Aquilante, F.; Lindh, R.; Bondo Pedersen, T. Unbiased auxiliary basis sets for accurate two-electron integral approximations. *J. Chem. Phys.* **2007**, *127*, 114107.
(83) Roos, B. O. The Complete Active Space SCF Method in a Fock-Matrix-Based Super-CI Formulation. *Int. J. Quantum Chem.* **1980**, *18*, 175−189.
(84) Siegbahn, P. E.; Almlöf, J.; Heiberg, A.; Roos, B. O. The complete active space SCF (CASSCF) method in a Newton-Raphson formulation with application to the HNO molecule. *J. Chem. Phys.* **1981**, *74*, 2384−2396.
(85) Feng, X.; Epifanovsky, E.; Gauss, J.; Krylov, A. I. Implementation of analytic gradients for CCSD and EOM-CCSD using Cholesky decomposition of the electron-repulsion integrals and their derivatives: Theory and benchmarks. *J. Chem. Phys.* **2019**, *151*, 014110.
(86) Dunlap, B. Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry. *J. Mol. Struct.: THEOCHEM* **2000**, *529*, 37−40.
(87) Reine, S.; Tellgren, E.; Krapp, A.; Kjærgaard, T.; Helgaker, T.; Jansik, B.; Høst, S.; Salek, P. Variational and robust density fitting of four-center two-electron integrals in local metrics. *J. Chem. Phys.* **2008**, *129*, 104101.
(88) OpenMP Architecture Review Board. *OpenMP Application Program Interface*, Version 4.5, 2015; https://www.openmp.org/wp-content/uploads/OpenMP-4.5-1115-F-web.pdf (accessed 2021-03-30).
(89) White, C. A.; Head-Gordon, M. A J matrix engine for density functional theory calculations. *J. Chem. Phys.* **1996**, *104*, 2620−2629.
(90) White, C. A.; Johnson, B. G.; Gill, P. M.; Head-Gordon, M. The continuous fast multipole method. *Chem. Phys. Lett.* **1994**, *230*, 8−16.
(91) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices. *J. Chem. Phys.* **1998**, *109*, 1663−1669.

https://doi.org/10.1021/acs.jctc.1c00327
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX