CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/130375/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Kreplin, David A., Knowles, Peter J. and Werner, Hans-Joachim 2020. MCSCF optimization revisited. II. Combined first- and second-order orbital optimization for large molecules. The Journal of Chemical Physics 152 (7) , 074102. 10.1063/1.5142241

Publishers page: http://dx.doi.org/10.1063/1.5142241

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

MCSCF optimization revisited. II. Combined first- and second-order orbital optimization for large molecules

David A. Kreplin,[1, a)] Peter J. Knowles,[2, b)] and Hans-Joachim Werner[1, c)]

[1)] *Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55,D-70569 Stuttgart, Germany*

[2)] *School of Chemistry, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, United Kingdom*

A new orbital optimization for the multiconfiguration self-consistent field (MCSCF) method is presented. This method combines a second-order (SO) algorithm for the optimization of the active orbitals with the first-order Super-CI (SCI) optimization of the remaining closed-virtual rotations and is denoted as SO-SCI method. The SO-SCI method improves the convergence significantly as compared to the conventional SCI method. In combination with density fitting, the intermediates from the gradient calculation can be reused to evaluate the two-electron integrals required for the active Hessian without introducing a large computational overhead. The orbitals and configuration interaction (CI) coefficients are optimized alternately, but the CI-orbital coupling is accounted for by the Limited Memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) quasi-Newton method. This further improves the speed of convergence. The method is applicable to large molecules. The efficiency and robustness of the presented method is demonstrated in benchmark calculations for 21 aromatic molecules as well as for various transition metal complexes with up to 826 electrons and 5154 basis functions.

I. INTRODUCTION

The multiconfiguration self-consistent field (MCSCF) method[1–45] plays an important role in the treatment of strongly correlated systems, in which the wavefunction is dominated by more than one electronic configuration. For example, a multiconfigurational treatment is needed even at a qualitative level to describe bond dissociation processes, excited states, or transition metal complexes with partly filled d-shells. Higher accuracy calculations are possible by adding dynamical correlation effects subsequently to the MCSCF calculation. This can be achieved through multireference perturbation theory (MRPT),[46–53] multireference configuration interaction (MRCI),[54–62] or multireference coupled cluster (MRCC).[63–66]

In the MCSCF method the molecular orbitals (MOs) and the configuration interaction (CI) coefficients of the electronic configurations are variationally optimized. The molecular orbitals can be classified into three subspaces: the inactive orbitals, which are doubly occupied in all configurations, the active orbitals with varying occupations, and the unoccupied virtual orbitals. Excited states can be most easily treated by a state-averaged approach, where all states share the same set of orbitals and the energy average of the considered states is optimized.[5,38,67]

Today, most MCSCF calculations are performed with a complete active space (CASSCF) approach, where the wavefunction is expanded in all possible configuration state functions (CSFs) within a preselected set of active orbitals.[6,68–72] In the following, CASSCF calculations with N_{el} electrons in N_{act} active orbitals are denoted CASSCF(N_{el}, N_{act}). However, the number of CSFs increases factorially with the number of active orbitals, and the largest calculations so far included 22 electrons in 22 active orbitals.[73] One way of tackling the exploding number of configuration is to divide the active space into subsets and to add occupation restrictions to each subspace, as for example in the restricted active space (RAS)[9,74] or the generalized active space (GAS)[75,76] methods. In addition to this, several approximate full CI (FCI) methods have recently been developed and integrated into CASSCF, as for example the FCI quantum Monte Carlo (FCIQMC) method,[77–79] the heat-bath CI,[80–82] or the density matrix renormalization group (DMRG) methods.[45,83–87] However, the approximations in the CI space remove the invariance of the MCSCF energy with respect to active-active orbital rotations. This can make the orbital optimization much more difficult, unless the active-active orbital rotations are neglected.[45]

The optimization of MCSCF wavefunctions can be extremely challenging due to strong couplings between the molecular orbitals and the CI coefficients. As a consequence, many MCSCF optimization methods have been developed in the past fifty years. An overview of the different approaches can be found in Ref. 88. The variety of MCSCF methods can be separated into two categories: first-order[1–16] and second-order methods.[17–45] In the latter, the energy changes and gradients decay quadratically near the final solution. In order to achieve quadratic

[a)] Electronic mail: kreplin@theochem.uni-stuttgart.de
[b)] Electronic mail: knowlespj@cardiff.ac.uk
[c)] Electronic mail: werner@theochem.uni-stuttgart.de

convergence it is necessary to treat the coupling between the CI coefficients and the orbitals explicitly. Methods which include this coupling in each iteration are denoted "coupled" or "one-step" optimizations, while methods in which the CI coefficients and orbitals are optimized alternately are denoted "uncoupled" or "two-step" optimizations.

In second-order optimization schemes the energy is in each macro-iteration approximated by a functional which is accurate to second-order in the orbital rotation parameters and the changes of the CI coefficients. This functional is iteratively minimized with respect to the orbitals and CI coefficients (micro-iterations), yielding near the final solution quadratic convergence of the energy in subsequent macro-iterations. Despite fast and robust convergence, second-order methods suffer from a serious deficiency: the construction of the exact orbital Hessian matrix requires the computation of all 2-electron repulsion integrals in the MO basis with up to two virtual orbitals. This computation scales formally as $O(N^5)$, where N is a measure for the molecular system size and the active space is assumed to be constant. The expensive integral evaluation limits the application of second-order MCSCF methods to medium sized molecules. The scaling can be formally reduced to $O(N^4)$ by avoiding the explicit construction of the Hessian and using density fitting approximations in the action of the orbital Hessian on a trial vector in each micro-iteration.[14] However, numerous density fitting integral evaluations are then required in each micro-iteration, leading to a large prefactor in the cost scaling. In first-order methods the orbitals and CI coefficients are optimized alternately: for a given set of orbitals and integrals, one first solves the CI eigenvalue equation. Subsequently the density matrices are computed and the orbitals are optimized for fixed CI coefficients (two-step or uncoupled optimization). The orbital optimization can be done by either a second-order method or by a more approximate approach. Most successful is the so-called Super-CI (SCI) method, which is based on the generalized Brillouin theorem.[89,90] It dates back to Grein and Chang[1–3] and Ruedenberg et al.[5] Originally, a single-excitation CI calculation was carried out in each MCSCF iteration, and the coefficients of the single excitations were used to update the orbitals. However, using the exact Hamiltonian in this method is even more expensive than using the full orbital Hessian, since it requires the same integrals as the latter and in addition the third-order reduced density matrix. The SCI method was strongly improved by Roos et al.[6,7,9] by using a perturbative treatment of the single excitations, based on a similar zeroth order Hamiltonian as used in complete active space second-order perturbation theory (CASPT2). Alternatively, it is also possible to use the Dyall Hamiltonian in the perturbative treatment.[11,15] Both variants use an approximate orbital Hessian, which can be built solely from closed-shell and active space Fock matrices. Evaluating these Fock matrices scales as $O(N^4)$, and their computation can be considerably accelerated with density fitting.[91,92] The scaling can be further reduced by integral screening techniques and/or local approximations.[93,94] For the CI part one also needs the 2-electron integrals in the active space, but these can be generated with very little extra cost together with the active space Fock matrix, if density fitting is used. Thus, the overall scaling is formally $O(N^4)$ (with a fixed active space), and therefore first-order methods can be applied to much larger molecules than second-order methods.

However, the convergence of the two-step optimizations can be problematic, since the decay of the gradient and energy may become extremely slow if there are strong couplings between the orbital and CI optimizations. The convergence can be significantly improved by introducing a convergence accelerator scheme, such as the direct inversion of the iterative subspace (DIIS) method[15,86] or a quasi-Newton (QN) approach.[9,42] Nevertheless, more than 100 iterations may be needed in unfavorable cases.

In the current paper we present a new algorithm which combines the advantages of first-order and second-order orbital optimization algorithms. All active orbitals are optimized using a second-order algorithm, which significantly improves the speed and robustness of convergence. This is combined with the SCI approach for the remaining inactive-virtual orbital rotations. The necessary integrals for the active space optimization can be efficiently computed using density fitting along with the Fock matrices without much additional cost. Furthermore, we found an improved way to accelerate the two-step optimization by using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) QN method. The algorithm utilizes the two-loop recursion scheme of the limited memory BFGS (L-BFGS) method,[95,96] which allows the initial Hessian to be updated in each iteration. The resulting method accounts approximately for the coupling between the CI-coefficients and the orbitals and considerably accelerates the two-step optimization independent of the orbital optimization.

The paper is structured as follows: we start with a brief review of the first- and second-order orbital optimizations in Sec. II, followed by the derivation of the combined first- and second-order orbital optimizations in Sec. II C. The L-BFGS convergence acceleration is discussed in Sec. III. All presented methods are compared in benchmark calculations in Sec. V.

II. ORBITAL OPTIMIZATIONS

We consider a normalized MCSCF wavefunction $|\Psi_n\rangle$ for an electronic state n

$$|\Psi_n\rangle = \sum_I |\Phi_I\rangle\, c_I^n, \qquad \langle\Psi_n|\Psi_n\rangle = \sum_I |c_I^n|^2 = 1, \quad (1)$$

where $|\Phi_I\rangle$ are orthonormal N-electron expansion functions, either spin-adapted configuration state functions

(CSFs) or Slater determinants, and c_I^n are the corresponding configuration interaction (CI) coefficients. The CSFs are built from a set of (spatial) orbitals $|r\rangle \equiv |\phi_r\rangle$, which are assumed to be real and orthonormal. For simplicity, we will in the following omit the state index n, and denote the initial wavefunction and energy of an iteration as $|0\rangle$ and E_0, respectively. The treatment of excited states using state-averaged MCSCF is discussed in section II E.

Unless otherwise noted, the MOs are labeled in following manner: k, l: any occupied orbitals, i, j: doubly occupied closed-shell orbitals, t, u, v, w: active orbitals, a, b: unoccupied (virtual) orbitals. We will assume that all occupied orbitals are optimized, even though freezing core orbitals is possible. The indices p, q, r, s refer to any orbitals.

Arbitrary changes of the molecular orbitals (subject to the orthonormality condition) can be described by an orthogonal transformation

$$|\tilde{r}\rangle = \sum_s |s\rangle\, U_{s\tilde{r}} \tag{2}$$

with

$$\mathbf{U} = \exp(\mathbf{R}) = 1 + \mathbf{R} + \frac{1}{2!}\mathbf{R}^2 + \dots\ , \tag{3}$$

where $R_{rk} = -R_{kr}$ are independent orbital rotation parameters. Virtual-virtual and inactive-inactive orbital rotations leave the energy invariant and are excluded. For CASSCF active-active rotations are also excluded since they can be treated by changes of the CI coefficients. Thus, the parameters R_{ti} and R_{ak} are sufficient for CASSCF wavefunctions, and all others are set to zero.

In the methods described in this paper, the orbitals and CI coefficients are optimized alternately (two-step optimizations). The coupling between the two steps is included approximately using QN approaches. In the following, we describe the orbital optimization methods for fixed CI coefficients.

A. Second-order orbital optimization

The change of the wavefunction by the orbital transformation in (3) can be expressed as

$$|\Psi\rangle = \exp(\hat{R})\,|0\rangle\,, \tag{4}$$

where the operator $\hat{R} = -\hat{R}^\dagger$ is defined as

$$\hat{R} = \sum_{r>k} R_{rk}(\hat{E}_{rk} - \hat{E}_{kr}). \tag{5}$$

The energy expectation value of the transformed wavefunction

$$E(\mathbf{R}) = \langle\Psi|\,\hat{H}\,|\Psi\rangle = \langle 0|\exp(-\hat{R})\hat{H}\exp(\hat{R})\,|0\rangle \tag{6}$$

can be expanded with help of the Baker-Campbell-Hausdorff (BCH) series

$$e^{-\hat{R}}\hat{H}e^{\hat{R}} = \hat{H} + [\hat{H},\hat{R}] + \frac{1}{2!}[[\hat{H},\hat{R}],\hat{R}] + \dots\ , \tag{7}$$

yielding a series of expectation values of commutators. By evaluating these expectation values one obtains a Taylor expansion in the rotation parameters R_{rk} (with $\mathbf{x}$ being a vector containing the non-redundant parameters R_{rk}):

$$\begin{aligned} E(\mathbf{R}) =& \langle 0|\,\hat{H}\,|0\rangle + \langle 0|\,[\hat{H},\hat{R}]\,|0\rangle + \frac{1}{2!}\langle 0|\,[[\hat{H},\hat{R}],\hat{R}]\,|0\rangle + \dots \\ =& E_0 + \mathbf{g}^\mathsf{T}\mathbf{x} + \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{h}\mathbf{x} + \dots\ . \end{aligned} \tag{8}$$

The single commutator term yields the scalar product of the orbital gradient with the parameters $\mathbf{x}$, while the double commutator term depends on the Hessian matrix $\mathbf{h}$. The explicit expressions for the energy, the gradient and the Hessian are shown in the Supplementary Material (SM). The same expressions for g_{rk} and $h_{rk,sl}$ can be obtained by direct expansion of the energy up to second-order in $\mathbf{R}$ and factorizing the resulting expression appropriately, as described in part I of this series.[42]

The parameters $\mathbf{x}$ can be determined iteratively by truncating the expansion (8) after the quadratic term and minimizing in each iteration the resulting second-order energy expression. This yields the linear Newton-Raphson (NR) equations

$$\mathbf{g} + \mathbf{h}\mathbf{x} = 0. \tag{9}$$

However, despite quadratic convergence near the final solution (for fixed CI coefficients), the radius of convergence of the NR method is quite small. A more robust numerical framework is the augmented Hessian (AH) method.[97] It is obtained by introducing a level-shift ϵ to the NR equations:

$$\mathbf{g} + (\mathbf{h} - \epsilon\mathbf{1})\mathbf{x} = 0 \quad \text{with} \quad \epsilon = \lambda^2\mathbf{x}^\mathsf{T}\mathbf{g}. \tag{10}$$

The step-length of $\mathbf{x}$ can be controlled by the damping parameter λ. This equation can be transformed into an eigenvalue problem:

$$\begin{pmatrix} 0 & \mathbf{g}^\mathsf{T} \\ \mathbf{g} & \mathbf{h}/\lambda \end{pmatrix} \begin{pmatrix} 1/\lambda \\ \mathbf{x} \end{pmatrix} = \nu \begin{pmatrix} 1/\lambda \\ \mathbf{x} \end{pmatrix}. \tag{11}$$

The diagonalization of the AH matrix yields the update step $\mathbf{x}$ and the level-shift $\epsilon = \lambda \cdot \nu$. It can be shown that the level-shifted Hessian is positive (semi) definite for the calculated ϵ. Due to the large dimension of the Hessian, the AH eigenvalue equation has to be solved iteratively, for example with a P-space Davidson method.[39,42] The damping parameter λ can be automatically adjusted by restricting the step length $|\mathbf{x}|$, and in this way convergence can be guaranteed. However, with poor starting guesses convergence can be rather slow. In the following,

we will denote the uncoupled second-order AH approach without further approximation as UC-AH method.

It is in principle also possible to include the orbital-CI coupling in the AH procedure, yielding a second-order method with quadratic convergence. However this is expensive, in particular for CASSCF calculations with large active spaces, and not further considered here. The better alternative is the method of Werner, Meyer and Knowles[39–41] (denoted WMK method), which includes higher order terms and converges much faster. A review and recent improvements of this method have been presented in part I of this series.[42]

B. The Super-CI optimization method

There have been many first-order orbital optimization methods developed in the past decades.[7–10,16] In this work, we follow the derivation of the Super-CI (SCI) method of Roos et al.[7–9] because of its wide range of successful applications and popularity. The SCI method is based on the generalized Brillouin theorem,[89,90] which states that for optimized orbitals

$$g_{rk} = 2\,\langle 0|\hat{H}|rk\rangle = 0. \tag{12}$$

Here $|rk\rangle$ are the so-called Brillouin states (internally contracted singly excited configurations)

$$|rk\rangle = (\hat{E}_{rk} - \hat{E}_{kr})\,|0\rangle\,. \tag{13}$$

For non-optimal orbitals, the matrix element $\langle 0|\hat{H}|rk\rangle$ corresponds to half of the orbital gradient g_{rk}.

The SCI wavefunction $|\mathrm{SCI}\rangle$ is defined as

$$|\mathrm{SCI}\rangle = (1+\hat{R})\,|0\rangle = |0\rangle + \sum_{r>k} R_{rk}\,|rk\rangle\,, \tag{14}$$

where the coefficients R_{rk} are equivalent to the rotation generators $\mathbf{R}$. This expansion is equivalent to a first-order approximation of the transformation shown in (4). The coefficients R_{rk} are determined by solving the generalized eigenvalue equations (summation over indices s, l is implied)

$$\begin{pmatrix} -\epsilon & \langle 0|\,\hat{H}\,|sl\rangle \\ \langle rk|\,\hat{H}\,|0\rangle & \langle rk|\,\hat{H} - E_0 - \epsilon\,|sl\rangle \end{pmatrix} \begin{pmatrix} 1 \\ R_{sl} \end{pmatrix} = 0. \tag{15}$$

Since the Brillouin states are non-orthogonal, the overlap matrix $\langle rk|sl\rangle$ has to be included. For CASSCF wavefunctions the overlap matrix becomes diagonal if natural active orbitals are used, i.e. if the 1-RDM is diagonal (c.f. SM).I don't think these are shown. David, please add overlap to the SM.

Computing the Hamiltonian elements $\langle rk|\hat{H}|sl\rangle$ exactly would require three particle RDMs and the same integrals as second-order optimization methods. In order to avoid their expensive computation and to achieve lower-order scaling, the terms $\langle rk|\,\hat{H} - E_0\,|sl\rangle$ in eq. (15) are replaced by $\langle rk|\hat{H}^{eff} - E^{(0)}|sl\rangle$,[7,9] where

$$\hat{H}^{eff} = \sum_{pq} F_{pq}\hat{E}_{pq}, \tag{16}$$

$$E^{(0)} = \langle 0|\hat{H}^{eff}|0\rangle = 2\sum_{i} F_{ii} + \sum_{tu} D_{tu}F_{tu}. \tag{17}$$

The definition of the effective Fock matrix F_{rs} is given in eq. (13) of the SM. Essentially, this corresponds to a perturbational treatment of the single excitations, similar to CASPT2. The Hylleraas functional for the SCI first-order wavefunction reads (summation over repeated indices implied):

$$\epsilon = 2R_{rk}\,\langle rk|\hat{H}|0\rangle + R_{rk}\,\langle rk|\hat{H}^{eff} - E^{(0)}|sl\rangle\,R_{sl}. \tag{18}$$

Its minimization with respect to the parameters R_{ri} leads to a first-order approximation of the parameters R_{rk} and the second-order energy ϵ, namely

$$\epsilon = R_{rk}\,\langle rk|\hat{H}|0\rangle\,, \tag{19}$$

$$0 = \langle rk|\hat{H}|0\rangle + \langle rk|\hat{H}^{eff} - E^{(0)}|sl\rangle\,R_{sl}. \tag{20}$$

This differs from the SCI eigenvalue equation only by the absence of the shift $-\epsilon\,\langle rk|sl\rangle$ in the second term of eq.(20). In praxis, this shift stabilizes convergence and is therefore included in the SCI method. The SCI matrix elements are given explicitly in the SM.

Each iteration of the SCI method starts with the computation of the closed-shell Fock matrix and the integrals $(tu|vw)$, which are needed to solve the CI eigenvalue problem. In our program the default is to carry out as many CI iterations as needed to reduce the CI gradient by a factor of 10 (typically 3-5 iterations). Subsequently, the RDMs, the active part of the Fock matrix, as well as the gradient $\mathbf{g}$ are computed. Finally, the rotation parameters R_{rk} are determined by solving the SCI eigenvalue problem. This is formally very similar to the AH equation and can be done iteratively by a P-space Davidson-algorithm[98] as described in Ref. 39 and 42. A level shift parameter λ can be used to restrict the step length, cf. eq. (11).

The Fock matrix and integrals $(tu|vw)$ are computed using efficient density fitting approximations (cf. SM). The computational effort of the SCI method then scales as $\mathcal{O}(N_{\mathrm{AO}}^3 N_{\mathrm{occ}})$ (if the size of the active space is assumed to be constant) and therefore allows the treatment of significantly larger systems than with second-order methods. However, convergence acceleration as described in Sec. III is vital for robust convergence. And even with acceleration methods, the SCI convergence can be slow, as shown for some examples in Sec. V.

C. Combined first- and second-order orbital optimization

In order to improve convergence of the SCI method, we now propose a method which combines a second-order optimization of all orbital rotations that involve active orbitals with a SCI like first-order optimization of the remaining closed-virtual rotations. The Hessian of the rotations involving active orbitals only requires the calculation of the integrals J^{tu}_{rs}= $(rs|tu)$ and K^{tu}_{rs}= $(rt|su)$ with two active indices t, u. The number of these integrals only scales as $\mathcal{O}(N^2_{\mathrm{AO}})$, if the active space is assumed to be independent of the molecular size. Their construction does not introduce a large computational overhead if density fitting is used (cf. SM).

To derive the combined first- and second-order optimization (in the following denoted SO-SCI method), we define two separate transformation operators similar to eq. (5). We refer to the rotations with active orbitals as *active rotations*. The operator for this active transformation is denoted $\hat{A}$. The remaining closed-virtual rotations are called *inactive rotations* and the associated transformation operator is $\hat{C}$:

$$\hat{A} = \sum_{t}^{act}\Big[\sum_{a}^{virt} R^A_{at}\hat{E}_{at} + \sum_{i}^{inact} R^A_{ti}\hat{E}_{ti} + \sum_{u(u<t)}^{act} R^A_{tu}(\hat{E}_{tu} - \hat{E}_{ut})\Big], \tag{21}$$

$$\hat{C} = \sum_{i}^{inact}\sum_{a} R^C_{ai}\hat{E}_{ai}. \tag{22}$$

For CASSCF active-active rotations are excluded, and the last term in eq. (21) disappears. Note that operators $\hat{E}_{ka}$ and $\hat{E}_{it}$ do not contribute, since their action on the wavefunction yields zero.

We now separate the rotations of the inactive rotations from the rotations with active orbitals:

$$|\Psi\rangle = \exp(\hat{A})\exp(\hat{C})\,|0\rangle\,. \tag{23}$$

Note that $\hat{A}$ and $\hat{C}$ do not commute, and therefore eq. (23) should be considered as an Ansatz. To stay consistent with this separation, the orbital transformation in equation (2) is defined accordingly:

$$\mathbf{U} = \exp(\mathbf{R}^C)\exp(\mathbf{R}^A). \tag{24}$$

The energy expanded with help of the BCH series then reads:

$$\begin{aligned} E(\mathbf{R}_A, \mathbf{R}_C) =\ & \langle 0|\,\hat{H}\,|0\rangle + \langle 0|\,[\hat{H},\hat{C}]\,|0\rangle + \langle 0|\,[\hat{H},\hat{A}]\,|0\rangle + \\ & + \frac{1}{2}\Big[\,\langle 0|\,[[\hat{H},\hat{A}],\hat{A}]\,|0\rangle + \langle 0|\,[[\hat{H},\hat{C}],\hat{C}]\,|0\rangle\,\Big] \\ & + \langle 0|\,[[\hat{H},\hat{A}],\hat{C}]\,|0\rangle + \ldots \qquad (25) \\ =\ & E_0 + \mathbf{g}^{\mathsf{T}}\mathbf{x} + \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{h}\mathbf{x} + \ldots, \qquad (26) \end{aligned}$$

where $\mathbf{x} = \begin{pmatrix}\mathbf{x}^A\\ \mathbf{x}^C\end{pmatrix} \equiv \begin{pmatrix}\mathbf{R}^A\\ \mathbf{R}^C\end{pmatrix}$. This yields for the gradient

$$g_{rk} = 2\,\langle 0|\hat{H}|rk\rangle\,. \tag{27}$$

The double commutator $\langle 0|[[\hat{H},\hat{A}],\hat{A}]|0\rangle$, which involves the Hessian block for the active rotations, can be computed without approximations from the integrals J^{tu}_{rs}, K^{tu}_{rs} with at least 2 active indices. However, the exact computation of the Hessian parts arising from the operator $\hat{C}$ would in addition require all integrals J^{ij}_{rs}, K^{ij}_{rs} with inactive labels i, j, leading to the higher scaling of second-order methods. To avoid this, the Hamiltonian operator in the double commutators that contain $\hat{C}$ is replaced by the effective SCI Hamiltonian $\hat{H}^{eff}$, i.e.,

$$\begin{aligned} \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{h}\mathbf{x} \approx & \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{h}^{eff}\mathbf{x} = \frac{1}{2}\,\langle 0|\,[[\hat{H},\hat{A}],\hat{A}]\,|0\rangle \\ & + \frac{1}{2}\,\langle 0|\,[[\hat{H}^{eff},\hat{C}],\hat{C}]\,|0\rangle + \langle 0|\,[[\hat{H}^{eff},\hat{A}],\hat{C}]\,|0\rangle\,. \end{aligned} \tag{28}$$

We had inconsistent notations $\hat{H}_{eff}$ or $\hat{H}^{eff}$. I now use the latter throughout. Also, I capitalized Hessian throughout. By comparing eq. (18) and eq. (28) one finds that the approximate Hessian elements $h^{eff}_{rk,sl}$ equal those of the effective SCI Hessian $h^{\mathrm{SCI}}_{rk,sl} = 2\,\langle rk|\hat{H}^{eff} - E^{(0)}|sl\rangle$ for all blocks with at most one active index, i.e.

$$h^{eff}_{ai,bj} = h^{SCI}_{ai,bj}, \quad h^{eff}_{ai,bu} = h^{SCI}_{ai,bu}, \quad h^{eff}_{ai,uj} = h^{SCI}_{ai,uj} \tag{29}$$

(cf. SM). This is not the case, however, for the Hessian blocks with 2 or more active indices, even if these would also be approximated by $\langle 0|[[\hat{H}^{eff},\hat{A}],\hat{A}]|0\rangle$. Thus, in order to determine the orbital rotation parameters $\mathbf{R}$, we solve an AH equation in which the exact Hessian is used for the blocks $h_{at,bu}$ and $h_{ti,uj}$, while the approximate SCI Hessian is used for the blocks shown in eq. (29). Furthermore, the SCI overlap contributions $\langle ai|bj\rangle = 2\delta_{ij}\delta_{bj}$ are taken into account for the closed-virtual rotations, since this improves convergence. The eigenvalue problem is solved iteratively with the P-space Davidson method (micro-iterations).[39,42] Only the P-space part of the Hessian is computed explicitly, while the product $\mathbf{h}^{eff}\mathbf{x}$ is computed directly from the integrals and density matrices in each micro-iteration. As in the SCI method, a level shift parameter λ is used to restrict the step length.

Convergence of the micro-iterations needed to solve the AH equations can be improved by generating after the initial CI of each macro-iterations natural active orbitals by diagonalizing the active 1-RDM. This keeps the singly excited configurations in eq. (13) orthogonal. Furthermore, block diagonalizing the Fock matrix in the inactive and virtual subspaces minimizes couplings via off-diagonal elements in the approximate SCI Hessian part. The integrals J^{tu}_{rs}, K^{tu}_{rs} as well as the CI vector(s) are transformed accordingly. The transformation of the CI

vectors only depends on the unitary transformation in the active space and can be done non-iteratively.[99]

Finally we note that instead of the expansion in (23) one could also make the *ansatz* $|\Psi\rangle = \exp(\hat{C})\exp(\hat{A})\,|0\rangle$ (note that $\hat{A}$ and $\hat{C}$ do not commute) or $|\Psi\rangle = \exp(\hat{A}+\hat{C})\,|0\rangle$. In both cases, the same Super-CI Hessian block $h^{SCI}_{ai,bj}$ is obtained, but the coupling blocks slightly change. In practice we found that the convergence of the three ansätze is nearly the same, with a slight advantage for the approach shown in equation (23), which has therefore been used for all calculations in this paper.

D. Using the WMK method for active-virtual orbital rotations

Much faster and more stable convergence than with the SO-AH method can be achieved by expanding the energy up to second-order in the orbital changes, which are represented by the matrix $\mathbf{T} = \mathbf{U} - \mathbf{1} = \mathbf{R} + \frac{1}{2!}\mathbf{R}^2 + \ldots$. The expansion $E^{(2)}(\mathbf{T})$ is of infinite order in $\mathbf{R}$, and is periodic in individual orbital rotations, as the exact energy.[39–42] Minimization of $E^{(2)}(\mathbf{T},\mathbf{c})$ with respect to the orbitals and CI coefficients $\mathbf{c}$ yields a set of non-linear equations, which can be solved iteratively (micro-iterations). This so-called WMK method converges very fast, often in only 3-4 macro-iterations. It therefore minimizes the number of integral transformations, but the micro-iterations needed to solve the non-linear equations are more expensive than in the uncoupled AH method. To a large extent, this is due to the additional CI steps needed in the micro-iterations, and the subsequent recomputation of the density matrices and other intermediates. For large molecules the method becomes very expensive, since it requires all integrals with two occupied orbitals, similar to the UC-AH method.

In the current work we have attempted to combine the WMK method for active orbital rotations with the SO-SCI method for the remaining orbital rotations. However, it turned out that including the inactive-active orbital rotations in this treatment does not lead to stable convergence, since the integrals J^{tu}_{rs}, K^{tu}_{rs} with at least 2 active indices are not sufficient to compute $E^{(2)}(\mathbf{T}^A)$ without further approximations (with $\mathbf{T}^A = \mathbf{R}^A + \frac{1}{2!}(\mathbf{R}^A)^2 + \ldots$). It is possible, however, to optimize the active-virtual orbitals along with the CI coefficients using the WMK method. This corresponds to a second-order WMK optimization with frozen inactive orbitals. Subsequently, the remaining orbital rotations are optimized using the SO-SCI method, as described in section II C. For this, the density matrices are recomputed with the CI coefficients obtained in the initial WMK optimization. However, the integrals and the Fock matrix are kept unchanged, to avoid another Fock-matrix and integral evaluation. Various tests have shown that recomputing the integrals has only a minor effect on the convergence and is not worth the effort. Thus, the couplings between the virtual-active and all remaining rotations are neglected entirely in this approximation.

We found that this hybrid WMK-SO-SCI improves convergence only in cases where the active-virtual orbital rotations are particularly large, and even then the savings are small. In many cases the additional cost for the additional CI-steps in the WMK optimization of the active orbitals exceeds the savings by the reduction of the number of macroiterations. The method is therefore not used by default, and not further considered in this paper.

E. State averaged MCSCF

Excited states are best treated with state-averaged MCSCF/CASSCF,[5,38,67] since this avoids root flipping problems in the optimization.[38] In state averaged MCSCF, the weighted energy average of several states is minimized

$$E_{\mathrm{av}} = \sum_n W_n E_n, \qquad \sum_n W_n = 1, \tag{30}$$

where W_n are the weights (which are mostly the same for all states, but other choices are possible, see, e.g., Ref. 100). Apart from computing several CI vectors in the CI steps, the only change in the algorithms is to use state-averaged density matrices

$$D_{tu} = \sum_n W_n \sum_{IJ} c^n_I c^n_J \langle \Phi_I | \hat{E}_{tu} | \Phi_J \rangle , \tag{31}$$

$$D_{tu,vw} = \sum_n W_n \sum_{IJ} c^n_I c^n_J \langle \Phi_I | \frac{1}{2}\left[\hat{E}_{tu,vw} + \hat{E}_{ut,vw}\right] | \Phi_J \rangle . \tag{32}$$

Since the state-averaged MCSCF method is based on the direct minimization of an energy functional, it is not possible to define a SCI wavefunction for it. Nevertheless, the SCI and SO-SCI methods as outlined above can straightforwardly be used, simply by replacing the state-specific densities by the state-averaged ones.

III. L-BFGS CONVERGENCE ACCELERATION

The two-step optimization of the MCSCF wavefunction, i.e. the alternating optimization of the CI coefficients and the orbitals, yields a first-order method even if a second-order orbital optimization method is used. The reason for this first-order convergence is the absence of the explicit coupling between the CI and the orbital optimization. This first-order convergence can be extremely slow, especially when approaching the minimum. In addition, a first-order orbital optimization further slows

down the convergence, which can lead to hundreds of iterations and may therefore require the calculation of hundreds of Fock matrices.

Already Malmqvist et al.[9] have mentioned that a QN approach can be used to approximately account for the orbital-CI coupling and thus accelerate the convergence, but details were not given. Such an approach has also been used in part I of this series[42] to speed up the micro-iterations in a second-order optimization. In this section we generalize this idea and show how the L-BFGS method can be used as a general convergence accelerator. The acceleration is based on a preconditioning of the gradient and a post-processing of the step resulting from an optimization algorithm, as proposed by Nocedal.[95] The resulting method can be easily combined with all presented orbital optimizations and can also be applied in other electronic structure methods.

In the L-BFGS method of Nocedal[95] the inverse of the QN Hessian is iteratively constructed by the following definition:

$$\left[\mathbf{h}_{n+1}^{QN}\right]^{-1} = \left(\mathbf{I} - \frac{\mathbf{x}_n\mathbf{y}_n^\mathsf{T}}{\rho_n}\right)\left[\mathbf{h}_n^{QN}\right]^{-1}\left(\mathbf{I} - \frac{\mathbf{y}_n\mathbf{x}_n^\mathsf{T}}{\rho_n}\right) + \frac{\mathbf{x}_n\mathbf{x}_n^\mathsf{T}}{\rho_n} \tag{33}$$

with the BFGS vectors of iteration $n+1$:

$$\mathbf{y}_n = \mathbf{g}_{n+1} - \mathbf{g}_n, \tag{34}$$

$$\rho_n = \mathbf{y}_n^\mathsf{T}\mathbf{x}_n. \tag{35}$$

Here $n+1$ is the current iteration, and n iterations have been carried out before. The gradient vector $\mathbf{g}_{n+1}$ is computed at the beginning of the current iteration. The optimization step $\mathbf{x}_{n+1}$ is calculated via the Newton-Raphson equation:

$$\mathbf{x}_{n+1} = -\left[\mathbf{h}_{n+1}^{QN}\right]^{-1}\mathbf{g}_{n+1}. \tag{36}$$

An update of the BFGS Hessian $\mathbf{h}^{QN}$ requires a consistent pair $\{\mathbf{x}_n, \mathbf{y}_n\}$ of the update vector $\mathbf{x}_n$ and gradient change $\mathbf{y}_n$. In the L-BFGS method, the set of the BFGS pairs $\{\mathbf{x}_i, \mathbf{y}_i\}$ can be limited to the last m recent vector pairs (our default value is m=10). The iterative construction of the inverse Hessian is stopped after m recursion steps with an initial inverse Hessian $\mathbf{h}_0^{-1}$. As shown in ref. 95, it is possible to unroll the recursive construction of the inverse L-BFGS Hessian $[\mathbf{h}_{n+1}^{QN}]^{-1}$ when only its action on a vector is needed. This so called two-loop recursion scheme is shown in Figure 1 and provides an efficient general implementation of the L-BFGS method. The first loop (lines 2 - 6) can be seen as a preconditioning of the gradient ($\mathbf{g}_{n+1} \rightarrow \bar{\mathbf{g}}_1$) and the second loop (lines 8 - 12) is a post-processing of the calculated step ($\bar{\mathbf{x}}_1 \rightarrow \mathbf{x}_{n+1}$). The update (line 7) returns a step from the preconditioned gradient ($\bar{\mathbf{g}}_1 \rightarrow \bar{\mathbf{x}}_1$). In the original L-BFGS method, the initial Hessian $\mathbf{h}_0$ is a scaled identity matrix, where the scaling is adjusted in each iteration.[96] However, as pointed out by Nocedal,[96] the

Input: Gradient $\mathbf{g}_{n+1}$ and the last m $\{\mathbf{y}_i\}, \{\mathbf{x}_i\}$, and $\{\rho_i\}$
Output: Step $\mathbf{x}_{n+1} = -[\mathbf{h}_{n+1}^{QN}]^{-1}\mathbf{g}_{n+1}$

1: $\bar{\mathbf{g}}_{m+1} = \mathbf{g}_{n+1}$
2: **for** $i = m, m-1, \ldots, 1$ **do**
3: $\quad j = i + n - m$
4: $\quad \alpha_i = \rho_j \mathbf{x}_j^\mathsf{T}\bar{\mathbf{g}}_{i+1}$ (store α_i)
5: $\quad \bar{\mathbf{g}}_i = \bar{\mathbf{g}}_{i+1} - \alpha_i\mathbf{y}_j$
6: **end for**
7: $\bar{\mathbf{x}}_1 = -\mathbf{h}_0^{-1}\bar{\mathbf{g}}_1$
8: **for** $i = 1, \ldots, m$ **do**
9: $\quad j = i + n - m$
10: $\quad \beta_i = \rho_j \mathbf{y}_j^\mathsf{T}\bar{\mathbf{x}}_i$
11: $\quad \bar{\mathbf{x}}_{i+1} = \bar{\mathbf{x}}_i + \mathbf{x}_j(-\alpha_i - \beta_i)$ (read α_i)
12: **end for**
13: $\mathbf{x}_{n+1} = \bar{\mathbf{x}}_{m+1}$

FIG. 1. The L-BFGS two-loop recursion algorithm for calculating the action of the inverse L-BFGS Hessian.[95]

choice of the initial Hessian is arbitrary as long as it stays positive definite. We use this flexibility and replace the step calculation in line 7 by the step returned from an AH calculation, carried out with the updated gradient vector $\bar{\mathbf{g}}_1$. A necessary criterion for convergence is that the underlying (approximate) Hessian is positive definite, otherwise the BFGS Hessian will lose its positive definite form. For example, the optimization method could be the SCI method, where the AH procedure ensures non-negative eigenvalues. Alternatively, either the exact orbital Hessian or the effective orbital Hessian defined in section II C can be used in the AH procedure.

It is also possible to include the CI gradient- and update vectors in the L-BFGS procedure. To test how well this method works for solving a CI eigenvalue equation we have applied it to the direct-CI method (without any orbital optimization). In this method, the CI update is normally computed as

$$\mathbf{g} = (\mathbf{H} - E\mathbf{1})\mathbf{c}, \quad \mathbf{c}^\mathsf{T}\mathbf{c} = 1, \tag{37}$$

$$\Delta c_I = -\frac{g_I}{H_{II} - E - \lambda}, \tag{38}$$

where H_{II} are the diagonal elements of the Hamiltonian, $E = \mathbf{c}^\mathsf{T}\mathbf{H}\mathbf{c}$ is the current energy, and λ a level shift for damping. The update can be improved by using the P-space method outlined in refs. 39 and 42. In this method a reduced Hamiltonian, built from a predefined set of P-space configurations and the current CI vector(s), is diagonalized in each iteration. The P-space includes the dominating configurations, and the P-space Hamiltonian is explicitly computed in this subspace. This yields the optimum P-space coefficients c_P and an improved gradient vector, which is zero in the P-space. Using this improved gradient vector, the coefficient of the remaining Q-space configurations are updated as in eq. (38). Optionally, further CI iterations can be carried out, and in each of these iterations the reduced Hamiltonian is augmented by a contracted Q-space function for each opti-

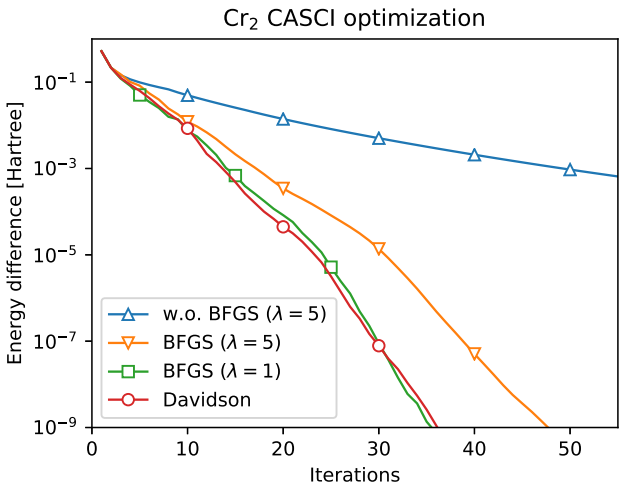


FIG. 2. Example: CASCI of the chromium dimer (r = 1.69Å), aug-cc-pVQZ basis, and full valence active space. Orbitals are obtained from the converged CASSCF solution. The difference to the final energy (DE) is shown for the non accelerated direct CI method with $\lambda = 5$ (blue), the BFGS accelerated method with $\lambda = 5$ (orange) and $\lambda = 1$ (green), as well as the conventional P-space Davidson method (red).

mized state. The BFGS vector pair $\{\mathbf{x}_n, \mathbf{y}_n\}$ is obtained after the last diagonalization of the reduced Hamiltonian, where consistent CI and gradient vectors are available. The CI update from the post-BFGS method is added to the CI vector, which is finally renormalized.

Figure 2 shows the convergence of the energy of a CAS-CI ground state optimization of the chromium dimer molecule using the full-valence CAS(12,12) active space (107216 Slater determinants, D_{2h} symmetry used). This is a particularly difficult case since there are multiple configurations strongly contributing to the wavefunction. We compare the direct-CI method with and without L-BFGS acceleration with the P-space Davidson method.[39,98] The direct-CI method without the acceleration requires a quite large damping parameter $\lambda = 5$ to obtain convergence. The L-BFGS accelerated version for $\lambda = 5$ is additionally shown and converges significantly faster. Furthermore, the L-BFGS allows a lower damping parameter of $\lambda = 1$ which again improves convergence. The L-BFGS accelerated direct-CI shows the very similar convergence behavior as the variational Davidson method. This is a surprising finding, since the Davidson method yields the best possible variational energy in the current space of expansion vectors. In this particular example, the numbers of expansion vectors in the L-BFGS and the Davidson method are kept equal.

We now describe in more detail how the L-BFGS convergence acceleration is implemented into the MCSCF two-step method. An overview of the single MCSCF iteration is presented in Figure 3. As we have seen in the direct-CI example, the L-BFGS acceleration is also capable of optimizing the CI coefficients. For this reason, the CI coefficients and the orbital rotation generators $\mathbf{R}$ are included in the L-BFGS algorithm to increase the amount of coupling. Hence, the displacements $\mathbf{x}_n$ and the change in the gradient $\mathbf{y}_n$ of the MCSCF problem are:

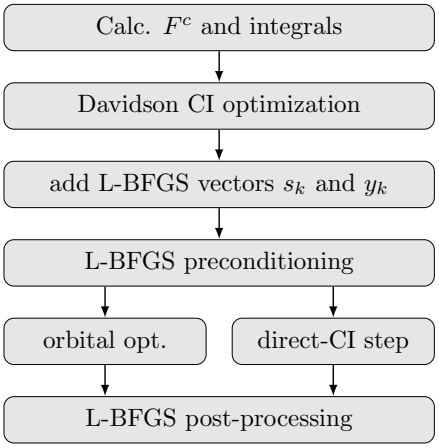


FIG. 3. Flowchart of one L-BFGS accelerated MCSCF iteration

$$\mathbf{x}_n = \begin{pmatrix} \mathbf{c}_{n+1} - \mathbf{c}_n \\ \mathbf{R}_n \end{pmatrix} \quad \text{and} \quad \mathbf{y}_n = \begin{pmatrix} \mathbf{g}^c_{n+1} - \mathbf{g}^c_n \\ \mathbf{g}^o_{n+1} - \mathbf{g}^o_n \end{pmatrix}. \tag{39}$$

The CI gradient $\mathbf{g}^c$ is obtained from the residual of the CI optimization, and the orbital gradient is calculated with the updated density. At this point, we have a consistent set of the step and the gradient for the BFGS vector pair $\{\mathbf{x}_n, \mathbf{y}_n\}$. However, particularly because of the in-iteration optimization of CI coefficients, the curvature condition of the BFGS method[96]

$$\rho_n = \mathbf{x}_n^\mathsf{T} \mathbf{y}_n > 0, \tag{40}$$

as well as the condition that the energy should decrease sufficiently,[96] are not necessarily satisfied. To maintain the positive definite form of the BFGS Hessian, a vector pair $\{\mathbf{x}_n, \mathbf{y}_n\}$ is only added to the L-BFGS vector set if the curvature condition is fulfilled. When the energy increased, we discard the current set of BFGS vectors and restart the extrapolation. Next, the orbital optimization and the direct-CI update are carried out with the preconditioned gradient. All of the orbital optimization methods described in Sec. II can be used. Even with the second-order UC-AH orbital optimization convergence is improved by the L-BFGS acceleration, since the orbital Hessian does not include the coupling with the CI coefficients. The post-processed step $\mathbf{R}$ of the orbital optimization is rescaled to be within a trust radius. How large is the trust radius? This is necessary because the BFGS method usually returns a good search direction but sometimes overshoots the step length. The iteration is finalized by transforming the orbitals according to eq. (3) and orthonormalizing the updated CI vector.

IV. STARTING GUESSES

For fast and reliable convergence of MCSCF calculations a reasonable starting guess for the orbitals is required. In particular, the initial active orbitals should have the the qualitatively correct character in order to avoid convergence to undesired solutions or local minima, which often happens in cases with many nearly degenerate states unless a good starting guess is used. In this work we use the "Automated Construction of Molecular Active Spaces from Atomic Valence Orbitals" (AVAS)[101] procedure to generate starting orbitals. However, instead of using converged restricted Hartree-Fock (ROHF) orbitals as input for this procedure, we use orbitals obtained from an atomic density guess as described in Ref. 94. In this method an approximate molecular density matrix is constructed by superposition of atomic density matrices, which are computed using a minimal basis of atomic orbitals that are stored in the basis set library. The effective occupation numbers are pre-optimized and also stored in a library. Using this density matrix a closed-shell Fock matrix is computed and then diagonalized, yielding the initial orbitals.

The AVAS method determines an orthogonal transformation of the input orbitals such that the overlap of the active orbitals with a specified set of atomic orbitals (AOs) (e.g. d-orbitals in transition metals, or p_π orbitals in conjugated or aromatic systems) is maximized. The target set of AOs has to be specified in the input. The transformation does not mix occupied and virtual MOs, but can mix closed- and open-shell orbitals. Usually, a Hartree-Fock configuration is assumed to determine the number of occupied orbitals.

Obviously, the AVAS the method is not fully automatic, since it requires some chemical intuition to select the target orbitals. This selection is closely related to the intuition needed to select an active space and sometimes requires some experimentation. Note that AVAS may generate more active orbitals than target functions have been specified, if the latter contribute significantly to two or more orbitals. However, for the CASSCF calculations presented in this paper, the choice of the target orbitals and active spaces was always straightforward. We found that the AVAS starting guess works very well, and avoids the need to carry out an ROHF calculation before the CASSCF. This is important since for typical MCSCF cases ROHF is often a very poor approximation and convergence may be difficult to achieve.

In principle, the described AVAS orbital guess can also be used in ROHF calculations. However, the number of open-shell orbitals in the ROHF is often smaller than the active space predicted by AVAS, and then one or more AVAS active orbitals may be doubly occupied in the ROHF wavefunction. In such cases the initial energy is not invariant to the order of the active AVAS orbitals, and we found in several calculations that the order of the initial orbitals was not consistent with that in the final optimized wavefunction. In such cases, AVAS may actually slow down the ROHF convergence, and in some cases convergence to the correct energy could not be achieved at all. Therefore, we used in all single-determinant calculations the orbitals from the atomic density guess without AVAS rotation. We hope that a solution to this problem can be found in future.

TABLE I. Results for the aromatics benchmark set.[52] The total number of MCSCF iterations and the total computation times (in minutes) are presented. All numbers are summed over the 21 calculations.

	avdz		avtz		avqz	
Orb. Opt	Iter.	Time	Iter.	Time	Iter.	Time
Without L-BFGS acceleration:						
SCI	1231	28.1	1261	101.6	1274	443.2
UC-AH	173	13.4	173	78.6	176	405.0
SO-SCI	538	9.6	569	40.2	571	177.8
With L-BFGS acceleration:						
SCI	435	10.5	442	35.8	447	152.9
UC-AH	139	11.9	140	70.3	140	353.0
SO-SCI	209	3.9	213	14.9	214	64.7
WMK	74	23.7	76	131.0	76	591.1

V. BENCHMARK CALCULATIONS

In this section we present benchmarks of CASSCF calculations in order to demonstrate the convergence properties and the efficiency of our new methods for different applications. These include calculations for 21 typical aromatic molecules and various large transition metal complexes.

The methods have been implemented in the Molpro software package.[? ?] Our default convergence criterion for the first-order methods requires that the CI and orbital gradients are lower than 10^{-5}. Additionally, the energy change between two successive iterations must be lower than 10^{-9}, which is almost every time the case when the gradient criterion is fulfilled. All calculations were run on a single dedicated computing node with two Xeon CPU E5-2650 v4 processors (2×12 cores, 2.20GHz). Unless otherwise noted 15 cores (MPI processes) were used. All times quoted are elapsed times.

A. Aromatic systems

In the first benchmark set the lowest π-π^* excitation energies are computed for 21 aromatic systems using SA-CASSCF with 2 states included. The benchmarks set has been introduced by by Menezes et al.[52] and has also been

used in part I of this series. All geometries were obtained from Ref. 52. The active orbitals are the π orbitals of the aromatic systems (in some cases including π orbitals of neighboring oxygen or nitrogen atoms), and the benchmark set comprises active spaces between CAS(6,6) and CAS(12,12). As in our previous calculations,[42] we slightly adjusted some of the original active spaces by removing doubly occupied orbitals. The starting orbitals were generated by the AVAS procedure,[101] based on atomic density orbital guesses as described in section IV. We did calculations with the aug-cc-pVDZ (avdz), aug-cc-pVTZ (avtz), and aug-cc-pVQZ (avqz) basis sets.[102] Molecular symmetry was not used.

The results are presented in Table I, where the numbers of iterations and the computation times are summed over all 21 calculations. We compare the SCI, SO-SCI, and UC-AH orbital optimizations, as discussed in Sec II. The calculations are done with and without the L-BFGS acceleration. For comparison, the results of the second-order method of Werner et al.[39–42] (WMK-method) are also shown.

For the molecules in this benchmark set the SO-SCI method converges about twice as fast as the SCI method. In terms of computation times, the savings are somewhat smaller due to the additional integral evaluations and Hessian constructions in the SO-SCI method. A further reduction of the number of iterations is achieved if the full second-order orbital optimization (UC-AH) is used. However, in this case the computation times are 2-3 times longer than for the SO-SCI method, which is due to the much more expensive integral transformations. The L-BFGS acceleration reduces the number of iterations and timings for the SCI and SO-SCI methods by more than a factor of 2. The acceleration is smaller for the UC-AH method, indicating that the L-BFGS procedure compensates not only for the missing orbital-CI couplings, but also for some of the approximations in the SCI or SO-SCI orbital optimizations. Overall, the SO-SCI method with L-BFGS acceleration is for all basis sets most efficient.

The WMK-method implemented as described in Ref. 42 requires by far the lowest number of iterations, because of its very rapid second-order convergence. For most molecules, only 3-4 iterations are required. Nevertheless, the computation times are considerably larger than in the first-order methods. This is due to the much higher cost of the WMK micro-iterations, since these include the orbital-CI coupling. Furthermore, for these rather simple cases the uncoupled first-order methods converge relatively fast, and therefore the reduction of the number of iterations by inclusion of the orbital-CI coupling in the WMK method does not outweigh the additional cost.

B. $[Cu_2O_2]^{2+}$ Isomerization

The next example is the isomerization of the $[(NH_3)_3Cu]_2O_2^{2+}$ complex from the bis(μ-oxo) to the μ-μ^2 : μ^2 peroxo structure. In the last two decades, this system gained attention because of the poor results of the CASPT2 method.[104–107] The CASPT2 results could be significantly improved by increasing the active space in a RASPT2 calculation.[9] In previous calculations,[42] we found for some structures a very strong coupling between the CI coefficients and orbitals for the CAS(16,14) active space introduced in Ref. 106. The coupling is due to a qualitative change of the weakly occupied $23a_g$ and $13b_g$ orbitals along the isomerization pathway. This strong coupling leads to an extremely slow convergence for the uncoupled two-step optimization and is therefore a good example to demonstrate the power of the L-BFGS convergence acceleration. The strong coupling and the associated convergence difficulties are removed when 2 more virtual orbitals are added to the active space, leading to CASSCF(16,16) calculations.

The isomerization pathway is modelled by a parameter F from the bis(μ-oxo) ($F = 0$) to the peroxo (F=100) structure, and six values of F were considered. The coordinates of the six structures were obtained from Ref. 106, where also more details on the active space are available. We did calculations for the ^1A_g ground state in the C_{2h} symmetry with the aug-cc-pVTZ basis set.[102] The starting orbitals for the first CASSCF calculation at $F = 0$ are obtained from an AVAS calculation. All further calculations are started with the final orbitals from the previous structure for the next smaller F value. We carried out calculations with and without the L-BFGS acceleration for the SCI, SO-SCI, and UC-AH optimizations.

Table II shows the number of iterations along the isomerization ($F = 0 - 100$) for the CASSCF(16,14) and CASSCF(16,16) calculations. We first discuss the CASSCF(16,14) calculations in the left part of the Table. If the L-BFGS acceleration is switched off, extremely slow convergence is obtained for all orbital optimization methods. Especially in the area $F = 60-100$, hundreds of iterations are required to obtain a gradient norm lower than 10^{-5}. This slow convergence can be clearly attributed to the absence of the CI-orbital coupling, since the pure second-order AH orbital optimization is converged in every step. Also, the CI optimization in the beginning of each macroiteration is converging quite fast, and so we can conclude that the reason for the slow convergence is the alternating optimization. When the L-BFGS acceleration is switched on, the convergence is strongly improved, and convergence is obtained for all structures in at most 20 iterations for the SO-SCI and UC-AH methods. The SCI converges slightly slower, but still with an acceptable speed.

With the (16,16) active space (41 410 450 Slater determinants) convergence is much faster and similar for

TABLE II. Number of MCSCF iterations of the isomerization of $[(NH_3)_3Cu]_2O_2^{2+}$ from the bis(μ-oxo) to the peroxo structure[103] with an active space of CAS(16,14) and CAS(16,16). The number of trial vectors $\sigma = Hc$ is shown for the CAS(16,16) case in parenthesis. Also, the total computation times summed over all six isomerization steps are presented in minutes.

	CAS(16,14)					
	without L-BFGS			with L-BFGS		
F	SCI	SO-SCI	UC-AH	SCI	SO-SCI	UC-AH
0	52	46	26	20	16	14
20	40	42	14	17	12	11
40	54	67	27	16	14	13
60	206	179	146	23	17	15
80	411	288	289	26	20	18
100	234	184	162	22	15	14
SUM	997	806	664	124	94	85
Time:	168.9	135.7	745.1	24.1	18.3	101.7

	CAS(16,16)			
	default		with minimal CI	
F	SCI	SO-SCI	SCI	SO-SCI
0	17 (102)	13 (91)	28 (46)	25 (41)
20	13 (95)	9 (67)	22 (32)	22 (32)
40	14 (87)	9 (66)	24 (33)	22 (31)
60	15 (83)	9 (61)	23 (32)	18 (27)
80	15 (75)	9 (46)	23 (26)	16 (23)
100	17 (80)	10 (48)	24 (27)	17 (24)
SUM	91 (522)	59 (379)	144 (196)	120 (178)
Time:	180.8	123.8	144.1	111.6

all values of F (right part of Table II). In this case the SO-SCI converges much faster than SCI, and also the number of CI steps is reduced by more than a factor of 2 by the SO-SCI relative to SCI. For this large active space the computational effort is dominated by the CI steps and the density matrix evaluations, and therefore the number of Hamiltonian actions on trial vectors ($\sigma = Hc$) is shown in parenthesis (the number of density matrix evaluations equals the number of iterations and is not included in these numbers). Obviously, for such large active spaces it is more important to minimize the number of CI steps than the number of orbital optimizations. We therefore did a second series of calculations in which the initial Davidson optimization of the CI vector in each macro-iteration was switched off once the L-BFGS was activated. For $F > 0$, this happened after the second iteration, for F=0 after the third iteration. Thus, after the first few macro-iterations, the CI optimization was done only with the L-BFGS acceleration method, similar to the example in Sec. III. This means that only one $\sigma = Hc$ step and one density matrix evaluation are necessary per macro iteration, and this is denoted "with minimal CI" in Table II. In this case the number of macro-iterations increases (by up to a factor of 2 for SO-SCI), but overall the number of CI steps is significantly reduced. As shown in Table II this reduces the computation times by 10-20%. Another possibility to increase the efficiency is to use graphical processing units (GPUs), but this requires the development of algorithms that depend on the available hardware.[108]

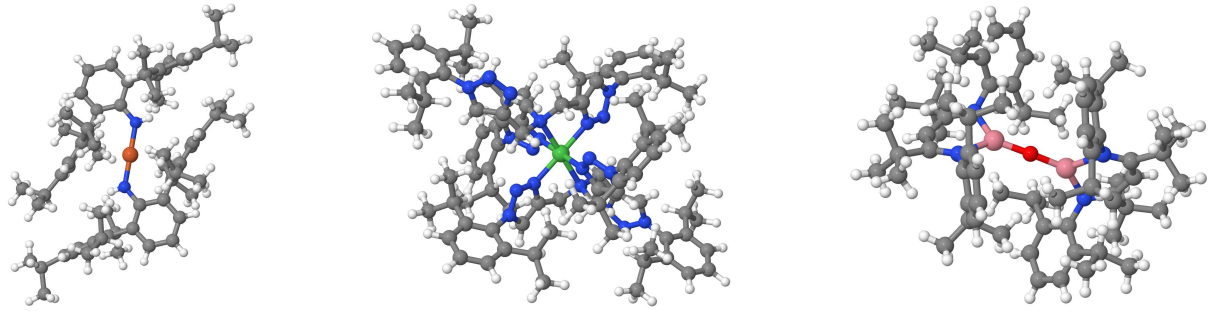
Unfortunately, it is difficult to predict how many CI steps in each iteration are optimal for a given molecule, but the minimal CI algorithm would certainly be advantageous for very large active spaces treated by the DMRG or FCIQMC methods.

C. Iron, Nickel and Cobalt complexes

The final benchmark calculations demonstrate the performance of the SCI and SO-SCI methods for three larger transition metal complexes. The first two complexes $FeC_{72}N_2H_{100}$ and $[NiC_{90}N_{20}H_{120}]^{2+}$ were obtained from Guo et al.[109] They have been previously used for benchmarking the PNO-NEVPT2[109] and PNO-CASPT2[52,111] methods. In both references, the preceding CASSCF implementation has been identified as a severe bottleneck. The third complex is the $Co_2ON_4C_{70}H_{106}$ system recently published by Roy et al.[110] The structures of all three systems are shown in Figure 4.

For the iron complex we calculated the triplet and quintet states, while for the nickel complex the singlet and the triplet states were optimized. All calculations were done with the def2-tzvp and the def2-tzvpp basis sets,[112] and the geometries were obtained from Ref. 109. We carried out single-determinant calculations (equivalent to ROHF) and double d-shell calculations, i.e. CASSCF(6,10) and CASSCF(8,10) for the iron and nickel complexes, respectively. More information about these choices can be found in the SM.

The results of the single-determinant calculations for the three complexes are presented in Table III, in which the convergence and timings of SCI, SO-CI, and ROHF calculations are compared. The latter calculations were done with the ROHF program in Molpro. All shown calculations were started with the atomic density orbital guess without AVAS. In case of the iron complex, we were not able to converge the ROHF calculations in less than hundred iterations. Also the SCI converged very slowly, and nearly hundred iterations were required for the iron and the cobalt complexes. The SO-SCI needed by far the lowest number of iterations, and in all cases it was significantly faster than the corresponding ROHF calculations. The computation time per SO-SCI iteration is slightly higher than in ROHF, since the two-electron integrals with at least two active indices have to be computed



$FeC_{72}N_2H_{100}$[109] $[NiC_{90}N_{20}H_{120}]^{2+}$[109] $Co_2ON_4C_{70}H_{106}$[110]

FIG. 4. Structures of the transition metal complexes for the calculations in Sec. V C.

TABLE III. Number of iterations (Iter.) and computation times (in hours) of the single-determinant calculations for large transition metal clusters.

		ROHF		SCI		SO-SCI		
Basis	State	Iter.	Time [h]	Iter.	Time [h]	Iter.	Time [h]	Energy
				$FeC_{72}N_2H_{100}$				
def2-tzvp	^3A_g		a)	98	11.5	41	4.4	-4156.068 370
(2939)	^5A_g		a)	83	9.7	27	3.0	-4156.160 091
def2-tzvpp	^3A_g		a)	98	16.6	42	6.6	-4156.120 185
(3785)	^5A_g		a)	91	15.4	27	4.3	-4156.211 440
				$[NiC_{90}N_{20}H_{120}]^{2+}$				
def2-tzvp	^3A	26	9.0	28	12.1	16	6.5	-6074.798 901
def2-tzvpp	^3A	26	12.7	27	16.9	16	9.5	-6074.841 763
				$Co_2ON_4C_{70}H_{106}$				
def2-tzvp	^7A	61	7.2	86	13.9	28	4.5	-5768.851 187
def2-tzvpp	^7A	62	10.8	87	21.2	29	7.1	-5768.892 008

a) No convergence after 100 iterations

along with the Fock matrices. However, the difference is not large and in the shown calculations compensated by a better convergence of SO-SCI. We therefore recommend to use the SO-SCI method for difficult open-shell Hartree-Fock calculations.

We note that for the quintet state of the iron complex, convergence could be significantly accelerated by using AVAS orbitals (based on the atomic density guess) with projection to the $3d_{\pm1}$, $3d_0$, and $3d_{+2}$ atomic orbitals only (the $3d_{-2}$ starting orbital being doubly occupied, and the N-Fe-N moiety on the z-axis, using C_{2h} symmetry). With these starting orbitals the SO-SCI converged in only 17 iterations, while convergence of the ROHF to the same final energy was only achieved after 150 iterations. However, for this starting guess we needed to know which d-orbital is doubly occupied in the optimized wavefunction, and this is not obvious in general.

Table IV presents the results for the double d-shell CASSCF calculations for the iron and the nickel complexes. Again, the SO-SCI needed the lowest computation time and the smallest number of iterations in all calculations, even though the SCI converged rather quickly as well.

Our last test system is the Co_2O complex from the recent work of Roy et al.,[110] using their DFT optimized geometry and the def2-tzvp and def2-tzvpp basis sets.[112] We carried out single determinant calculations for the septet high-spin state (cf. Table III), as well as CASSCF(14,10) and CASSCF(14,14) calculations for the singlet and triplet states (cf. SM for more information about these choices). The convergence of the energy for the triplet state with the SCI and the SO-SCI methods is shown in Fig. 5. In the first five iterations, the convergence of both methods is very similar. However, the convergence of SCI slows down considerably between iteration 10 and 55, and 84 iterations are required until convergence. The SO-SCI also slows down after iteration 10, but not as strongly as SCI, and 48 iterations are required to achieve convergence (cf. Table V).

With the larger CAS(14,14) active space the SCI and SO-SCI methods both converged in a reasonable number of iterations (see Table V), and significantly lower ener-

TABLE IV. Number of MCSCF iterations, timings (in hours) and energies (in Hartree) for the CASSCF calculations of the iron and the nickel complexes.

$FeC_{72}N_2H_{100}$ - CAS(6,10)						
		SCI		SO-SCI		
Basis	State	It.	Time [h]	It.	Time [h]	Energy
def2-tzvp	^3A_g	23	2.7	20	2.2	-4156.154 789
(2939)	^5A_g	38	4.4	32	3.5	-4156.227 647
def2-tzvpp	^3A_g	25	4.2	21	3.4	-4156.206 566
(3785)	^5A_g	40	6.7	34	5.6	-4156.279 034

$[NiC_{90}N_{20}H_{120}]^{2+}$ - CAS(8,10)						
		SCI		SO-SCI		
Basis	State	It.	Time [h]	It.	Time [h]	Energy
def2-tzvp	^{1}A	22	9.8	19	8.2	-6074.846 689
(4175)	^{3}A	25	11.5	17	7.4	-6074.923 467
def2-tzvpp	^{1}A	22	14.2	19	12.1	-6074.890 582
(5154)	^{3}A	25	16.5	17	10.9	-6074.966 405

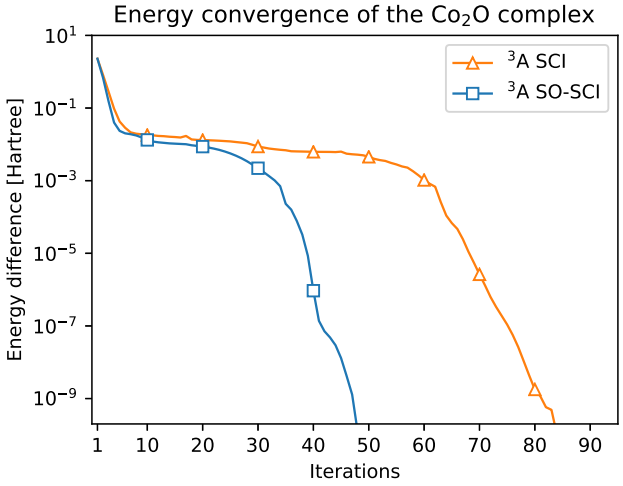


FIG. 5. Energy convergence of the triplet calculation of the cobalt complex with def2-tzvp and CAS(14,10). Shown is the difference to the converged energy in Table V.

TABLE V. Number of MCSCF iterations, timings (in hours) and energies (in Hartree) for the CASSCF calculation of $Co_2ON_4C_{70}H_{106}$ with CAS(14,10) and CAS(14,14).

		SCI		SO-SCI		
Basis	State	It.	Time [h]	It.	Time [h]	Energy
CAS(14,10)[a]						
def2-tzvp	^{1}A	56	8.2	33	4.1	-5768.857 959
(3051)	^{3}A	87	13.5	49	6.1	-5768.857 641
def2-tzvpp	^{1}A	50	13.2	34	8.5	-5768.898 833
(3937)	^{3}A	85	23.5	46	12.4	-5768.898 515
CAS(14,14)[b]						
def2-tzvp	^{1}A	40	6.3	34	5.3	-5768.941 538
(3051)	^{3}A	40	6.3	32	4.9	-5768.940 887
def2-tzvpp	^{1}A	39	9.2	32	8.2	-5768.982 734
(3937)	^{3}A	40	9.4	35	8.3	-5768.982 084

[a] AVAS guess with projection to Fe(3d), O(2p)
[b] AVAS guess with projection to Fe(3d,4d_0,4d_{+2}), O(2p)

gies were obtained. Even the calculations with the larger def2-tzvpp basis set (5154 CGTOs) required less than 10 hours computation time on a single workstation.

So far, we have never seen a case where the SCI required less iterations than the SO-SCI, and also the SO-SCI computation times were always smaller than the SCI ones.

VI. CONCLUSIONS

We proposed and benchmarked a new MCSCF orbital optimization method denoted SO-SCI, which combines a second-order (SO) optimization of the active orbitals with a first-order Super-CI (SCI) treatment for the inactive-virtual orbital rotations. The Hessian construction in the second-order optimization of the active orbitals requires the two-electron integrals $(rs|tu)$ and $(rt|su)$ in the MO basis with at least two active orbitals t, u, but using density fitting approximations these can be computed at relatively low cost along with the closed-shell Fock matrix. The method is compared with the pure SCI method which only requires Fock matrices [apart from the all active integrals $(tu|vw)$], as well as with the much more expensive second-order Augmented Hessian (AH) method, which requires all integrals $(rs|kl)$ and $(rk|sl)$ (k, l: all occupied orbitals).

In all three methods the orbitals and CI coefficients are optimized alternately. Due to the absence of a direct CI – orbital coupling in the effective Hessians, this leads to first-order convergence, which can in some cases be extremely slow. To solve this problem, we presented a convergence acceleration algorithm based on the L-BFGS method of Nocedal.[95] This strongly improves convergence, in particular in cases with strong orbital-CI couplings. It also improves the convergence of the first-order orbital optimization schemes, and it is demonstrated that with L-BFGS the SO-SCI method converges almost as quickly as the much more expensive AH method. In all cases tested so far, the SO-SCI method showed faster convergence than the SCI method and also the smallest computation times of the three methods. We have thus presented a framework for MCSCF optimization that is

generally applicable, and is highly efficient for many different types of applications, including cases with many closed-shell orbitals, large CI expansions, and challenging convergence.

The new MCSCF methods were tested for 21 aromatic molecules and 4 transition metal clusters. The largest calculations included 5154 basis functions and 413 optimized occupied orbitals and could still be done "overnight" on a single computer node. In all our CASSCF calculations the starting guesses were generated with the AVAS procedure[101] based an orbital guess from atomic densities.[94] If properly used, this method provides qualitatively correct starting orbitals and avoids the need of any preceding Hartree-Fock calculations. This is of paramount importance for typical MCSCF applications where Hartree-Fock is expected to perform poorly.

We also demonstrated that the SO-SCI method often converges much faster and more robustly than the standard ROHF method for single-determinant optimizations. It is therefore strongly recommended to use the SO-SCI method for Hartree-Fock calculations on large open-shell transition metal clusters. We also plan a generalization of the SO-SCI method for unrestricted Hartree-Fock (UHF) and Kohn Sham (UKS) calculations.

SUPPLEMENTARY MATERIAL

See supplementary material for explicit equations for the gradients and (approximate) Hessians, a description of our density fitting implementation, and details about the choice of active spaces for the transition metal clusters.

[1] F. Grein and T. C. Chang, Chem. Phys. Lett. **12**, 44 (1971).
[2] F. Grein and A. Banerjee, Chem. Phys. Lett. **31**, 281 (1975).
[3] F. Grein and A. Banerjee, Int. J. Quantum Chem. **9**, 147 (1975).
[4] A. Banerjee and F. Grein, Int. J. Quantum Chem. **10**, 123 (1976).
[5] K. Ruedenberg, L. M. Cheung, and S. T. Elbert, Int. J. Quantum Chem. **16**, 1069 (1979).
[6] B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, Chem. Phys. **48**, 157 (1980).
[7] B. O. Roos, Int. J. Quantum Chem. **18**, 175 (1980).
[8] P. E. M. Siegbahn, J. Almlöf, A. Heiberg, and B. O. Roos, J. Chem. Phys. **74**, 2384 (1981).
[9] P. Å. Malmqvist, A. Rendell, and B. O. Roos, J. Phys. Chem. **94**, 5477 (1990).
[10] G. Chaban, M. W. Schmidt, and M. S. Gordon, Theor. Chem. Acc. **97**, 88 (1997).
[11] C. Angeli, S. Evangelisti, R. Cimiraglia, and D. Maynau, J. Chem. Phys. **117**, 10525 (2002).
[12] J. E. Bates and T. Shiozaki, J. Chem. Phys. **142**, 044112 (2015).
[13] E. G. Hohenstein, N. Luehr, I. S. Ufimtsev, and T. J. Martínez, J. Chem. Phys. **142**, 224103 (2015).
[14] R. D. Reynolds, T. Yanai, and T. Shiozaki, J. Chem. Phys. **149**, 014106 (2018).
[15] C. Kollmar, K. Sivalingam, B. Helmich-Paris, C. Angeli, and F. Neese, J. Comput. Chem. **40**, 1463 (2019).
[16] U. Meier and V. Staemmler, Theor. Chim. Acta **76**, 95 (1989).
[17] E. Dalgaard and P. Jørgensen, J. Chem. Phys. **69**, 3833 (1978).
[18] E. Dalgaard, Chem. Phys. Lett. **65**, 559 (1979).
[19] D. L. Yeager, P. Albertsen, and P. Jørgensen, J. Chem. Phys. **73**, 2811 (1980).
[20] D. L. Yeager and P. Jørgensen, Mol. Phys. **39**, 587 (1980).
[21] P. Jørgensen, P. Albertsen, and D. L. Yeager, J. Chem. Phys. **72**, 6466 (1980).
[22] P. Jørgensen, J. Olsen, and D. L. Yeager, J. Chem. Phys. **75**, 5802 (1981).
[23] A. Igawa, D. L. Yeager, and H. Fukutome, J. Chem. Phys. **76**, 5388 (1982).
[24] J. Olsen, P. Jørgensen, and D. L. Yeager, J. Chem. Phys. **76**, 527 (1982).
[25] J. Olsen, P. Jørgensen, and D. L. Yeager, J. Chem. Phys. **77**, 356 (1982).
[26] D. L. Yeager, D. Lynch, J. Nichols, P. Jørgensen, and J. Olsen, J. Phys. Chem. **86**, 2140 (1982).
[27] J. Olsen, D. L. Yeager, and P. Jørgensen, in *Adv. Chem. Phys.*, Vol. 54 (John Wiley & Sons, Inc., 1983) pp. 1–176.
[28] P. Jørgensen, P. Swanstrøm, and D. L. Yeager, J. Chem. Phys. **78**, 347 (1983).
[29] B. H. Lengsfield III, J. Chem. Phys. **73**, 382 (1980).
[30] B. H. Lengsfield III, J. Chem. Phys. **77**, 4073 (1982).
[31] B. H. Lengsfield III and B. Liu, J. Chem. Phys. **75**, 478 (1981).
[32] H. J. A. Jensen and H. Ågren, Chem. Phys. Lett. **110**, 140 (1984).
[33] H. J. A. Jensen and P. Jørgensen, J. Chem. Phys. **80**, 1204 (1984).
[34] H. J. A. Jensen and H. Ågren, Chem. Phys. **104**, 229 (1986).
[35] H. J. A. Jensen, P. Jørgensen, and H. Ågren, J. Chem. Phys. **87**, 451 (1987).
[36] R. Shepard and J. Simons, Int. J. Quantum Chem. **18**, 211 (1980).
[37] H.-J. Werner and W. Meyer, J. Chem. Phys. **73**, 2342 (1980).
[38] H.-J. Werner and W. Meyer, J. Chem. Phys. **74**, 5794 (1981).
[39] H.-J. Werner and P. J. Knowles, J. Chem. Phys. **82**, 5053 (1985).
[40] P. J. Knowles and H.-J. Werner, Chem. Phys. Lett. **115**, 259 (1985).
[41] H.-J. Werner, in *Adv. Chem. Phys.*, Vol. 69 (John Wiley & Sons, Inc., 1987) pp. 1–62.
[42] D. A. Kreplin, P. J. Knowles, and H.-J. Werner, J. Chem. Phys. **150**, 194106 (2019).
[43] F. Lipparini and J. Gauss, J. Chem. Theory Comput. **12**, 4284 (2016).
[44] Q. Sun, J. Yang, and G. K.-L. Chan, Chem. Phys. Lett. **683**, 291 (2017).
[45] Y. Ma, S. Knecht, S. Keller, and M. Reiher, J. Chem. Theory Comput. **13**, 2533 (2017).
[46] B. O. Roos, P. Linse, P. E. Siegbahn, and M. R. Blomberg, Chem. Phys. **66**, 197 (1982).
[47] J. Finley, P.-Å. Malmqvist, B. O. Roos, and L. Serrano-Andrés, Chem. Phys. Lett. **288**, 299 (1998).
[48] K. Andersson, P.-Å. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).
[49] H.-J. Werner, Mol. Phys. **89**, 645 (1996).
[50] P. Celani and H.-J. Werner, J. Chem. Phys. **112**, 5546 (2000).
[51] C. Angeli, M. Pastore, and R. Cimiraglia, Theor. Chem. Acc. **117**, 743 (2007).
[52] F. Menezes, D. Kats, and H.-J. Werner, J. Chem. Phys. **145**, 124115 (2016).
[53] D. Kats and H.-J. Werner, J. Chem. Phys. **150**, 214107 (2019).
[54] P. E. Siegbahn, J. Chem. Phys. **72**, 1647 (1980).
[55] V. R. Saunders and J. H. van Lenthe, Mol. Phys. **48**, 923 (1983).
[56] H.-J. Werner and P. J. Knowles, J. Chem. Phys. **89**, 5803 (1988).
[57] P. J. Knowles and H.-J. Werner, Chem. Phys. Lett. **145**, 514 (1988).
[58] H. Werner and P. J. Knowles, Theor. Chim. Acta **78**, 175 (1990).
[59] P. J. Knowles and H.-J. Werner, Theor. Chim. Acta **84**, 95 (1992).

[60] P. Celani, H. Stoll, H.-J. Werner, and P. J. Knowles, Mol. Phys. **102**, 2369 (2004).
[61] K. R. Shamasundar, G. Knizia, and H.-J. Werner, J. Chem. Phys. **135**, 054101 (2011).
[62] P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka, and R. Shepard, Chem. Rev. **112**, 108 (2012).
[63] F. A. Evangelista, M. Hanauer, A. Köhn, and J. Gauss, J. Chem. Phys. **136**, 204108 (2012).
[64] M. Hanauer and A. Köhn, Chem. Phys. **401**, 50 (2012).
[65] M. Hanauer and A. Köhn, J. Chem. Phys. **137**, 131103 (2012).
[66] M. Hanauer and A. Köhn, J. Chem. Phys. **136**, 204107 (2012).
[67] K. K. Docken and J. Hinze, J. Chem. Phys. **57**, 4928 (1972).
[68] K. Ruedenberg and K. R. Sundberg, in *Quantum Sci. Methods Struct.*, edited by J.-L. Calais, O. Goscinski, J. Linderberg, and Y. Öhrn (Springer US, Boston, MA, 1976) pp. 505–515.
[69] L. M. Cheung, K. R. Sundberg, and K. Ruedenberg, Int. J. Quantum Chem. **16**, 1103 (1979).
[70] K. Ruedenberg, M. W. Schmidt, M. M. Gilbert, and S. Elbert, Chem. Phys. **71**, 41 (1982).
[71] B. O. Roos, in *Adv. Chem. Phys.*, Vol. 69 (John Wiley & Sons, Inc., 1987) pp. 399–445.
[72] J. Olsen, Int. J. Quantum Chem. **111**, 3267 (2011).
[73] K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. de Jong, J. Chem. Phys. **147**, 184111 (2017).
[74] J. Olsen, B. O. Roos, P. Jørgensen, and H. J. A. Jensen, J. Chem. Phys. **89**, 2185 (1988).
[75] D. Ma, G. Li Manni, and L. Gagliardi, J. Chem. Phys. **135**, 044128 (2011).
[76] K. D. Vogiatzis, G. Li Manni, S. J. Stoneburner, D. Ma, and L. Gagliardi, J. Chem. Theory Comput. **11**, 3010 (2015).
[77] G. H. Booth, A. J. W. Thom, and A. Alavi, J. Chem. Phys. **131**, 54106 (2009).
[78] R. E. Thomas, Q. Sun, A. Alavi, and G. H. Booth, J. Chem. Theory Comput. **11**, 5316 (2015).
[79] G. Li Manni, S. D. Smart, and A. Alavi, J. Chem. Theory Comput. **12**, 1245 (2016).
[80] A. A. Holmes, N. M. Tubman, and C. J. Umrigar, J. Chem. Theory Comput. **12**, 3674 (2016).
[81] S. Sharma, A. A. Holmes, G. Jeanmairet, A. Alavi, and C. J. Umrigar, J. Chem. Theory Comput. **13**, 1595 (2017).
[82] J. E. T. Smith, B. Mussard, A. A. Holmes, and S. Sharma, J. Chem. Theory Comput. **13**, 5468 (2017).
[83] D. Zgid and M. Nooijen, J. Chem. Phys. **128**, 144116 (2008).
[84] D. Ghosh, J. Hachmann, T. Yanai, and G. K.-L. Chan, J. Chem. Phys. **128**, 144117 (2008).
[85] G. K.-L. Chan and S. Sharma, Annu. Rev. Phys. Chem. **62**, 465 (2011).
[86] T. Yanai, Y. Kurashige, D. Ghosh, and G. K.-L. Chan, Int. J. Quantum Chem. **109**, 2178 (2009).
[87] Y. Ma and H. Ma, J. Chem. Phys. **138**, 224105 (2013).
[88] M. W. Schmidt and M. S. Gordon, Annu. Rev. Phys. Chem. **49**, 233 (1998).
[89] B. Levy and G. Berthier, Int. J. Quantum Chem. **2**, 307 (1968).
[90] B. Levy and G. Berthier, Int. J. Quantum Chem. **3**, 247 (1969).
[91] F. Aquilante, T. B. Pedersen, R. Lindh, B. O. Roos, A. Sánchez de Merás, and H. Koch, J. Chem. Phys. **129**, 024113 (2008).
[92] W. Györffy, T. Shiozaki, G. Knizia, and H.-J. Werner, J. Chem. Phys. **138**, 104104 (2013).
[93] R. Polly, H.-J. Werner, F. R. Manby, and P. J. Knowles, Mol. Phys. **102**, 2311 (2004).
[94] C. Köppl and H.-J. Werner, J. Chem. Theory Comput. **12**, 3122 (2016).
[95] J. Nocedal, Math. Comput. **35**, 773 (1980).
[96] J. Nocedal and S. Wright, *Numerical optimization* (Springer Science & Business Media, 2006).
[97] D. R. Yarkony, Chem. Phys. Lett. **77**, 634 (1981).
[98] E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[99] G. J. Atchity and K. Ruedenberg, J. Chem. Phys. **111**, 2910 (1999).
[100] M. P. Deskevich, D. J. Nesbitt, and H.-J. Werner, J. Chem. Phys. **120**, 7281 (2004).
[101] E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan, and G. Knizia, J. Chem. Theory Comput. **13**, 4063 (2017).
[102] R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).
[103] B. F. Gherman and C. J. Cramer, Coord. Chem. Rev. **253**, 723 (2009).
[104] M. Flock and K. Pierloot, J. Phys. Chem. A **103**, 95 (1999).
[105] M. F. Rode and H.-J. Werner, Theor. Chem. Acc. **114**, 309 (2005).
[106] C. J. Cramer, M. Włoch, P. Piecuch, C. Puzzarini, and L. Gagliardi, J. Phys. Chem. A **110**, 1991 (2006).
[107] P. Å. Malmqvist, K. Pierloot, A. R. M. Shahi, C. J. Cramer, and L. Gagliardi, J. Chem. Phys. **128**, 204109 (2008).
[108] B. S. Fales and B. G. Levine, J. Chem. Theory Comput. **11**, 4708 (2015).
[109] Y. Guo, K. Sivalingam, E. F. Valeev, and F. Neese, J. Chem. Phys. **144**, 094111 (2016).
[110] L. Roy, M. H. Al-Afyouni, D. E. DeRosha, B. Mondal, I. M. DiMucci, K. M. Lancaster, J. Shearer, E. Bill, W. W. Brennessel, F. Neese, S. Ye, and P. L. Holland, Chem. Sci. **10**, 918 (2019).
[111] F. Menezes, *A local complete active space 2nd-order perturbation theory method*, Ph.D. thesis, University Stuttgart (2016).
[112] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[113] B. Helmich-Paris, J. Chem. Phys. **150**, 174121 (2019).
[114] K. Andersson and B. O. Roos, Chem. Phys. Lett. **191**, 507 (1992).